Social Comparisons in the Classroom Revisited: Insights Into Underlying Processes Using Immersive Virtual Reality as a Research Tool

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von

Lisa Christina Hasenbein, M.Sc.

aus Friedberg

Tübingen

2021

1. Betreuer:	Prof. Dr. Ulrich Trautwein
2. Betreuer:	Prof. Dr. Richard Göllner
Tag der mündlichen Prüfung:	10.12.2021
Dekan:	Prof. Dr. Josef Schmid
1. Gutachter:	Prof. Dr. Richard Göllner
2. Gutachter:	Prof. Dr. Benjamin Nagengast

ACKNOWLEDGMENTS

This work would not have been the same without the invaluable support I received from the many different people who encouraged me and helped me along my academic journey with their honest feedback, inspiring nature, thoughtful advice, and trusted companionship.

First, I would like to thank my supervisors Prof. Dr. Richard Göllner and Prof. Dr. Ulrich Trautwein. Thank you for all your guidance and support throughout these past 3 years. Thank you for challenging me many times and providing an environment that allowed me to learn and grow so much. You both invested a tremendous amount of resources in supporting my dissertation project, and your constant belief in me and my ideas as well as your genuine interest in the progress of my dissertation were a great source of motivation.

Thank you to all my (former) colleagues at the Hector Research Institute of Education Sciences and Psychology and the LEAD Graduate School & Research Network. Thank you for your open doors and for sharing your knowledge and experience with me. Special thanks to Patricia, Laura, and Ann-Kathrin—you helped me navigate this doctoral journey from the first day to the last, and I am deeply grateful for all the laughs we shared and for your constructive advice and pep talks in all stages of this dissertation.

Moreover, I would like to thank all the people who invested their expertise and time to make this research with a virtual reality classroom an actual reality and supported me to get the most out of it. A big thank you to our collaboration partners at the HdM Stuttgart, Prof. Dr. Jens-Uwe Hahn, Stephan, Sandra, and Sophie, and the VRE + HCI project group, particularly Philipp, Hong, and Efe, who introduced me to the fascinating world of eye-tracking research.

I would like to thank the whole KLUG project team. Even though not related to my dissertation, working with you was a fundamental part of my doctoral experience. I am deeply grateful for everything I learned from you and this project that we got off the ground.

Finally, I am beyond blessed with an amazing family and friends who have supported me from near and afar. Thank you, especially to Franzi for making Tübingen feel like home so quickly and for being there throughout all my ups and downs. Thank you to my parents and my sister Katrin—I would not be where I am today without you; thank you for giving me the roots and wings that can carry me through anything. And last but not the very bit least, thank you Carsten—thank you for being you, for always keeping me in balance, for being understanding and supportive, and for having my back no matter what. Danke!

ABSTRACT

The beliefs that students hold about their academic abilities (a.k.a. students' academic self-concepts) play a central role in the success of their learning and academic trajectories. Importantly, students' academic self-concepts are not formed in a vacuum; particularly social comparisons in the classroom are considered a major determinant of students' self-evaluations (Dijkstra et al., 2008; Trautwein & Möller, 2016). Most prominently, in educational psychology research, social comparisons have been assumed to be the cause behind the well-known Big-Fish-Little-Pond effect (BFLPE; Marsh, 1987), a highly robust and generalizable pattern of negative effects of higher class-average (or school-average) achievement on students' academic self-concept while controlling for individual achievement (Marsh et al., 2017; Marsh & Seaton, 2015). However, most BFLPE studies have not provided information about the actual mechanisms that underlie the proposed effects (Dai & Rinn, 2008; Huguet et al., 2009). Research in social psychology, in turn, has yielded a large body of experimental studies that allow in-depth insights into the mechanisms behind social comparisons (Buunk & Gibbons, 2007; Suls et al., 2002; Suls & Wheeler, 2000). However, the respective experimental studies are predominantly lab-based, and it remains unclear how the respective findings apply to actual classroom situations. In other words, the exact mechanisms behind social comparisons in the classroom and how students process social information while learning are still a black box.

The goal of the present dissertation was to open this black box and revisit social comparisons in the classroom by using immersive virtual reality (IVR) as an experimental tool. IVR technology provides an opportunity to integrate ecological validity (as in BFLPE studies based on real-world classroom data) and experimental control (as in social psychology studies on social comparisons) to gain authentic and yet standardized insights into social comparison processes in the classroom (Blascovich et al., 2002; Fox et al., 2009). To this end, in the present dissertation, a theoretical model to examine social comparisons in the classroom was proposed and implemented by means of an IVR classroom that allowed (a) an experimental manipulation of classmates' performance-related behavior, (b) standardized insights into students' cognitive and behavioral responses to respective social comparison information, and ultimately, (c) causal conclusions to be drawn about effects on students' self-evaluations.

Therefore, the present dissertation had two objectives pertaining to the theoretical as well as the methodological advancement of research on social comparisons in the classroom. First, pursuing a more in-depth theoretical understanding of social comparisons and the respective processing of social information in the classroom, the present dissertation aimed to

identify *covert* as well as *overt* social comparison behaviors that reflect students' *cognitive* and *behavioral* responses to social comparison information in an IVR classroom and ultimately explain individual differences in students' self-concepts. Second, the present dissertation aimed to provide insights into how IVR classrooms can be used as an experimental tool to gain insights into classroom processes, such as social comparisons and beyond. To address these objectives, the present dissertation drew on three empirical studies with an IVR classroom. The three studies used data from N = 381 sixth-graders who experienced an IVR classroom scenario with a class of virtual peer learners. The virtual classmates displayed experimentally varied hand-raising behavior to simulate distinct class performance levels.

The first study (*Does a 15-minute exposure to strong classmates affect students' self-concept? An experimental test of the Big-Fish-Little-Pond effect using an immersive virtual reality classroom*) experimentally tested (a) the extent to which students recognized their virtual classmates' implicit performance-related behavior (i.e., hand-raising) as social comparison information and (b) how respective perceptions explained differences in students' self-evaluations. In line with the BFLPE, the results showed a negative effect of higher performing classmates on students' situational self-concept. Moreover, the results showed that classmates' hand-raising behavior had a positive effect on students' perceptions of the class performance level, and the perceived class performance level fully mediated the effect of classmates' hand-raising on students' situational self-concept. The study thereby contributed to a more in-depth theoretical understanding of social comparisons in the classroom as it (a) provided experimental evidence for the BFLPE based on classmates' performance-related behavior and (b) showed that students' *covert* cognitive responses to social comparison information (i.e., their interpretations of their classmates' performance-related behavior) explained the BFLPE. By reproducing the empirically well-established BFLPE in the IVR classroom setting, the study additionally provided evidence for the validity of the IVR classroom as an experimental tool.

The second study (*Do students actively seek comparisons with others? Using eye-movement data from a virtual reality classroom to uncover social information processing*) used eye-tracking data from the IVR classroom as an overt behavioral indicator of the extent to which students actively attended and responded to different levels of their virtual classmates' performance-related behavior (i.e., hand-raising). Analyses of students' gaze data indicated that students actively processed classmates' performance-related behavior and attended and responded particularly to a clear minority or majority of hand-raising classmates. In line with assumptions associated with the BFLPE, students who paid more attention to social comparison information (as indicated by increased visual attention paid to classmates) showed lower situational self-concepts. Taken together, the study contributed to a theoretically more nuanced understanding of social comparisons in the classroom and demonstrated how students' *overt* behavioral responses to social comparison information—measured by visual attention paid to peer learners—can provide insights into the mechanisms that underlie the BFLPE.

The third study (*Configuring an immersive virtual reality classroom for educational research and practice: Implications from students' gaze-based attention networks*) focused on the configuration of an IVR classroom as an experimental tool and examined three salient classroom features (namely, the location of a student's seat in the classroom, the style with which virtual avatars were represented, and virtual classmates' performance-related behavior) to examine how these IVR classroom configurations affect students' processing of information presented in the IVR classroom. Results based on students' gaze data showed that the IVR configurations were systematically associated with differences in students' gaze-based attention networks. Moreover, interindividual differences in students' gaze-based attentional networks during instruction (e.g., a gaze focused on virtual classmates compared with gazes centering on instructional content) were related to students' self-reported interest in the IVR lesson, situational self-concept, and performance after the IVR lesson. The study thereby provided insights into how an IVR classroom can be used as a methodological approach for classroom research, specifically concerning (a) the role of different IVR classroom configurations for students' processing of information and (b) the use of sophisticated gaze analysis to gain insights into students' learning experiences in the IVR classroom.

By using an IVR classroom as an experimentally controlled yet authentic research setting, the present dissertation provides novel insights into social comparison processes in the classroom. The present dissertation advances the theoretical understanding of social comparison processes in the classroom by describing students' social comparison behavior via cognitive and behavioral responses to social information that ultimately explain individual differences in students' self-concept. Moreover, the present dissertation demonstrates how IVR classrooms and the corresponding standardized process data can be used to gain insights into classroom processes, such as social comparisons. This dissertation thereby provides implications for research on both social comparisons in the classroom and the use of IVR as an experimental tool in educational and social psychology research in general.

ZUSAMMENFASSUNG

Die Überzeugungen, die Lernende über ihre akademischen Fähigkeiten haben (auch bekannt als akademisches Selbstkonzept), spielen eine zentrale Rolle für ihren Lernerfolg und ihre akademische Laufbahn. Das akademische Selbstkonzept bildet sich allerdings nicht im luftleeren Raum; insbesondere soziale Vergleiche im Klassenzimmer gelten als wichtige Determinante der entsprechenden Selbsteinschätzungen von Lernenden (Dijkstra et al., 2008; Trautwein & Möller, 2016). In der pädagogisch-psychologischen Forschung wird angenommen, dass soziale Vergleiche die Ursache für den bekannten Big-Fish-Little-Pond Effekt (BFLPE; Marsh, 1987) sind, ein äußerst robustes und verallgemeinerbares Muster negativer Auswirkungen von höherer durchschnittlicher Klassen-/Schulleistungen auf das akademische Selbstkonzept von Lernenden unter Kontrolle der individuellen Leistung (Marsh et al., 2017; Marsh & Seaton, 2015). Die meisten BFLPE-Studien liefern jedoch keine Belege für die tatsächlichen Mechanismen, die dem Effekt zugrunde liegen (Dai & Rinn, 2008; Huguet et al., 2009). Im Gegenzug bietet die sozialpsychologische Forschung eine große Zahl an experimentellen Studien, die tiefere Einblicke in die Mechanismen sozialer Vergleiche ermöglichen (Buunk & Gibbons, 2007; Suls et al., 2002; Suls & Wheeler, 2000). Entsprechende experimentellen Studien finden jedoch überwiegend im Labor statt und es bleibt unklar, wie sich die Ergebnisse auf tatsächliche Situationen im Klassenzimmer übertragen lassen. Mit anderen Worten: Die genauen Mechanismen sozialer Vergleiche im Klassenzimmer und die Art und Weise, wie Lernende soziale Informationen während des Unterrichts verarbeiten, sind noch eine Blackbox.

Die vorliegende Dissertation hatte zum Ziel, diese Blackbox zu öffnen und soziale Vergleiche im Klassenzimmer durch den Einsatz von immersiver virtueller Realität (IVR) als experimentelles Instrument neu zu untersuchen. IVR-Technologie bietet die Möglichkeit, ökologische Validität (s. BFLPE-Studien basierend auf realen Klassenzimmerdaten) und experimentelle Kontrolle (s. sozialpsychologische Studien zu sozialen Vergleichen) zu kombinieren, um authentische und dennoch standardisierte Einblicke in soziale Vergleichsprozesse im Klassenzimmer zu gewinnen (Blascovich et al., 2002; Fox et al., 2009). Im Rahmen der vorliegenden Dissertation wurde ein theoretisches Modell zur Untersuchung sozialer Vergleiche im Klassenzimmer vorgeschlagen und mittels eines IVR-Klassenzimmers umgesetzt, das (a) eine experimentelle Manipulation des leistungsbezogenen Verhaltens von Mitschülerinnen und Mitschülern, (b) standardisierte Einblicke in die kognitiven und verhaltensbezogenen Reaktionen der Lernenden auf die jeweiligen sozialen Vergleichs-

informationen und schließlich (c) kausale Schlussfolgerungen über die Auswirkungen auf die Selbsteinschätzungen der Lernenden ermöglicht.

Die vorliegende Dissertation verfolgte daher zwei Ziele, die sowohl eine theoretische als auch eine methodische Weiterentwicklung der Forschung zu sozialen Vergleichen im Klassenzimmer betreffen. Erstens sollte ein tieferes theoretisches Verständnis sozialer Vergleiche und der entsprechenden Verarbeitung sozialer Informationen im Klassenzimmer erlangt werden. Die vorliegende Dissertation hatte zum Ziel, *intrapersonale* und *beobachtbare* Aspekte des sozialen Vergleichsverhaltens zu identifizieren, welche die *kognitiven* und *verhaltensbezogenen* Reaktionen der Lernenden auf soziale Vergleichsinformationen in einem IVR-Klassenzimmer widerspiegeln und letztlich individuelle Unterschiede im Selbstkonzept der Lernenden erklären. Zweitens sollte die vorliegende Dissertation Erkenntnisse darüber liefern, wie ein IVR-Klassenzimmer als experimentelles Instrument genutzt werden kann, um Einblicke in Unterrichtsprozesse wie soziale Vergleiche zu gewinnen. Die vorliegende Dissertation stützte sich dazu auf drei empirische Studien mit einem IVR-Klassenzimmer. Die drei Studien verwendeten Daten von N = 381 Sechstklässlerinnen und Sechstklässlern, die eine IVR-Unterrichtseinheit mit einer Klasse virtueller Mitschülerinnen und Mitschüler erlebten. Die virtuellen Mitschülerinnen und Mitschüler zeigten experimentell variiertes Meldeverhalten, um unterschiedliche Leistungsniveaus der Klasse zu simulieren.

In der ersten Studie (*Does a 15-minute exposure to strong classmates affect students' self-concept? An experimental test of the Big-Fish-Little-Pond effect using an immersive virtual reality classroom*) wurde experimentell untersucht, (a) inwieweit Lernende das implizite leistungsbezogene Verhalten ihrer virtuellen Klassenkameradinnen und Klassenkameraden (d.h. deren Meldeverhalten) als soziale Vergleichsinformation erkannten und (b) inwieweit die jeweiligen Wahrnehmungen Unterschiede in der Selbsteinschätzung der Lernenden erklärten. In Übereinstimmung mit dem BFLPE zeigten die Ergebnisse einen negativen Effekt von leistungsstärkeren Mitschülerinnen und Mitschülern auf das situative Selbstkonzept der Lernenden. Darüber hinaus zeigten die Ergebnisse, dass das Meldeverhalten in der Klasse einen positiven Effekt auf die Wahrnehmung des Leistungsniveaus der Klasse hatte; das wahrgenommene Leistungsniveau der Klasse erklärte den Effekt des Meldeverhaltens auf das situative Selbstkonzept der Lernenden vollständig. Die Studie trug damit zu einem vertieften theoretischen Verständnis sozialer Vergleiche im Klassenzimmer bei, indem sie (a) auf der Grundlage des leistungsbezogenen Verhaltens der Mitschülerinnen und Mitschüler einen experimentelle Beleg für den BFLPE lieferte und (b) zeigte, dass die kognitive Reaktion der Lernenden auf die sozialen Vergleichsinformationen (d.h. ihre Interpretation des leistungsbezogenen Verhaltens in der Klasse) den BFLPE erklärte. Durch die Reproduktion des empirisch gut belegten BFLPE in der IVR-Klassenzimmerumgebung lieferte die Studie zusätzlich Beweise für die Validität des IVR-Klassenzimmers als experimentelles Instrument.

In der zweiten Studie (*Do students actively seek comparisons with others? Using eye-movement data from a virtual reality classroom to uncover social information processing*) wurden Blickbewegungsdaten aus dem IVR-Klassenzimmer als verhaltensbezogener Indikator für die Reaktion von Lernenden auf das leistungsbezogene Verhalten (d.h. das Meldeverhalten) ihrer virtuellen Klassenkameradinnen und Klassenkameraden verwendet. Die Analyse der Blickbewegungsdaten zeigte, dass die Lernenden das leistungsbezogene Verhalten in der Klasse aktiv verarbeiteten und insbesondere auf eine deutliche Minder- oder Mehrheit von sich meldenden Mitschülerinnen und Mitschülern reagierten. In Übereinstimmung mit den dem BFLPE zugrunde liegenden Annahmen, zeigten Lernende, die mehr auf soziale Vergleichsinformationen achteten, ein geringeres situationales Selbstkonzept. Insgesamt trug die Studie so zu einem theoretisch differenzierteren Verständnis sozialer Vergleiche im Klassenzimmer bei und zeigte, wie die beobachtbare Verhaltensreaktion von Lernenden auf soziale Vergleichsinformationen—gemessen durch die visuelle Aufmerksamkeit auf Mitschülerinnen und Mitschüler—Einblicke in die zugrunde liegenden Mechanismen des BFLPE liefern kann.

Die dritte Studie (*Configuring an immersive virtual reality classroom for educational research and practice: Implications from students' gaze-based attention networks*) konzentrierte sich auf die Konfiguration eines IVR-Klassenzimmers als experimentelles Instrument. Die Studie nahm drei zentrale Klassenzimmermerkmale (die Sitzposition der Lernenden im Klassenzimmer, den Visualisierungsstil der virtuellen Avatare und das leistungsbezogene Verhalten der virtuellen Mitschülerinnen und Mitschüler) in den Blick, um zu untersuchen, wie diese die Verarbeitung der im IVR-Klassenzimmer präsentierten Informationen durch die Lernenden beeinflussen. Die Ergebnisse basierend auf den Blickbewegungsdaten der Lernenden zeigten, dass die IVR-Konfigurationen systematisch mit Unterschieden in der blickbasierten Aufmerksamkeitsverteilung der Lernenden im Klassenzimmer zusammenhingen. Darüber hinaus standen interindividuelle Unterschiede in der blickbasierten Aufmerksamkeitsverteilung der Lernenden während des Unterrichts (z.B. Blickfokus auf virtuelle Mitschülerinnen und Mitschülern im Vergleich zum Blickfokus auf dem Lernmaterial) in Zusammenhang mit dem von Lernenden berichteten Interesse an der IVR-Unterrichtseinheit, ihrem situativen Selbstkonzept und ihrer Leistung nach der IVR-Unterrichtseinheit. Die Studie lieferte somit Erkenntnisse darüber, wie ein IVR-Klassenzimmer als methodischer Ansatz für die Unterrichtsforschung genutzt werden kann, insbesondere im Hinblick auf (a) die Rolle verschiedener Konfigurationen des IVR-Klassenzimmer für die Informationsverarbeitung der Lernenden und (b) die Nutzung und Analyse von Blickbewegungsdaten als blickbasierte Aufmerksamkeitsverteilungen, um Einblicke in die Erfahrung der Lernenden im IVR-Klassenzimmer zu gewinnen.

Durch die Verwendung eines IVR-Klassenzimmers als experimentell kontrollierbare und dennoch authentische Forschungsumgebung lieferte die vorliegende Dissertation neue Erkenntnisse über soziale Vergleichsprozesse im Klassenzimmer. Die vorliegende Dissertation konnte das theoretische Verständnis sozialer Vergleichsprozesse im Klassenzimmer erweitern, indem sie das soziale Vergleichsverhalten von Lernenden durch kognitive und verhaltensbezogene Reaktionen auf soziale Informationen beschreibt, die letztlich individuelle Unterschiede im Selbstkonzept der Lernenden erklären. Darüber hinaus konnte die vorliegende Dissertation zeigen, wie ein IVR-Klassenzimmer und entsprechende standardisierte Prozessdaten genutzt werden können, um Einblicke in Klassenzimmerprozesse wie soziale Vergleiche zu gewinnen. Dies hat Implikationen sowohl für die Forschung zu sozialen Vergleichen im Klassenzimmer als auch für den Einsatz von IVR als experimentelles Instrument in der pädagogischen und sozialpsychologischen Forschung im Allgemeinen.

CONTENTS

1

INTRODUCTION AND THEORETICAL BACKGROUND

1 Introduction and Theoretical Background

"Tell me something about yourself. What are you good at? And what not so much?" when asked this question, everyone will immediately be able to list a number of things without having to think too much. Whatever the answers to these questions are, these beliefs we hold about ourselves and our abilities[1] matter: Whether consciously or not, what we think we are good (or not good) at affects how we feel (e.g., Cross et al., 2003), what we are interested in and enjoy doing (e.g., Arens et al., 2019; Trautwein et al., 2006), our educational choices, and the professional paths we choose (e.g., Göllner et al., 2018; von Keyserlingk et al., 2020). It is therefore crucial to understand how we arrive at these beliefs about ourselves and our abilities.

A critical point to consider is that human perception and evaluation is relative (see, e.g., Kahneman & Miller, 1986). Generally speaking, how we perceive and evaluate something is—sometimes knowingly but oftentimes unwittingly—based on a comparison with a certain standard. I would like to use a phenomenon from perceptual psychology to illustrate this. The so-called Ebbinghaus illusion (also referred to as Titchener Circles; see Figure 1) shows two sets of circles consisting of a center circle surrounded by six circles that are either considerably smaller or larger in size compared with the center circle. The center circle is the same size in both panels—and yet we perceive it differently.

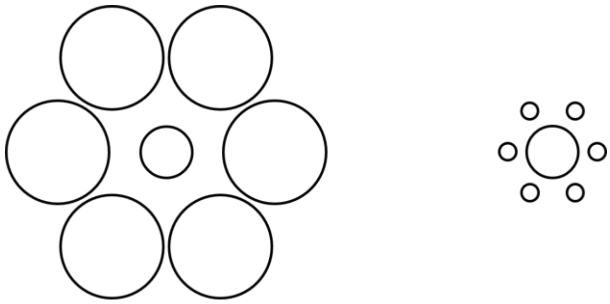


Figure 1. The Ebbinghaus illusion (adapted from Massaro & Anderson, 1971).

Of course, this is different from humans' perceptions of themselves: We do not observe ourselves from an outside point of view, and in many situations, our and others' characteristics are not explicitly displayed and cannot be objectively measured. And yet, this perceptual phenomenon offers a great visual representation of how the relativity of self-evaluations can be understood, emphasizing the role of subjective perceptions. Imagine that the sizes of these

[1] There is a great deal of variety in the terms that refer to what was introduced here as "beliefs about ourselves and our abilities." In this introductory section, I use this description and the term *self-evaluation* for all related constructs (e.g., self-beliefs, self-perception, self-view, self-esteem, self-concept, self-image, self-schema, self-worth; for an overview, see, e.g., Schunk & Pajares, 2005).

circles indicate levels of competence. The self as the center circle is the same in the two different scenarios, and yet how this self and one's respective abilities are perceived is not always the same. We compare our abilities with a variety of comparison targets (see, e.g., Trautwein & Möller, 2016, with regard to the educational context). For example, let the surrounding circles be other domains, past experiences, or other people—each of them with a specific level of status or competence. Particularly social comparisons with other people have been found to play a central role in shaping the beliefs we hold about ourselves (see fundamental theoretical works by Festinger, 1954; Frank, 1985, 2013; Hyman, 1942). As Baumeister and Bushman (2011) put it, "In fact, if you grew up on a deserted island and never met other human beings, you might hardly have a 'self' at all in the usual sense" (p. 60).

In educational psychology, the effect of certain reference groups on students' evaluations of their own academic abilities is best known as the Big-Fish-Little-Pond effect (BFLPE; Marsh, 1987). Using Figure 1 to explain the BFLPE, the center circle would represent a student with a certain performance level. Being placed in a class with a higher performance level (represented by the larger surrounding circles in the left panel), the student will evaluate their own abilities as worse (i.e., the center circle seems smaller) compared with when the student is placed in a class with an overall lower performance level (represented by the smaller surrounding circles in the right panel). Similar to the center circle seeming smaller or larger depending on the surrounding circles, the student evaluates their own abilities as better or worse depending on the performance levels of the students they are surrounded by—and presumably compare themselves with. In the two scenarios, when measured objectively, the size of the center circle—or the student's ability—is the same, but the student's perceptions and evaluations of their ability levels in the two scenarios are different.

There is a large body of empirical research supporting the BFLPE (see, e.g., reviews by Fang et al., 2018; Marsh et al., 2017; Marsh & Seaton, 2015). However, despite the large body of supporting evidence, BFLPE research leaves open the central question about the underlying mechanisms: What happens in the classroom that ultimately leads to individual differences in students' evaluations of their abilities? This question becomes particularly apparent when simultaneously considering conceptualizations from social psychology that highlight the active nature of the self (see, e.g., the review by Mischel & Morf, 2003). Again, speaking in the terms used in Figure 1: Whereas BFLPE research is not so much different from observing this phenomenon from the outside (i.e., using a student's individual achievement level in relation to the average level of achievement in the class to predict differences in self-evaluations), the center circle is in fact an active person who—according to the theoretical assumptions underlying the BFLPE—compares themself with others. However, this aspect has largely been neglected in typical BFLPE research. It is widely acknowledged that the reference group makes the difference, but is it because the teacher points out the differences among the students? Or do students use grades and explicitly communicated grade point averages as the decisive standard to make comparisons with? Do students rely on their peers' comments about their own performance, and do comments—such as "This smarty-pants knows it all, once again" in contrast to "Oh, you silly, you don't understand a word, do you?"—from peers make all the difference? Or are social comparisons in fact so inherent to human nature that students themselves seek comparisons with their peers even if the respective information they use for these comparisons might sometimes be only implicitly provided by their reference group (e.g., in their peers' performance-related behavior)?

Providing an answer to questions as such is not a simple endeavor: Classroom situations are complex and dynamic, and isolating and manipulating single variables (e.g., teachers' feedback, peers' performance-related behavior) to observe the respective effects comes with certain challenges. Whereas the lab experiments that are often used in social psychology research on social comparisons allow insights as such, such research designs cannot capture the authenticity of what happens in a real classroom. The aim of the present dissertation is to use an immersive virtual reality classroom as a research tool to get the best of both worlds—experimental control from lab settings and authenticity from real-world research—in order to gain insights into exactly how social comparisons in the classroom affect students' beliefs about their own academic abilities. Whereas the potential of immersive virtual environments is evident (see, e.g., Bainbridge, 2007; Blascovich et al., 2002; Fox et al., 2009), not many studies have made use of this potential in social and educational research to date, specifically with regard to classroom research. Against this background, the present dissertation additionally aims to provide insights into how immersive virtual reality classrooms can be used for research purposes, specifically to gain insights into classroom processes, such as social comparisons.

The present dissertation is structured as follows: Chapter 1 explains the dissertation's theoretical background in more detail. Using a definition of the central outcome, students' beliefs about their own abilities (i.e., their academic self-concept; Chapter 1.1), I take a closer look at social comparisons in the classroom as one of the major determinants of students' self-concept and outline the central open questions on the relationships between individual differences in self-concept and social comparison processes in the classroom (Chapter 1.2). I then introduce immersive virtual reality as a research tool that provides a promising avenue for answering the respective questions (Chapter 1.3). Based on this, Chapter 2 describes the dissertation’s overall aims and research questions. To address these, three empirical studies are presented in Chapters 3 to 5. Chapter 6 closes with a general discussion of the three studies with regard to the overall aims and research questions of this dissertation. More specifically, I use the closing chapter to discuss theoretical and methodological contributions with regard to research on social comparison processes in the classroom and the use of immersive virtual reality as a research tool (Chapter 6.1), the dissertation’s overall strengths and limitations (Chapter 6.2), as well as implications for future research and practice (Chapter 6.3) before closing with a general conclusion (Chapter 6.4).

1.1 Why Social Comparisons in the Classroom Matter: Students' Academic Self-Concept

"I am good at mathematics," "I struggle when I try to learn foreign languages," "I can quickly solve technical problems"—these are all expressions of students' beliefs about their abilities in a certain domain. These beliefs are also referred to as academic self-concept, which is one of the central constructs in educational psychology. In a nutshell, the present dissertation defines academic self-concept as individual beliefs about one's abilities in a certain academic domain (Shavelson et al., 1976). Academic self-concept is therefore one part of a more general and overarching self-evaluation (e.g., "All in all, I am happy with who I am"), which is typically referred to as self-esteem or self-worth (e.g., Marsh & Yeung, 1997). Moreover, self-concept needs to be differentiated from self-efficacy, which also describes individual competence beliefs but is prospectively oriented toward the future, whereas self-concept is retrospective and refers to aggregated self-evaluations that are based on past experiences (Bong & Skaalvik, 2003; Marsh, Pekrun, et al., 2019).

In the following, I first explain the relevance of academic self-concept (Chapter 1.1.1), before going into more detail about how exactly academic self-concept can be defined by considering both structural conceptualizations from educational psychology and more process-oriented conceptualizations from publications in social psychology (Chapter 1.1.2). On the basis of these conceptualizations, I then outline the determinants of academic self-concept and introduce the role of social comparisons (Chapter 1.1.3).

1.1.1 The Role of Academic Self-Concept in Student Learning

In educational psychology, there is a large body of research on students' academic self-concept. On the one hand, the importance of self-concept pertains to the general notion that self-concept is related to well-being (Cross et al., 2003). Developing a positive concept of one's own abilities is considered an important part of individual development and plays a central role in positive psychology (Marsh & Craven, 2006; Marsh et al., 2017; Schunk & Pajares, 2005; Seligman & Csikszentmihalyi, 2000). On the other hand, the importance of academic self-concept pertains to the fact that it is associated with achievement-related outcomes.

Students' beliefs about their abilities in a certain domain (i.e., their domain-specific academic self-concept) were found to be related to actual achievement early on (Marsh, 1990a; Wylie, 1979). Ever since that time, empirical evidence of the association between academic self-concept and achievement has been growing. There is a large body of research suggesting a positive effect of academic self-concept on subsequent achievement (i.e., the so-called skill development model; Baumeister et al., 2003; Marsh & Craven, 2006). The most extensive meta-analytic review in this regard was conducted by Valentine et al. (2004), who looked at 60 studies involving over 50,000 students and found substantive positive effects of academic self-concept on subsequent achievement—particularly for domain-specific measures of academic self-concept and respective achievement scores and also when controlling for prior achievement. At the same time, there is support for the so-called self-enhancement model, which suggests an effect of academic achievement on academic self-concept (e.g., Guay et al., 2003; Huang, 2011).[2] Integrating the two perspectives, a number of longitudinal studies have provided robust evidence of reciprocal effects between self-concept and academic achievement (reciprocal effects model; Marsh & Craven, 2006; Marsh & Martin, 2011).

So how exactly are beliefs about one's own abilities associated with differences in achievement? The short answer is that academic self-concept is related to factors that promote learning or are associated with effective learning processes. The respective evidence can be theoretically integrated using the expectancy-value model (Eccles, 1983; Wigfield & Eccles, 1992). The model suggests that academic self-concept is closely related to students' expectancy of success; more specifically, it is considered a mediator between students' performance-related experiences and their motivational and learning behaviors. In this vein, for instance, Helmke (1992) found that high mathematics self-concept is related to greater engagement and willingness to exert effort. Similarly, Dickhauser and Reinhard (2006) concluded that the predictive power of self-concept for favorable learning outcomes is related to "effortful information processing" (p. 499). Moreover, self-concept has been found to be related to domain-specific motivation and the value that a person attributes to specific domains (Arens et al., 2019; Brisson et al., 2017; Cambria et al., 2017; Guay et al., 2010) as well as a person's interest in specific subjects (Köller et al., 2000; Trautwein et al., 2006). Importantly, the associations of academic self-concept in a certain domain with a student's respective interest and motivation in that domain do not end when students leave the classroom: Academic self-concept has been found to be related to academic course selection (Marsh & Yeung, 1997; von Keyserlingk et al., 2020) and overall academic trajectories (Göllner et al., 2018; Nagengast & Marsh, 2012).

Considering its far-reaching effects on students' learning and academic trajectories, it is thus no wonder that academic self-concept is of central interest to researchers in educational

[2] Whereas I presented these findings here as clear indications of causal relationships, self-evaluations are not purely objective, and therefore, the respective associations need to be interpreted with caution in terms of causality.

psychology. The last few decades have produced vast numbers of theoretical and empirical studies that have been aimed at understanding how academic self-concept can be conceptualized and how its development and the emergence of respective differences can be explained. The following subchapters provide an overview.

1.1.2 Conceptualizations of (Academic) Self-Concept

The origin of self-concept research is often attributed to the work of William James (James, 1892/1999). James (1892/1999) distinguished between the "self as a knower" (i.e., the person as an actor and thinker themself) and the "self as known" (i.e., all aspects about the person that the person can be aware of, e.g., abilities and dispositions, social perceptions of others, possessions, and relationships). James' theory can be summed up with the sentence "I know Me," whereby *I* represents the "self as a knower" and *Me* represents the "self as known." Most interestingly for subsequent conceptualizations of self-concept, James described the *Me* in that sentence as hierarchical and multidimensional and argued that people's beliefs about themselves are based on single experiences of success and failure and their standing in relation to others (James, 1892/1999). Whereas not all of James' theoretical assumptions have been reinforced by empirical research in the decades that followed, these assumptions are still reflected in the most prominent conceptualizations of self-concept to date.

Notably, self-concept has inspired a large body of research in both educational psychology and social psychology. The conceptualizations and methodologies used in these two disciplines are not mutually exclusive, but they are systematically different with regard to their overall research focus: Speaking in James' (1892/1999) terms, research in educational psychology has been more interested in the *Me* part of "I know Me" (i.e., the theoretical and empirical structure of self-concept), whereas research in social psychology is driven by a more process-oriented focus, focusing on the *I* as an active and dynamic self (see, e.g., the review by Mischel & Morf, 2003).

Academic Self-Concept in Educational Psychology: A Structural Definition

The starting point for research in educational psychology on self-concept in the way we know it today is typically set in the 1970s (see, e.g., Marsh & Craven, 1997). In 1976, Shavelson et al. wrote a review of self-concept research to this date. They lamented about what was later called the "dustbowl phase" (Hattie, 1992; Marsh & Craven, 1997) of self-concept research as it was missing a rigorous theoretical basis regarding the structure of self-concept, which was often reflected in inconsistent measurements and methodology. Going back to theoretical descriptions by James (1892/1999), Shavelson et al. (1976) suggested a multidimensional and hierarchical structure for self-concept. The so-called Shavelson model (Shavelson et al., 1976; see Figure 2) suggests that individuals organize their self-concepts in different areas, whereby one of these areas is academic self-concept, which is further divided into different subjects. Some adaptions to the model have been made, such as the Marsh/Shavelson model (Marsh, 1990c; Marsh & Shavelson, 1985), which suggests verbal (e.g., languages, history) and mathematical (e.g., mathematics, physics, chemistry) academic self-concepts as two subordinate factors instead of one general academic self-concept. In general, following Shavelson et al.'s first structural self-concept definition, research on self-concept has led to fairly established conceptualizations; the multidimensionality and therefore the domain-specificity of self-concept as well as the notion of self-concept's hierarchical structure (i.e., more general self-concept on top and more differentiated the lower you get) has been backed up by lots of empirical evidence and is generally accepted today (see, e.g., Marsh, 1990b; Marsh & Craven, 1997; Schunk & Pajares, 2005; Trautwein & Möller, 2016).

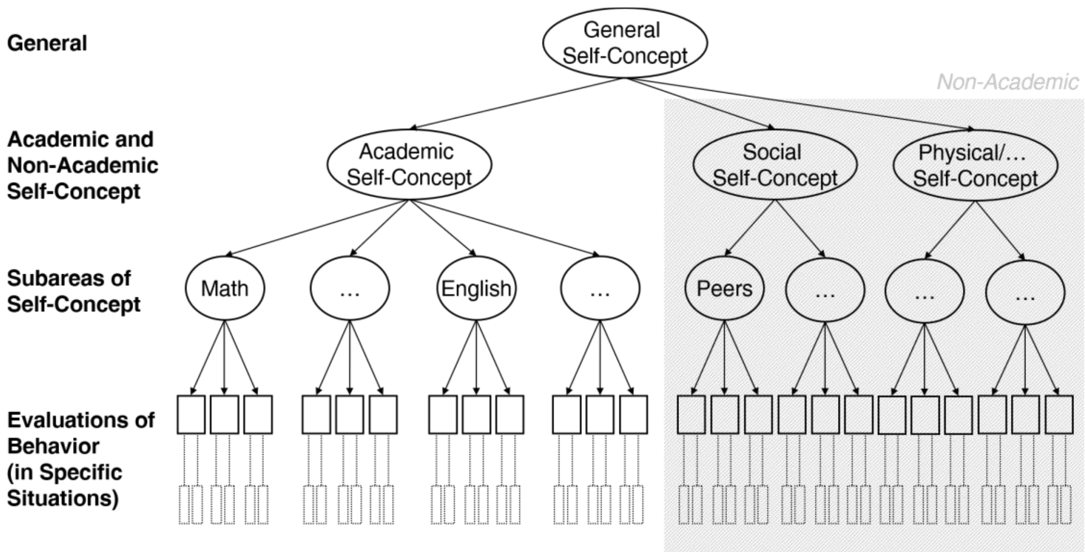


Figure 2. The Shavelson model (adapted from Shavelson et al., 1976, p. 413).

When Shavelson et al. (1976) proposed the multidimensional and hierarchical structure of self-concept, they noted that the structure of self-concept gets more differentiated with age from childhood to adulthood. As would be expected from a Piagetian perspective (see Piaget, 1960), self-concept has been found to develop from concrete-operational descriptions of the self to more abstract ones, accompanied by an increasing integration of different sources of information, thus increasing the differentiation of self-concept with time (Cairns et al., 1990; Harter, 1986, 1998, 1999). Importantly, if a student thinks, "I am not good at mathematics," it is unlikely to be only a rational description of their own abilities but includes an evaluation thereof (i.e., self-concept has both descriptive and evaluative components; Shavelson et al., 1976). The evaluative component of self-concept is argued to be particularly reinforced in academic settings that naturally have an evaluative character; not only is performance assessed and witnessed on an almost daily basis, but students are also repeatedly told that their performance matters whether it is for their grade on the next exam or their general academic and life trajectories. As Shavelson et al. (1976) put it: “Self-concept is inferred from a person’s response to situations” (p. 411). More specifically, the lowest level of the Shavelson model points to the role of single experiences of success and failure (i.e., “evaluations of behavior in specific situations”; Shavelson et al., 1976, p. 413).

Aiming to identify the role of single situations and experiences and their relationships to more stable (academic) self-conceptualizations, it is important to note that research in educational psychology typically considers the stability and situation-specificity of self-concept more from a structural/methodological perspective. For instance, newer developments, such as the Nested Marsh/Shavelson model (Brunner et al., 2010), and respective discussions (Braun et al., 2020) demonstrate that questions about how more general high-level and more specific components of self-concept are related and are ideally modeled are still open and under discussion. I would like to highlight that the present dissertation does not aim or claim to contribute to these questions but is rather aimed at identifying the processes that lead to the emergence of individual differences in self-concept in specific situations. To this end, the present dissertation focuses primarily on the lowest level of the Shavelson model (i.e., self-evaluations in specific situations; Shavelson et al., 1976). In order to zoom in on this level, I draw on conceptualizations of self-concept from social psychology, which applies a more process-oriented approach to integrate more stable and simultaneous situation-specific facets of self-concept, highlighting the role of the social environment in specific situations and the respective processing of social information in these.

Self-Concept in Research From Social Psychology: A Process-Oriented Definition

Social psychological conceptualizations of self-concept were visibly influenced by two central developments in the 20th century. First, at the beginning of the century, a new research stream emerged, highlighting the substantial influence of interactions with the social environment for individuals (symbolic interactionism; Baldwin, 1897; Cooley, 1902; Mead, 1934). Second, with the cognitive revolution in the mid-20th century, psychological research began to focus on information processing after behaviorism had dominated most psychological research in the preceding decades (see, e.g., Newell & Simon, 1972; Skinner, 1985). With this shift came even greater interest in the specific role of information from the social environment for self-concept.

The dynamic self-concept model (Markus & Wurf, 1987; see Figure 3) summarizes how research in social psychology has typically conceptualized self-concept.[3] The model puts a major focus on the processing of (social) information as a source of knowledge about the self (see similar theoretical assumptions and conceptualizations by Filipp, 1979; Hannover, 1997; Swann, 1987; Tesser, 1988). Markus and Wurf (1987) suggested that people's self-concept is dynamic in the sense that people have a relatively stable image of themselves (based on past experiences and perceptions of themselves as well as beliefs about how they would like to be). However, different situations activate different aspects of this self-concept (see also Markus & Kunda, 1986). As Demo (1992) summarized it:

> To understand self-concept, we must conceptualize it as a moving baseline with fluctuations across situations […]. This involves recognizing that the self-concept is simultaneously a complex structure and a process, that it is stable, but that it is also dynamic. (p. 304)

Markus and Wurf (1987) labeled this dynamic and situation-specific facet of self-concept the "working self," which shapes and is shaped by responses to the present (social) environment, whereby they distinguished between *intra*personal and *inter*personal behavior in response to the social environment. *Intra*personal behavior refers to the identification and interpretation of self-relevant information and affect regulation; *inter*personal behavior concerns interaction strategies with the social environment and certain situational choices (Markus & Wurf, 1987). Given, for instance, students in a class, the model suggests that students process the prevailing (social) circumstances in the classroom environment differently depending on what they identify as relevant for themselves, in terms of the attributions and interpretations they draw, and with regard to their emotional and motivational responses to those circumstances (i.e., intrapersonal behavior). Moreover, the model proposes that students' attention to distinct aspects of the situation and engagement in different interactions with the social environment (i.e., interpersonal behavior) depend on—and ultimately reciprocally shape—their self-concept.

[3] Social psychological conceptualizations of self-concept refer to a general self-concept (similar to the apex of the Shavelson model; Shavelson et al., 1976) and a wide variety of social contexts. The present dissertation applies the inherent process-oriented conceptualizations to students' (domain-specific) academic self-concepts and the classroom context.

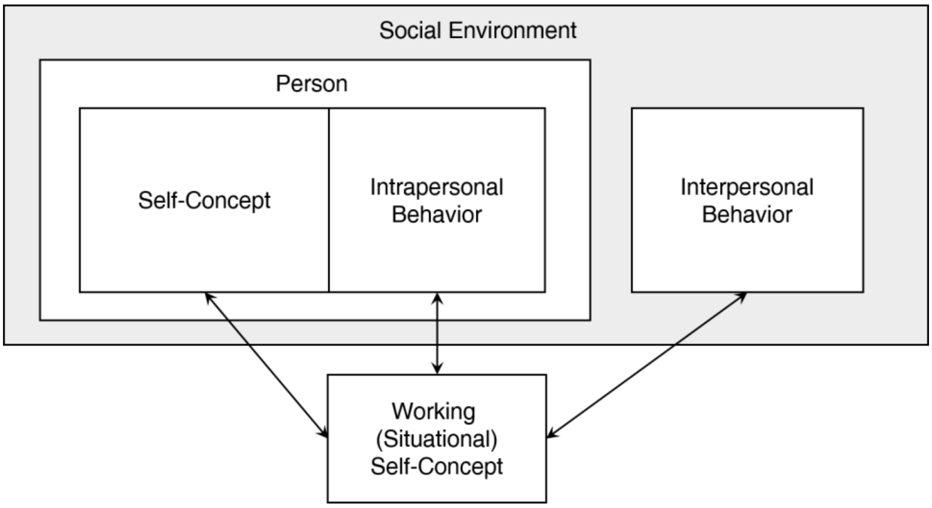


Figure 3. Dynamic self-concept (adapted from Markus & Wurf, 1987, p. 315).

Notably, both structural and more process-oriented conceptualizations of self-concept include the notion that individuals' self-concept is based on specific situational experiences. Particularly with regard to the emergence of individual differences in self-concept, it is crucial to understand how exactly certain situations and—speaking in terms of the dynamic self-concept model—interactions with the social environment and respective inter- and intrapersonal behavior are related to differences in self-concept (i.e., what aspects of the social environment impact self-concept and how). The following chapter provides an overview of the roles that situational cues and other factors play in determining students' academic self-concept.

1.1.3 Determinants of Students' Academic Self-Concept

As noted in the previous chapter, academic self-concept is assumed to be determined by single experiences of success and failure. The Self-Memory System (Conway, 2005; Conway & Pleydell-Pearce, 2000) implies that people experience numerous self-defining moments (among others, so-called first-time experiences) throughout their lives, whereby repeated similar experiences reinforce respective beliefs about the self. This idea is in line with models from personality psychology (see, e.g., the TESSERA framework by Wrzus & Roberts, 2017) that suggest that personality development can be based on single triggering situations and associated expectancies and responses, whereby long-term changes in personality result from repeated short-term experiences. For instance, students could fail to successfully answer the physics teacher's question once or they could be among the top 10% of students in a yearly mathematics competition, and they would probably remember these experiences, which would have a negative or positive effect, respectively, on their (situational) academic self-concept.

However, if students failed to answer the physics teacher's question repeatedly and ended up among the top 10% of the yearly mathematics competition numerous times in a row, the effect on their academic self-concept would manifest itself and become even more pronounced. Importantly with regard to the classroom context, the more constant environments are in terms of their expectancies (e.g., role-appropriate behavior), the more likely it is that similar repeated experiences and corresponding long-term development will occur (Wrzus & Roberts, 2017). In the educational context, which typically places a great deal of importance on achievement and is therefore relatively constant in its expectations (see, e.g., Weinstein, 1991), experiences of success or failure as described above are likely to repeatedly lead to similar responses, which consequently manifest themselves in more stable self-evaluations.

However, as was already noted by James (1892/1999), academic self-concept is considered a *subjective* and *relative* evaluation of one's own abilities. In other words, what students regard as a success or failure depends on a number of individual factors. For instance, two students who have the same ability in mathematics and who both score among the top 10% on a math test might still have different self-concepts in this domain. One place where people's *subjective* perceptions of their own abilities may originate is gender differences. Boys have been found to report higher self-concepts in STEM-related domains. However, these differences are only partly due to actual differences in achievement (see, e.g., Trautwein & Möller, 2016). Explanations for these findings typically evolve around gender stereotypes and distinct expectancies of boys and girls (e.g., communicated by teachers and parents; Friedrich et al., 2015; Frome & Eccles, 1998; Harter, 1998; Tiedemann, 2000), suggesting that people generally attribute more talent to boys in STEM subjects, whereas girls with the same performance are typically perceived as more diligent rather than talented, and this difference is reflected in higher self-concepts among boys. On a more general level, the importance of certain expectancies and attributions of success or failure has also been highlighted by intervention studies that are aimed at fostering individual self-concepts. O'Mara et al. (2006) conducted a large meta-analysis of 145 studies, including a total of 200 self-concept interventions. The authors found the largest effect sizes (d > 1.50) for self-concept interventions utilizing praise and (particularly attributional) feedback, emphasizing that self-evaluations are substantially shaped by expectancies and reinforcements communicated by significant others as well as people's own attributions for their own behavior or performance.

Most importantly—like everything humans perceive and evaluate—the evaluation of one's own abilities is not just subjective but *relative* and therefore dependent on the standard to which it is compared (see, e.g., Kahneman & Miller, 1986; Morina, 2021). In a typical classroom situation, students are presented with a wide range of pieces of information coming from different sources (e.g., classmates, the teacher, instructional materials), happening simultaneously, and changing at a high rate of immediacy, oftentimes unpredictably (Doyle, 2006). Numerous pieces of information in the classroom are potentially relevant to self-concept and are a source for respective comparisons. Consequently, there are a number of factors that determine how students evaluate their own abilities (Trautwein & Möller, 2016). In order to evaluate their ability in a specific domain, students compare it with the abilities of other students (i.e., social comparisons; Festinger, 1954) and with certain standards, such as the score needed to pass (i.e., criterial comparisons), with their own achievement at a previous timepoint, such as on the last test (i.e., temporal comparisons; Albert, 1977), and with their own performance in another subject (i.e., dimensional comparisons; Möller & Marsh, 2013).

Whereas dimensional and temporal comparisons refer to intraindividual comparisons, social comparisons are based on the idea that was already proposed by James (1892/1999): that the social environment plays a role in one's perceptions and evaluations of one's own abilities, and more specifically, that one's self-perceptions of ability in specific situations are made in relation to others. This idea is inherent in conceptualizations of self-concept from educational psychology (e.g., Shavelson et al., 1976) and has been particularly reinforced by conceptualizations of self-concept from social psychology (e.g., Markus & Wurf, 1987). On the basis of these theoretical conceptualizations, the present dissertation focuses on social comparisons in the classroom as a major determinant of students' academic self-concept, which will be described in Chapter 2.

1.2 Social Comparisons in the Classroom

In 1954, Festinger posited his influential social comparison theory, postulating that the self is always a social concept and that other people play a major role in shaping the beliefs we hold about ourselves (in line with earlier theoretical work by Cooley, 1902; Mead, 1934). Festinger (1954) described different reasons for social comparisons, noting that one or potentially the most central motive is self-evaluation.[4,5] Social comparisons have hence received plenty of attention from research in educational and social psychology as a major determinant of students' evaluations of their academic abilities (e.g., Blanton et al., 1999; Dumas et al., 2005; Huguet et al., 2001).

Social comparisons in classrooms are commonplace from preschool onwards (Dijkstra et al., 2008; Frey & Ruble, 1990; Suls & Mullen, 1982). When students compare themselves with their peers, they can make their comparisons with either higher achieving peers (upward) or with lower achieving counterparts (downward). Both types of comparisons can have either contrastive or assimilative consequences, ultimately determining whether students will evaluate their own abilities as better or worse (e.g., Marsh, Trautwein, et al., 2008; Mussweiler et al., 2004). *Contrast effects* in the context of social comparisons describe the phenomenon of perceiving one's own abilities as different from the comparison target, consequently leading to lower self-concept in the case of upward comparisons. *Assimilation effects*, in turn, mean that individuals perceive comparison targets to some degree as role models and evaluate their own abilities as similar, which results in higher self-concept in the case of upward comparisons.

In the following, I summarize existing research on social comparisons in the classroom. I first introduce the Big-Fish-Little-Pond effect, which primarily examines the consequences of social comparisons in the classroom and has received a great deal of attention from research in educational psychology (Chapter 1.2.1). I then move to a more process-oriented—primarily social psychologically grounded—perspective by providing an overview of approaches used to gain insights into the actual mechanisms of social comparisons (Chapter 1.2.2). Finally, on the basis of both perspectives, I outline central open questions regarding social comparison processes in the classroom (Chapter 1.2.3).

[4] Festinger (1954) theorized that humans refer to others to evaluate themselves when they are uncertain about their abilities. Research in the following decades has agreed that this is true but is an understatement as social comparisons are more ubiquitous for the purpose of self-evaluation than Festinger initially claimed: It is assumed that people have a fundamental tendency to evaluate their own abilities in relation to others and do so not only in times of uncertainty (see, e.g., Suls & Wheeler, 2000).

[5] Other related motives include self-assessment, self-enhancement, self-improvement, and self-verification (see, e.g., Sedikides & Strube, 1997).

1.2.1 The Big-Fish-Little-Pond Effect

The Big-Fish-Little-Pond effect (BFLPE; see original studies by Marsh & Parker, 1984; Marsh, 1987) suggests a negative effect of class-average achievement on students' academic self-concept, controlling for individual achievement (see Figure 4 for an illustration of the theoretical model). To illustrate this concept, let's take two students named Ella and Bob who have the same abilities in mathematics. Say that Ella's class is very good in mathematics, whereas Bob is in a class with a lower average performance level in mathematics. With their performance level, Ella belongs to the lower performing students in her class, whereas Bob is among the top students in his class. In BFLPE language, Bob is the big fish in the pond, and Bob's self-concept will be significantly higher compared with Ella's despite the fact that they have the same abilities.

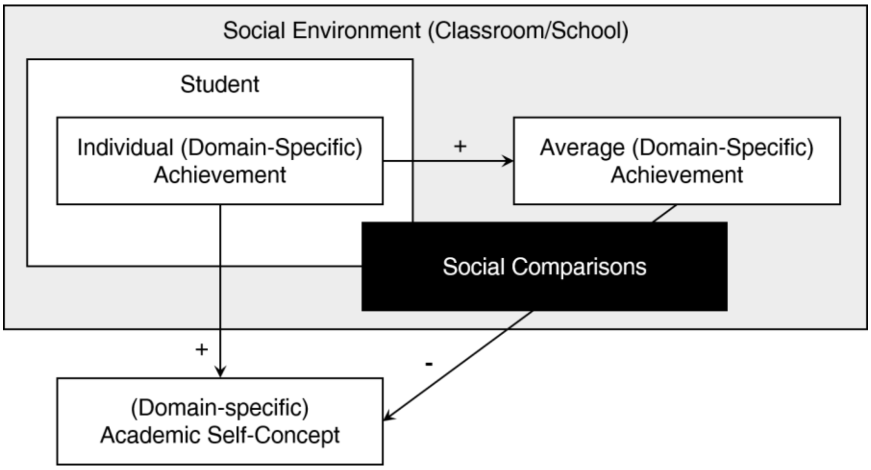


Figure 4. The Big-Fish-Little-Pond effect (BFLPE). Social comparisons are in the black box because they are assumed to be the underlying mechanism but have not been directly examined in classical BFLPE research.

Especially research in educational psychology has produced a large number of studies that have examined and provided evidence for the BFLPE in the last few decades (see the meta-analysis by Fang et al., 2018; and respective reviews by Marsh et al., 2017; Marsh & Seaton, 2015). For instance, the BFLPE has been found in samples from different countries and age groups (Cheng et al., 2014; Loyalka et al., 2018; Marsh & Hau, 2003; Seaton et al., 2009), with regard to school transitions (Becker & Neumann, 2016, 2018; Trautwein & Baeriswyl, 2007), or with regard to gifted education and subsequent (extracurricular) ability grouping (Dai et al., 2013; Herrmann et al., 2016). The generalizability of the BFLPE has been demonstrated in particular by a number of large-scale studies using data from PISA, including several hundred thousand students from countries all over the world (Marsh & Hau, 2003; Marsh, Parker, et al., 2019; Nagengast & Marsh, 2012; Seaton et al., 2009, 2010).

The BFPLE typically assumes that social comparisons and subsequent contrast effects are the underlying mechanism, which leads to the repeatedly found pattern of results (see, e.g., Huguet et al., 2009; Marsh et al., 2014). Notably, contrary to the BFLPE, studies have also provided evidence for assimilative and therefore positive effects of upward social comparisons on subsequent self-evaluations in the sense that learners perceive their own competence as higher when they are surrounded by higher achieving peers or are placed in high-achieving schools (Brewer & Weber, 1994; Cialdini & Richardson, 1980; Pelham & Wachsmuth, 1995). Theoretically speaking, there is a sound basis for the finding that students “bask in reflected glory” (Cialdini & Richardson, 1980, p. 406) of their higher achieving peers, but the empirical evidence for this has not always been consistent (Hattie, 2002; Nash, 2003; Televantou et al., 2015). Marsh and colleagues (Marsh, 1984; Marsh, 1987; Marsh, Seaton, et al., 2008) have argued that the BFLPE presents a net effect of the posited negative contrast effect and a counterbalancing positive effect of upward comparisons (i.e., assimilation or reflected glory effect). Positive effects of upward comparisons in the sense of reflected glory or assimilative peer spillover effects have been found to be comparably smaller and are typically outweighed by the contrastive effects, resulting in the overall negative BFLPE (e.g., Dicke et al., 2018; Marsh et al., 2000; Seaton et al., 2008; Trautwein et al., 2009).

Accompanying the growing body of studies providing evidence for the BFLPE, the question of potential moderators has also received considerable attention in the last few decades. With regard to the overall composition of the class or school, for instance, the degree of ability stratification has been found to influence the extent of the BFLPE (Lohbeck & Möller, 2017; Parker et al., 2018; Salchegger, 2016). With respect to classroom processes, studies have shown that teachers’ feedback and instructional practices (Roy et al., 2015; Schwabe et al., 2019), teachers’ grading practices (Trautwein et al., 2006), and comments from peers (Gest et al., 2008) influence the BFLPE. Finally, with regard to individual student characteristics, the BFLPE has been found to be affected by personality traits (specifically narcissism and neuroticism; Jonkmann et al., 2012), achievement goals (Wouters et al., 2013), gender (Plieninger & Dickhäuser, 2013; Preckel & Brüll, 2008; Thijs et al., 2010), and individual achievement (Huguet et al., 2009; Marsh et al., 2014; Trautwein et al., 2009). However, the results across different studies have been mixed, and consistent evidence of moderating variables that clearly reinforce or attenuate or at least neutralize the BFLPE is missing (Seaton et al., 2010; Seaton et al., 2011). Against this background, Marsh et al. (2021) argued that social comparisons are a universal evolutionary process and that the BFLPE is hence generalizable across diverse student and contextual characteristics (see also Marsh & Seaton, 2015).

With regard to the empirical support for the BFLPE, two points should be noted: First, BFLPE research is mostly based on cross-sectional data, relying on correlational analyses of students' self-reported self-concepts in relation to individual and class-average achievement scores. Second, the BFLPE assumes—without explicitly examining it—that social comparisons are the reason for the observed pattern of effects (see the black box in Figure 4). Concerning the first issue, there have been theoretical and methodological advances over time (see, e.g., reviews by Marsh & Seaton, 2015; Marsh, Seaton, et al., 2008), such as the use of multilevel models to differentiate between the effects that occur at the school, class, and student levels. Also, longitudinal designs have been particularly good for providing some more defensible support with regard to the BFLPE's causality (Marsh et al., 1991; Marsh et al., 2001; Marsh et al., 2000; Marsh, Pekrun, et al., 2019; Pekrun et al., 2019). However, to date and to the best of my knowledge, Zell and Alicke (2009a) conducted the only truly experimental study on the BFLPE by randomly assigning students to different comparison conditions and testing the resulting effects of different frames of reference on students' self-evaluations (see also Zell & Alicke, 2010).

Speaking to the second issue mentioned above, Dai and Rinn (2008) critiqued BFLPE research to this day for being too restrictive in its assumption of social comparisons as the BFLPE's underlying mechanism. The authors suggested a broader conception of social comparison effects on academic self-concept and highlighted the lack of direct evidence that social comparisons are the underlying mechanisms (Dai & Rinn, 2008):

> The most problematic aspect of the BFLPE paradigm is that social comparison is inferred, not observed or measured; and explanation and interpretation of data is based on blanket assumptions rather than direct evidence. (p. 297)

In fact, studies that have explicitly examined the role of social comparisons in the BFLPE are scarce. For instance, Marsh, Trautwein, et al. (2008) included the achievement of an individually selected classmate as a comparison target (in addition to the typically used class-average achievement) in their analyses and found distinct effects of both frames of reference in line with the BFLPE. Similarly, Huguet et al. (2009) included the achievement of individually selected classmates as comparison targets and additionally assessed students' perceived relative standing in the class, both of which significantly mediated the BFLPE. In order to demonstrate the role that direct social comparisons play in determining the BFLPE, Marsh et al. (2014) added individual rankings of students' ability (i.e., how students thought they compared with other students in their class) to their analyses of the BFLPE and found that it substantially reduced the BFLPE.

In sum, the BFLPE is one of the best-established findings in educational psychology, and social comparisons as its underlying mechanism are widely acknowledged. However, on the basis of existing research, it remains unclear how exactly these social comparisons in the classroom proceed. To this end, the following Chapter 1.2 describes findings from social psychology studies—particularly those based on experimental designs—that focus on the mechanisms of social comparisons and provide respective insights into the underlying processes of social comparison effects.

1.2.2 Mechanisms of Social Comparisons

Research from social psychology on the self has mostly been oriented toward understanding when and why people compare themselves and with whom (Buunk & Gibbons, 2007; Buunk & Mussweiler, 2001; Suls & Wheeler, 2000). The field has developed different paradigms, approaches, and applications to study social comparisons (see a review of the early decades by Wood, 1996). In order to understand individual comparison behaviors and attitudes in different situations, different sets of questionnaires—for instance, about individual social orientation (e.g., Gibbons & Buunk, 1999) or comparison choices in the classroom (Blanton et al., 1999; Dumas et al., 2005; Huguet et al., 2001)—as well as narrative methods, such as interviews or diary methods based on retrospective recall (see, e.g., Buunk et al., 2007), and in situ measurement, such as experience sampling in a naturalistic setting, have been employed (see the review by Arigo et al., 2019). A large portion of social psychology work on social comparisons has been based on experimental designs, which provide systematic insights into the processes that underlie social comparisons.

The large body of experimental studies on social comparisons can be broadly grouped into studies examining (a) the *selection* of social comparison targets and (b) *reactions* to social comparison information. More specifically, experimental research on social comparisons is characterized by five overall approaches (see Table 1 for an overview).[6] First, for quite a long time after Festinger's theory was published in 1954, the major focus was comparison level choice. These studies typically used the so-called rank order paradigm (Wheeler, 1966), in which participants must complete a test, are subsequently presented with an order of scores,

[6] The summary and classification provided here is not based on a systematic literature review but is rather aimed at a phenomenological description of the most important experimental approaches for the purpose of the present dissertation.

and are asked to select the score to which they want to compare their own score (e.g., Smith & Insko, 1987; Wheeler et al., 1969). The primary research interest behind the respective studies was to gain insights into (upward vs. downward) comparison choices and respective motives. Findings generally have suggested that people tend to compare themselves with slightly better others, which supports their inherent desire to improve (see, e.g., the review by Gruder, 1977). In a similar vein, a large body of studies have used fictitious scenarios with given comparison targets to investigate the selection of comparison targets. These studies typically outline a situation that participants have to imagine (e.g., that they just received a certain grade on an exam). Participants are then presented with detailed descriptions of potential comparison targets and are asked to indicate their desired comparison target (e.g., Helgeson & Mickelson, 1995; Mussweiler, 2001b; Ray et al., 2017). Similar to the comparison-level choice studies described above, these studies have been used to investigate social comparison choices but with a greater focus placed on individual characteristics as drivers of social comparison choices aside from (upward or downward) selection in terms of performance levels. Findings have indicated what Mussweiler (2003) called a selective accessibility mechanism, suggesting that people prefer to compare themselves with socially similar (or psychologically close) individuals with whom they identify.

Second, these aforementioned scenario studies have been intensively used to examine comparison effects with a given comparison target and manipulated comparison information. In these cases, participants are presented with (a description of) a comparison target that is typically experimentally varied across groups and are asked to evaluate themselves in relation to it (e.g., Kiviniemi et al., 2008; Mitchell & Schmidt, 2014; Mussweiler et al., 2004; Mussweiler & Strack, 2000; Ruble et al., 1980; Stanton et al., 1999; Tesser et al., 1988). The primary research interest in these studies has been not so much the choice of comparison because researchers explicitly took this away from the participant in the research design. Rather, the interest has been on insights into the antecedents of contrastive versus assimilative social comparison effects and the subsequent consequences for self-evaluations. The results of these studies have shown that people tend to use contextual information that seems close/similar to their own selves to evaluate themselves more favorably (e.g., Kessels & Hannover, 2004; Mussweiler, 2003; Wheeler et al., 1997).[7] With regard to social comparisons in the classroom, for instance, students' gender and race but also their achievement levels and general

[7] According to social categorization theory, belonging to one or more groups is an important part of identity, and therefore categorizations that follow the principle of similarity (in-group) and dissimilarity (out-group) are a fundamental aspect of human nature that guide social behaviors (e.g., Turner et al., 1987).

psychological closeness have been found to be important factors that are related to a perception of dis/similarity among students (see the review by Dijkstra et al., 2008).

Third, fictitious scenarios with a given comparison target have been combined with priming manipulations. Such studies have focused in particular on situational influences, such as students' vulnerability to social comparisons or perceived dis/similarities between the self and the comparison target. Consequently, in these designs, a certain mood toward social comparisons is typically induced in participants (e.g., making them recall their last failure or success; Aspinwall & Taylor, 1993; Gibbons, 1986), or participants are instructed to focus on either similarities or dissimilarities with their fictitious comparison target in the scenario (e.g., Collins, 1996; Mussweiler, 2001b). Priming tasks with a focus on dis/similarities have been found to be very effective with regard to self-concept(-related) outcomes (see the reviews by Cross et al., 2011; Oyserman & Lee, 2008), and the overall findings have suggested that comparisons with targets that are perceived as more similar tend to produce less adverse comparative effects than comparisons with dissimilar people (in line with the so-called similarity hypothesis proposed by Festinger, 1954; see the selective accessibility mechanism proposed by Mussweiler, 2003).

Fourth, one of the prototypical research designs that is used to examine the effects of social comparisons has been based on manipulated performance feedback, whereby researchers ask participants to complete a task and subsequently confront them with manipulated information about their performance (e.g., that they scored among the top 10% of all participants) and assess self-evaluations or achievement in a follow-up task (e.g., Bannister, 1986; Klein, 1997, 2003; Lyubomirsky & Ross, 1997; Pyszczynski et al., 1985; Rijsman, 1974; Zell & Alicke, 2009b). The respective studies have examined different outcomes in addition to self-evaluations and achievement (e.g., achievement goals and emotions; Ilies & Judge, 2005; O'Connell, 1980). Notably, the respective findings have shown mixed results, suggesting positive as well as—in line with the BFLPE—negative effects of upward comparisons.

Table 1. Overview of Central Experimental Approaches That are Applied to Examine Social Comparisons

Approach	Type of comparison	Type of comparison target	Type of comparison information	Main research interest
Comparison (level) choice	Forced	Specific	Performance levels (rank order) or individual characteristics and performance (mostly scenario-based)	Comparison selection: Comparison motives and (upward vs. downward) choices
Manipulated comparison information	Forced	Specific	Individual characteristics and performance (mostly scenario-based)	Comparison reaction: Antecedents of assimilation versus contrast effects, effects on self-evaluations
(Similarity) priming	Forced	Specific	Individual characteristics and performance (mostly scenario-based)	Comparison reaction: Antecedents of assimilation versus contrast effects
Manipulated performance feedback	Forced	Generalized	Average performance level (mostly related to task-specific performance)	Comparison reaction: Effects on self-evaluations
Physiological response	Forced	Specific	Individual characteristics and performance (mostly based on specific stimuli)	Comparison reaction: Behavioral indicators of comparison information processing

Note. The type of "comparison target" refers to what is interchangeably labeled a "comparison standard" or "frame of reference" in different studies. "Scenario-based" comparison information refers to the commonly used approach of scenario studies that outline fictitious situations and comparison targets that participants have to imagine to examine the mechanisms behind their selection of and reactions to social comparisons.

Finally, newer developments include physiological measures to gain insights into people's responses to social comparison information across different domains. The respective studies typically present different stimuli to participants and aim to infer participants' processing of and response to social comparison information from reaction times and error rates to certain self-related statements (e.g., Muller & Butera, 2007; Rullo et al., 2015), from eye movements and visual attention patterns when examining self-relevant information (Bauer, Schneider, Waldorf, Adolph, et al., 2017; Bauer, Schneider, Waldorf, Braks, et al., 2017; Lou et al., 2019; Michinov et al., 2015), or from different measures of brain and heart activity[8] and changes in skin conductance (Burnside & Ullsperger, 2020; Kedia et al., 2014; Scheepers, 2009; Scheepers & Ellemers, 2005). The primary research aim of these studies has been to provide evidence for existing findings based on behavioral and physiological data, considering that most social psychology research on social comparisons has relied on self-reports. The majority of these studies have examined social comparisons in non-education-related contexts, for instance, regarding body satisfaction or social status groups. With regard to social comparisons in an educational context, for instance, Michinov et al. (2015) used eye-tracking data from a web-based synchronous brainstorming task to examine students' visual attention to their partner's ideas and the subsequent effects on their self-evaluations and their own achievement.

In sum, social psychology research has provided a wide array of insights into when, why, and with whom people compare themselves to evaluate themselves. However, most of this research has been lab-based, and it is unclear how it generalizes to a—way more complex and dynamic—classroom setting. With regard to social comparisons in the classroom, there are two important distinctions that need to be considered: First, experimental approaches to examine social comparisons always examine forced comparisons (see the overview in Table 1)[9]. This is not surprising based on the fact that social comparisons are the focus of experimental manipulations and are therefore explicitly targeted. However, in a typical classroom situation, students are unlikely to be explicitly asked to compare themselves with their classmates, but this presumably happens naturally in most cases. Second, experimental studies of social comparisons mostly work with comparisons with specific others (e.g., single comparison choices, single manipulated comparison targets; see the overview in Table 1). Whereas students in the classroom presumably compare themselves with specific classmates as well, the question

[8] The cited studies used electroencephalograms (EEG), functional magnetic resonance imaging (fMRI), impedance-cardiographic signals (ICG), electrocardiographic signals (EKG), and blood pressure, respectively.

[9] "Forced" social comparisons are also referred to as "(situationally) imposed" social comparisons compared with so-called "naturally occurring" (also labeled "self-engendered" or "spontaneous") social comparisons (see, e.g., Dai & Rinn, 2008).

that remains is how much their self-evaluations are influenced by the impressions they get from the class' overall dynamics (i.e., more generalized others) rather than comparisons with single classmates. This distinction is crucial, considering that BFLPE research, which assumes that social comparisons are the mechanism that underlies the observed effects, almost exclusively examines social comparisons with a generalized other (i.e., the class-average performance level; Marsh, Trautwein, et al., 2008).

1.2.3 The Black Box of Social Information Processing in the Classroom

As outlined in the previous chapters, there is compelling evidence for effects of the social environment at school on students' self-evaluations (see BFLPE; Chapter 1.2.1), and a large body of experimental research has provided insights into the mechanisms behind social comparisons in general (Chapter 1.2.2). Both BFLPE research and social psychology research on social comparisons agree with Nowak et al. (2000), who summarized that

> The thoughts and feelings that populate the self-system […] are frequently derived from social experiences, revolving to a considerable degree around real and imagined relationships with specific and generalized others. (p. 39)

However, with regard to the classroom as the social environment (and academic self-concept as the primary "thoughts and feelings populating the self-system"), the exact processes mentioned in this statement are still quite a black box. In general, a black box describes a complex entity for which we know the inputs and the—presumably causally linked—outputs, but not the inner workings (Bunge, 1963). With regard to social comparisons in the classroom, these "unknown inner workings" primarily concern the question of how exactly self-evaluations are derived from social experiences in the classroom, in other words, how exactly students process social information provided by specific and generalized others in the classroom.

It has long been argued that it is crucial to understand how people make selective use of social comparison information given the vast amount of social comparison information that they typically face every day (e.g., Goethals, 1986; Suls, 1986; Suls & Wheeler, 2000). Social psychology research has provided insights into the respective mechanisms behind social comparisons (see Chapter 1.2.2). However, as such studies have primarily been based in laboratory settings, such studies cannot—and do not aim to as a matter of fact—reflect the complexity and different sorts of dynamics in real classrooms. On the basis of existing evidence, I argue that the black box of social information processing in the classroom can be roughly described by two distinct but related types of processes, which concern students' *cognitive* and *behavioral* responses to social comparison information.

First, with regard to students' *cognitive* responses to social comparison information, the central question is what students interpret as social comparison information and how. BFLPE research typically regards the reference group as a generalized other in the sense of an average performance level. In comparison, existing social comparison choice studies in real-world classrooms (Blanton et al., 1999; Dumas et al., 2005; Huguet et al., 2001) have asked students to nominate one or two classmates with whom they mostly compare themselves in class. Whereas these studies have undoubtedly been able to present a more authentic choice of social comparison targets compared with other lab-based experimental designs, the question that remains is to what extent comparisons with the classmates who were nominated as comparison choices in these studies are "critical" for students' self-evaluations compared with social comparison information from more generalized others that is, for instance, implicitly included in the class' overall level of performance or students' displayed performance-related behavior.

Second, speaking to students' *behavioral* responses to social comparison information, existing experiments that are aimed at examining social comparisons among students have primarily relied on forced comparisons with a given comparison target. With regard to the classroom context, it thus remains unclear how social comparisons naturally proceed and influence students' self-concept: Do students actively compare themselves, or are they rather compared with others by the teacher and their peer learners? The respective insights into students' active behavioral responses to social comparison information are particularly important against the background that the focus of social comparisons (i.e., comparing oneself against a standard versus a standard being compared against oneself) has been found to impact whether comparisons have more contrastive or assimilative consequences (Mussweiler, 2001a). In other words, social comparisons that are explicitly presented to a student might be based on different mechanisms and might consequently lead to different outcomes compared with social comparisons that naturally occur in a classroom situation.

In order to gain insights into the black box and the respective "inner workings" of social comparisons in the classroom, with the present dissertation, I propose a theoretical model for examining social comparisons in the classroom that (a) describes the mechanisms that underlie social information processing in the classroom and consequently (b) provides a framework to guide research that is aimed at relating the social comparison processes in the classroom to characteristics of the social environment as well as to students' self-evaluations (see Figure 5). The model integrates central perspectives on social comparisons from social and educational psychology. More specifically, it uses a process-oriented definition of self-concept (Markus &

Wurf, 1987; see Figure 3) to describe social comparisons as the mechanisms that underlie the BFLPE (see social comparisons as the black box in Figure 4).

Reflecting the conceptualization of self-concept by Markus and Wurf (1987), the black box of social comparisons is described by the two aforementioned related but distinct mechanisms behind the processing of social comparison information, namely, (a) *intra*personal processes on the side of the student and (b) *inter*personal processes with the social environment (i.e., classmates). *Intra*personal behavior refers to all *covert* behavior, which cannot be directly observed and reflects students' cognitive responses to social comparison information, such as their interpretation of and affective responses to their classmates' (performance-related) behavior. *Inter*personal behavior refers to students' behavioral responses to social comparison information and therefore to *overt* behavior that can be observed, such as visual attention to peer learners or actual interactions with them.[10]

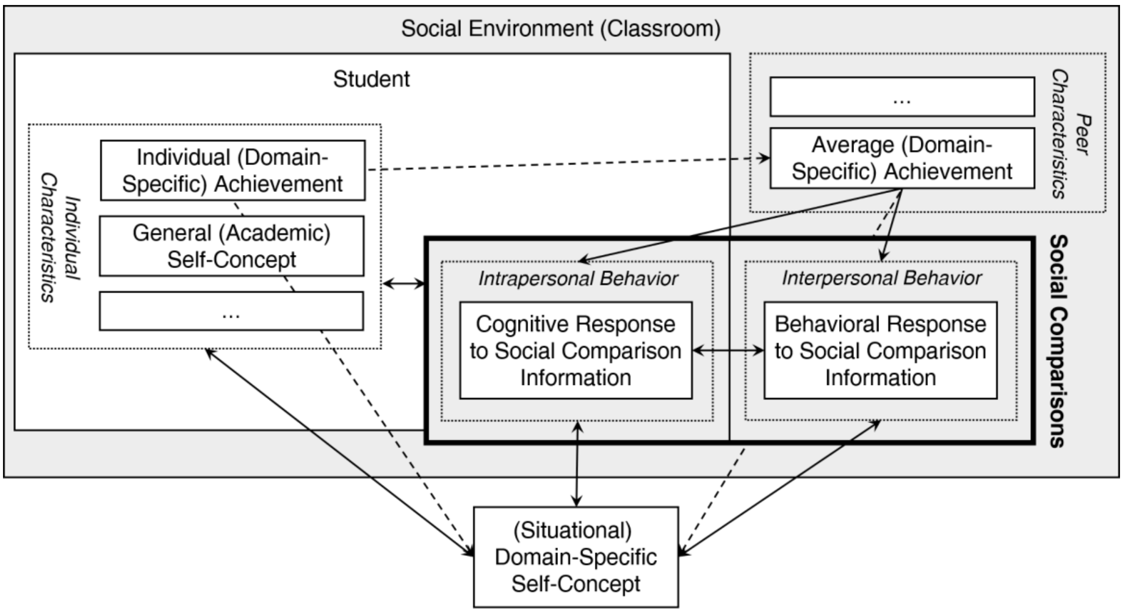


Figure 5. Theoretical model for examining social comparisons in the classroom. Dashed arrows represent the relationships that are typically examined in BFLPE research.

[10] Markus and Wurf (1987) distinguished between intra- and interpersonal behavior primarily on the basis of the fact that some parts of information processing are purely related to the self (i.e., the processing of self-relevant information as intrapersonal behavior), and other parts of information processing are related to the information provided by the social environment in general (i.e., the processing of the social environment as interpersonal behavior). Based on this argument, both types of behavior are to some extent covert processes and cannot be disentangled. Whereas I also highlight the link between intra- and interpersonal behaviors, I aim to make a clearer distinction between the two and consequently focus only on the behaviors that are clearly either intra- or interpersonal in the sense of covert and overt behaviors.

In cognitive psychology, these processes that I suggest describe the black box of social comparison processes in the classroom are referred to—even if not always defined and differentiated so clearly—as (behavioral indicators of) social cognition (e.g., Fiske, 1993; Forgas, 1983; Higgins & Bargh, 1987; Sherman et al., 1989; Smith & Semin, 2007). On the basis of the respective conceptualizations from cognitive psychology (see the review by Cross et al., 2011), it is crucial to consider both (a) students' cognitive (i.e., intrapersonal) responses to and their respective interpretations of the social information that is provided as well as (b) their observable (i.e., interpersonal) behavioral responses to it. The relationship between these two processes is particularly interesting in the context of social comparisons in the classroom against the background that when students aim to compare their own abilities during instruction, they aim to assess something that they cannot necessarily observe directly but rather need to infer from certain (performance-related) behaviors of their peers (see similar arguments by De Jaegher et al., 2010; Goethals & Darley, 1977; Suls & Mullen, 1982).

In line with the BFLPE and its underlying assumptions, the model suggests that social comparison information originates primarily from peer learners as it is mainly characterized by peers' average achievement and respective performance-related behavior. Based on the fact that self-concept is "systematically implicated in all aspects of social information processing" (Markus & Wurf, 1987, p. 301), the proposed theoretical model conceptualizes situational self-concept as a construct that is related to both intra- and interpersonal social comparison behavior in response to the social comparison information provided by peers. In line with Markus and Wurf (1987), situational—in their wording: dynamic—self-concept is conceptualized as distinct from but also closely related to more stable self-beliefs (e.g., general academic self-concept) and individual characteristics.

Importantly, in order to gain insights into the black box of social comparisons and the subsequent processing of social information in a classroom as proposed here, one needs to work in an authentic environment (as opposed to the lab-based settings typically used in experimental research on social comparisons) and at the same time be able to control the comparison information that is provided (as opposed to the plethora of situational influences that field-based research typically cannot disentangle). To this end, the following chapters introduce immersive virtual reality as a tool for experimental classroom research.

1.3 Immersive Virtual Reality as a Research Tool

> It's a very interesting kind of reality. […] The thing is, however real the physical world is – which we never can really know – the virtual world is exactly as real, and achieves the same status. But at the same time, it also has this infinity of possibility that you don't have in the physical world: in the physical world, you can't suddenly turn this building into a tulip; it's just impossible. But in the virtual world you can. (Lanier, 1989, p. 8)

Considering that virtual worlds have evolved tremendously within the past 30 years based on fast-paced technical advances in the field of software and hardware development, what Lanier noted back in 1989 is now true more than ever: Virtual realities, that is, computer-generated simulated environments, allow for realistic perceptions and seemingly real interactions within an artificial virtual world, and the designers and programmers of these worlds can make almost anything a reality (e.g., Blascovich et al., 2002).

These two aspects—authenticity on the one hand and controllability on the other—are what has made virtual realities so popular in many different domains (Bellini et al., 2016; Slater & Sanchez-Vives, 2016). In the educational field, virtual realities have received a great deal of interest as an instructional tool. The respective learning applications typically make use of virtual realities to provide learners with experiences that they cannot or not so easily have in real life (Howard et al., 2021; Johnston et al., 2017; Karutz & Bailenson, 2015; Seidel & Chatelier, 2013). For instance, virtual reality applications are often used to simulate complex or risky procedures in engineering, military, or medical training (Alhalabi, 2016; Bric et al., 2016; Collaco et al., 2021; Moro et al., 2017; Webster, 2015) or to allow students to go on virtual field trips to far-away places or to do advanced science experiments in a safe environment without leaving their usual classrooms (Cheng & Tsai, 2019; de Jong et al., 2013; Fauville et al., 2020; Makransky et al., 2020; Makransky, Terkildsen, et al., 2019; Queiroz et al., 2018).

For educational psychology research, the integration of authenticity and experimental control moreover presents the unprecedented opportunity to implement research designs that allow researchers to revisit phenomena that have been subject to studies for a long time. However, the respective studies have either suffered from a lack of experimental control in field settings or missed the authenticity of a real-world setting when placed in the lab (Blascovich et al., 2002; Fox et al., 2009). In this vein, the present dissertation uses immersive virtual reality as an experimental tool to overcome the so-called "experimental control-mundane realism trade-off" (Blascovich et al., 2002, p. 103) with regard to research on social comparisons in classrooms. On the basis of the theoretical model that I am proposing for examining social comparisons in the classroom (see Figure 5, Chapter 1.2.3), I will use the following chapters to outline three critical aspects with regard to the use of immersive virtual reality as a research tool: first, the transformation and authentic simulation of real-life classrooms as social environments in an immersive virtual reality setting (Chapter 1.3.1); second, the potential of the experimental control and manipulation of the immersive virtual reality classroom setting, especially the social comparison information provided by peers (Chapter 1.3.2); and finally, the affordances of process data from the immersive virtual reality environment to gain insights into social comparison processes in the classroom (Chapter 1.3.3). [11]

1.3.1 Authenticity: Immersive Virtual Simulations of Classroom Realities

Ella and Bob are sitting in their classroom. They are surrounded by their 23 classmates, and the class is waiting for the teacher to enter. This school year, they are in a new class that will be taught by a new mathematics teacher, and they are curious what it will be like. The teacher enters the classroom, and as the class realizes that the lesson is about to start, conversations slowly fade and the class gets quiet. "So, let's dive right in," the teacher announces, "you can see today's topic written here on the blackboard." Ella freezes. She doesn't understand a word of what is written on the blackboard, but it definitely sounds very complicated to her. She looks at Bob, sitting next to her and seeming way more comfortable with the situation. "Who has an idea what this means?" the teacher asks, pointing to what's written on the blackboard. Bob immediately raises his hand. "Oh no, I should probably also know this, shouldn't I?" Ella gets worried and turns around to see her other classmates' reactions...

When aiming to use virtual reality as a research tool, the central and most important goal is to transform classroom situations as outlined above into authentic virtual reality experiences that make students react as they would in the real world (see, e.g., Williams, 2010). What I summarize here as *authentic* (i.e., ecologically valid) experiences in virtual reality environments can be described on a number of different dimensions. However, no unified definition for any of them can be found in the existing literature to date (Heeter, 1992; International Society for Presence Research, 2000; Lee, 2004; Lombard & Ditton, 1997; Lombard et al., 2009; Schubert et al., 2001; Slater, 1999; Witmer & Singer, 1998). I therefore outline three central constructs that have repeatedly appeared in respective discussions and sum

[11] I focus on critical features regarding the implementation of IVR as a tool for experimental research from a conceptual perspective. The necessary technical background and respective details regarding the programming of the virtual reality classroom environment or the extraction of process data are provided in the Method sections of the empirical studies described in Chapters 3 to 5.

up the definition of what I refer to as *authenticity* in virtual reality experiences in the present dissertation: immersion, presence, and realism.

Immersion primarily refers to the technical features and the multimodality of virtual reality with respect to the extent to which the virtual environment places users with all their senses in an extensively surrounding and vivid illusion of reality (Sanchez-Vives & Slater, 2005; Slater & Wilbur, 1997). In order to achieve high levels of immersion, state-of-the art virtual reality set-ups use so-called head-mounted displays (HMDs), which enclose the entire eye area and are designed in such a way that the actual screens that present the virtual reality cannot be perceived as such (as well as any other visual stimuli from the physical world). Moreover, modern HMDs are equipped with noise-canceling headphones to additionally shut out audio stimuli from the real world as much as possible (see, e.g., Fox et al., 2009; Radianti et al., 2020). In this vein, immersion has also been defined as a sort of attentional involvement in the virtual reality environment, based on the complete shutting out of stimuli from the physical environment and the respective attentional focus on the virtual reality environment (International Society for Presence Research, 2000; Wirth et al., 2007; Witmer & Singer, 1998). As Wirth et al. (2007) pointed out, immersion and the respective attentional involvement in the virtual reality environment are the prerequisites for a virtual reality experience that will lead to authentic behavioral responses from users. Against this background, I explicitly highlight the immersive character of virtual reality classrooms as experimental research tools and refer to them as immersive virtual reality (IVR) classroom environments throughout this dissertation.[12]

With regard to students' actual experience in the IVR classroom, high-quality IVR environments are expected to give users an exhaustive sense of *presence*, which describes the psychological state and subjective perception of being in the virtual environment (Oh et al., 2018; Schubert et al., 2001; Slater & Wilbur, 1997). As Slater and Wilbur (1997) put it, a high experience of presence makes users describe IVR environments as "places visited rather than as images seen" (p. 3). There are many different definitions and assessments of presence (Cummings & Bailenson, 2016; Lombard et al., 2009; Schuemie et al., 2001), but they usually share the notion of presence referring to (a) a spatial perception of actually *being in* the virtual

[12] The term virtual reality (VR; without the added emphasis on immersion) is often used interchangeably to describe *immersive* virtual reality technology with head-mounted displays and noise-canceling headphones. The present dissertation highlights the immersive character because, per definition (see, e.g., Sanchez-Vivez & Slater, 2005), *nonimmersive* VR environments typically describe simulations without an all-encompassing sensory input (e.g., a computer game presented on a screen). Importantly, nonimmersive VR environments can be experienced from a first-person perspective and lead to great attentional involvement as well (e.g., being completely absorbed in a computer game), but technologically enabled full immersion leads to a more authentic experience in the simulated VR environment.

environment (Lombard et al., 2009; Schubert et al., 2001; Slater & Wilbur, 1997; Steuer, 1992) and (b) a social perception of *being with* another in the virtual environment and a respective response to or interaction with virtual actors (see, e.g., the review by Oh et al., 2018). Spatial presence is also described as "place illusion" (Slater, 2009, p. 3549), which—beyond making users feel like they exist in the virtual environment (Slater, 1999; Slater & Steed, 2000; Wirth et al., 2007)—describes the perception of oneself as an active being in the virtual world, that is, that one is able to move and act as one does in the physical world (Schubert et al., 2001; Zahorik & Jenison, 1998), particularly when the IVR environment is experienced from a first-person perspective (Lim & Reeves, 2009). Social presence refers to perceiving that one is in an IVR environment together with other people who appear in the IVR environment (i.e., avatars) and are controlled by real people or a computer (i.e., virtual agents). Whereas some argue that social presence necessarily involves (verbal or nonverbal) communication with other people in the IVR environment (Heeter, 1992; Schilbach et al., 2006), it is mostly suggested that social presence describes the pure sense of not being alone in the IVR environment but rather together with another entity (Biocca et al., 2003; Bulu, 2012; Lee, 2004; Lombard et al., 2009; Nass & Moon, 2000).

Notably, these spatial and social perceptions of IVR are affected by users' comparisons with the real world and presumably undergo a certain reality judgment (Lombard et al., 2009; Schubert et al., 2001). *Realism*, as the third feature characterizing the authenticity of an IVR experience, refers to the "realness" (Schubert et al., 2001, p. 271) of the IVR environment in terms of (a) how similar the events in IVR are compared with possible events in the physical world and (b) to what extent the objects and people in IVR resemble those in the physical world (International Society for Presence Research, 2000; Schubert et al., 2001). Importantly, it is not necessary for both aforementioned aspects to be fulfilled in order to lead to a high degree of realism and subsequently to a successful illusion of plausibility (Slater, 2009). For instance, the events in a science-fiction IVR may seem very unrealistic, but the realistic representation and behavior of the characters may nevertheless generate a plausible IVR experience (e.g., Lombard et al., 2009). In a similar vein, research has demonstrated that a more realistic (i.e., human-like) visual representation of avatars does not always lead to more favorable evaluations by users (MacDorman et al., 2009; Mathur & Reichling, 2016; Strait et al., 2015). This phenomenon is commonly referred to as the Uncanny-Valley effect (Mori et al., 2012), suggesting that when avatars are very realistic but not *completely* realistic, they tend to produce adverse perceptions and reactions among users as the little flaws in these nearly perfectly human virtual counterparts evoke a creepy feeling (i.e., often compared to the perception of zombies). Importantly, beyond their appearance, the sounds and behaviors of virtual agents play an important role (Bailenson et al., 2005; Blascovich, 2002; Heidicker et al., 2017; Ho & MacDorman, 2010; MacDorman, 2006). With regard to the educational use of IVRs, for instance, virtual agents have been implemented as pedagogical agents, and beneficial effects on students' learning and their perceptions of the virtual agents have also been found for nonhuman or less human-like visual representations (Hudson & Hurter, 2016; Makransky, Wismer, et al., 2019).

In sum, by stimulating as many sensory modalities as possible and by simultaneously suppressing stimuli from the physical environment (e.g., Sanchez-Vives & Slater, 2005), state-of-the-art IVR environments promote high levels of (spatial and social) presence and a well-functioning plausibility illusion. Against the background that there is no evolutionary difference in the processing of or response to stimuli on the basis of their (natural or virtual) origin (Horvath & Lombard, 2010; International Society for Presence Research, 2000), it is assumed that—given an authentic IVR experience—humans will experience and naturally respond to social influences in an IVR environment without "actual physical presence" (Blascovich, 2002, p. 127). In fact, there is a substantial body of empirical evidence showing that users' behavior in IVRs is similar to real-life behavior (see Parsons, 2015, for a review on the ecological validity of IVR environments). With regard to social influences in an IVR environment, research has provided evidence that general human behaviors as observed in a real-life setting could be reproduced in an IVR environment, such as individual differences in intimacy and interpersonal distance (Bailenson et al., 2001), social facilitation and inhibition effects (Hoyt et al., 2003; Zanbaka et al., 2006), mimicking behavior and empathy (Bailenson & Yee, 2005; Cummings et al., 2021), obedience to authority and the elicitation of distress (replicating the classic Milgram, 1963, experiment; Slater et al., 2006), and the evoking of specific emotions among participants (Diemer et al., 2015; Hirt et al., 2020; Huang & Bailenson, 2019).

Concerning the educational context, most of the research comparing IVR with real-world learning scenarios has been interested in students' learning with IVR compared with traditional (media) formats (see meta-analyses and reviews by Howard et al., 2021; Merchant et al., 2014; Mikropoulos & Natsis, 2011). Studies that have explicitly compared students' behavior—beyond their learning with different materials—in IVR and real-life classrooms are scarce and limited to students' reactions to distractions in an IVR classroom setting. There is evidence that students' individually different reactions to distractions in real-world classrooms are similar in an IVR classroom, and the respective differences (associated with ADHD diagnoses) can be reproduced in an IVR classroom simulation (e.g., Adams et al., 2009; Bioulac et al., 2012; Nolin et al., 2016; Pollak et al., 2010; Rizzo et al., 2006). In general, speaking to the use of IVR classrooms as an authentic research environment for students, it is important to note that children have been found to be particularly (both cognitively and behaviorally) responsive to IVR environments compared with adults: Children tend to perceive the simulations as more real and feel a higher level of presence, which makes them act more spontaneously while thinking less about the world outside of the IVR environment (Bailey & Bailenson, 2017; Baumgartner et al., 2008; Hite et al., 2019; Sharar et al., 2007; Southgate et al., 2017; Stavropoulos et al., 2017).

1.3.2 Experimental Control: Manipulating Immersive Virtual Classroom Realities

As outlined in the previous chapter, state-of-the-art IVR technology makes it possible to create authentic IVR classroom scenarios for students in which they respond naturally to different situational stimuli. Going back to the exemplary classroom situation outlined in the beginning of Chapter 1.3.1 (*Ella gets worried and turns around to see her other classmates' reactions...*), it can therefore be assumed that Ella would show a similar reaction and refer to her virtual peer learners if she experienced the situation as an IVR classroom simulation. This is a necessary precondition for IVR classrooms as an experimental tool. In addition, IVR classrooms provide full experimental control for researchers. Most importantly with regard to the use of IVR classrooms to investigate social comparison effects in the classroom, the social environment and peer learners' behavior can be manipulated. This means that researchers can systematically vary how the scenario outlined above unfolds. Imagine, for instance, option (a) *Most of Ella's classmates smile and eagerly raise their hands to answer the teacher's question* compared with option (b) *Most of Ella's classmates stare at their desks or cautiously look around*. Against the background that all other factors (e.g., the teacher's feedback) could be held exactly the same in the outlined scenarios, this presents an unprecedented opportunity to examine the effects of single classroom features in an experimentally controlled and yet authentic manner. In combination with the random assignment of students to one of the IVR classroom scenarios and an appropriate assessment of individual differences in baseline and outcome measures, IVR as a research tool presents a strong research design to draw causal conclusions on the basis of authentic classroom situations (see, e.g., perspectives on causal inference by West & Thoemmes, 2010).

Despite this obviously high potential of IVRs to integrate ecological and internal validity into a research design, it must be emphasized that, to date, only a limited number of studies have used IVR environments in experimental research in social or educational psychology (Parsons, 2015; Schnotz, 2016). Particularly with regard to social comparisons—or even more broadly speaking, compositional or contextual effects—in the classroom, there are only a few studies that have provided a starting point for designing a respective experimental design. In the following, I summarize the methodological approach of three types of studies situated in IVR classroom settings that I argue provide important groundwork with regard to the use of IVR classrooms to gain insights into the black box of social comparisons.

The first studies that explicitly used an IVR classroom as an experimental tool were conducted by clinical psychology researchers who used IVR classrooms to (a) assess effects of attention deficit hyperactivity disorder (ADHD) on student learning in a standardized classroom situation and subsequently aimed to (b) design IVR classrooms to support the rehabilitation and training of students suffering from ADHD in simulated classroom environments (Adams et al., 2009; Bioulac et al., 2012; Gutiérrez-Maldonado et al., 2009; Mangalmurti et al., 2020; Moreau et al., 2006; Nolin et al., 2016; Parsons et al., 2007; Pollak et al., 2010; Pollak et al., 2009; Rizzo et al., 2006; Rizzo et al., 2000). Most of these studies go back to Rizzo et al. (2000), who designed an IVR classroom scenario in which common distracting events (e.g., classroom noise, activities outside the window) were systematically varied. Students experienced one of the IVR classroom scenarios and had to complete a number of attention-related tasks that were presented on the blackboard or by the virtual teacher in the IVR classroom. Respective studies found that more distractions in the IVR classroom, including distracting peer behavior, led to significantly lower performances for students with ADHD compared with students who were not suspected to have ADHD, suggesting that the IVR environment is an ecologically valid measure of respective individual differences.

Second, aside from the use of IVR classrooms in the context of ADHD assessment and treatment, studies have transformed classrooms into an IVR environment to examine the effects of different classroom features more generally, such as seating arrangements and social dynamics. The first and probably most prominent study in this regard consists of three experiments conducted by Bailenson et al. (2008), who varied students' positions in the IVR classroom (i.e., more in the center vs. more peripheral and closer to the teacher vs. farther away) and the behavior of virtual peer learners (i.e., model students vs. more distracting students) to systematically examine the respective effects on students' learning outcomes. In a similar vein, Blume et al. (2019) examined the effects of the position of a seat that was proximal to the teacher compared with in the back of the classroom with randomly appearing distracting events and distracting behavior by classmates. Results of both studies indicate that peer learners in an IVR classroom have a substantial effect on student learning, either indirectly by showing better learning outcomes in the front row (Blume et al., 2019) or explicitly by finding better learning outcomes for students who were surrounded by virtual model students (i.e., less distracting peer learners; Bailenson et al., 2008).

Finally, with regard to social comparisons, there are two IVR studies that have provided some insights into effects of more or less high-performing virtual agents. Christy and Fox (2014) implemented leaderboards in an IVR classroom, showing the current task scores of every student in a virtual class so that participating students could see how they were performing relative to their virtual classmates. The authors were specifically interested in gender effects and manipulated the ratio of high-achieving boys versus girls on the leaderboard while keeping all other factors constant. The findings provided evidence of contrastive social comparisons with the same gender and subsequent negative effects on performance and self-evaluations. Moreover, whereas not situated in an IVR classroom environment, there is an IVR study from the field of economics that examined social comparison effects with a more or less powerful co-worker on participants' productivity (Bönsch et al., 2017; Gürerk et al., 2019). In this study, participants had to do a sorting task at a virtual conveyor belt with a virtual agent doing the same task in their field of view. Findings showed that the presence of a more powerful virtual counterpart led to a significant increase in participants' productivity.

Taken together, existing studies provide evidence that peer effects can be experimentally manipulated in an IVR classroom setting. Particularly the two last studies mentioned above (Christy & Fox, 2014; Gürerk et al., 2019) provide experimental evidence for social comparison effects in an IVR setting. However, their methodological approaches were similar to existing experimental approaches in real-world laboratories that have presented individuals with manipulated social comparison information (see Chapter 1.2.2). In other words, the IVR environments undoubtedly provided a more authentic research environment compared with scenario studies or similar approaches. However, it needs to be noted that the achievement level of potential comparison targets was explicitly presented in the IVR environments via the leaderboard (Christy & Fox, 2014) or on a screen with individual productivity scores for the virtual agent and the participant (Bönsch et al., 2017; Gürerk et al., 2019). Whereas participants in the IVR studies were not explicitly instructed to compare themselves (i.e., no forced comparisons), the experimental approach still somewhat encouraged social comparisons, and there was no need for participants to infer more implicitly provided social comparison information from their social counterparts in the IVR setting. To this end and to the best of my knowledge, there are no studies that have used an IVR classroom to examine naturally occurring social comparisons and the subsequent effects on self-evaluations in a classroom setting in which students, for instance, had to infer social comparison information from more implicit performance-related peer behavior (compared with explicitly displayed performance scores).

Figure 6. Immersive virtual reality classroom with different avatar representation styles. The top image shows a more cartoon-styled representation of avatars, the bottom a more realistic one.

Generally, the goal for the use of IVR classrooms as an experimental tool would be (a) to design IVR classroom scenarios that are as authentic as possible (see Chapter 1.3.1) and (b) at the same time only implement and manipulate cues in the IVR environment that are important to the current research question. With regard to the latter, it needs to be considered that there are countless aspects in an IVR classroom that can be manipulated or at least have to be decided on by researchers. For instance, Cheryan et al. (2011) highlighted the importance of an unbiased IVR classroom design that included only neutral objects (e.g., water bottles, lamps, plants, subject-unspecific pictures or posters on the wall) to avoid potential effects of domain-specific stereotypes (e.g., induced by subject-specific posters, pictures of famous male/female representatives of the respective domain). However, even when aiming for an overall relatively simple and yet authentic IVR classroom design (see Figure 6 for the one used in the present dissertation), there are a number of decisions that researchers need to make about generic factors (e.g., the seating arrangement) as well as more specific design questions (e.g., the style used to represent virtual avatars).

The critical point to consider for all configuration or manipulation decisions in an IVR classroom is how they affect the processing of the information that is provided in the IVR environment as this determines not just users' perceptions but ultimately the successful use of the IVR classroom for the specific research purposes. The following chapter outlines how IVRs provide insights into the respective processing of information in the IVR environment.

1.3.3 Opening the Black Box: IVR Classrooms as Sources of Standardized Process Data

Returning to the exemplary situation from the beginning of Chapters 1.3.1 and 1.3.2 once more: *When Ella turns around to see her other classmates' reactions, she finds* (depending on the experimental manipulation) *the majority of them* (a) *smiling and eagerly raising their hands to answer the teacher's question* or (b) *staring at their desks or cautiously looking around*. Depending on what Ella sees when she looks around the classroom, she will respond to and interpret it in a certain way (e.g., *the others all smile knowingly and raise their hands because they want to show the new teacher that they already know the answer* vs. *the others seem to share my sentiment and don't know the answer to the teacher's question*). This example illustrates the different kinds of social information processing associated with social comparisons in the classroom (see Figure 5, Chapter 1.2.3). On the one hand, there is Ella's covert social comparison behavior, reflected in her intrapersonal cognitive response to the social comparison information (i.e., her interpretation of what she sees). On the other hand, there is Ella's overt social comparison behavior, reflected in her interpersonal behavioral response to the social comparison information (i.e., her looking around to see her classmates' reactions to the teacher's question).

IVR environments provide the opportunity to examine these cognitive and behavioral processes in a standardized environment. In other words, whereas students' self-reports always capture the student's subjective perception of social comparison information, IVRs as experimental environments make it possible to ensure that all students experience the exact same classroom situation from an identical perspective, and respective self-reports of how students interpret social comparison information therefore refer to a standardized comparison target (compared with various situational influences in typical field research). Moreover, it is assumed that students' internal and therefore covert responses to social environments (i.e., their social cognition) are to a large extent reflected in their overt behavioral responses to social cues, which are directly observable (e.g., De Jaegher et al., 2010; Lou et al., 2019). IVR technology makes it possible to collect standardized process data that reflect these overt social comparison behaviors, which I will describe in more detail in the following.

Using technology-based measures to gain insights into social information processing and social cognition, psychologists have used, for instance, neuroimaging techniques (Gallese et al., 2004; Kedia et al., 2014; Lieberman, 2005; Posner & Rothbart, 2007; Schilbach et al., 2008) or other physiological signals, such as heart activity and electrodermal activity (Burnside & Ullsperger, 2020; Poh et al., 2010; Scheepers, 2009; Scheepers & Ellemers, 2005). Notably, such measurements are primarily useful for answering very specific research questions because, otherwise, the potential for inference is very large, and the respective conclusions are often accused of suffering from a lack of validity (Girard & Cohn, 2016). A special and slightly different approach is eye-movement analysis, which allows researchers to integrate contextual information (i.e., not just the physiological reaction but the source of the reaction by determining which object the person's gaze is focused on). Against this background, the analysis of gaze data presents a (if not *the* most) promising avenue to gain unbiased and yet interpretable insights into (social) information processing and is the focus of the following explanations.

Eye movements have been used to answer a variety of research questions in different scientific fields (Kowler, 2011; Lai et al., 2013). The central argument in almost all eye-tracking literature is the so-called *eye-mind link*, suggesting that eye movements directly reflect cognitive processes (Just & Carpenter, 1976; Rayner, 1998; Reichle et al., 2012). On the basis of this assumption, in the context of teaching and learning, researchers have made use of eye tracking primarily to investigate learners' reactions to certain instructional material and different types of tasks (Jarodzka et al., 2017; Lai et al., 2013; Olney et al., 2015; Scheiter & Eitel, 2017; Strohmaier et al., 2020). With regard to the classroom context, a large amount of research has focused on how teachers notice, interpret, and react to what is happening in the classroom; however, studies examining how students perceive whole classroom situations—and not just single learning materials and types of tasks—are scarce (Jarodzka et al., 2021; Kaakinen, 2021). This is not surprising considering that it presents quite a challenge in real-world classrooms to record eye movements in a standardized way.[13] IVR technology provides the ideal opportunity to move out of traditional lab settings and examine authentic gaze

[13] Recent advances in mobile eye-tracking technology have made the collection of eye-tracking data in real-life environments much easier. However, the interpretation of the respective data from natural settings still faces a great challenge considering that situational and contextual influences cannot be controlled in the field.

behavior, including head movements and orientation in a 3D space (Foulsham et al., 2011; Tatler & Land, 2015) while tracking every move and gaze of students in an experimentally controlled environment (Bailenson et al., 2004; Bailenson et al., 2002). Using the benefits of IVR environments as standardized environments, established and also more sophisticated methods of analyzing eye movements can be applied to eye-tracking data from IVR classroom scenarios and used as indicators of students' perceptions of (social comparison) information (Holmqvist et al., 2011; Hutmacher, 2019; Jarodzka et al., 2017). Importantly, the interpretation of eye-tracking data is context-specific, and the appropriate analysis technique therefore has to be chosen carefully for each research design (Kaakinen, 2021). Table 2 provides an overview of different approaches that I will outline and discuss with respect to their application to identifying social comparison processes in an IVR classroom in the following.

Table 2. Different Approaches for Analyzing Eye-Tracking Data With Regard to (Social) Information Processing in the Classroom

Overall approach	Analysis examples	Temporal focus	Contextual integration	Interpretation
Eye movements and eye-related data	Fixations	Milliseconds	n/a	Cognitive processing effort
	Saccades	Milliseconds	n/a	Search behavior
	Pupil diameter	Milliseconds	n/a	Cognitive load, affective arousal
Visual attention on objects of interest	Gaze shifts to OOI	Aggregated	Object(s) of interest	Attentional focus shift to OOI
	Time on OOI	Aggregated	Object(s) of interest	Duration of attentional focus on OOI
Visual attention patterns	Scanpaths	Aggregated	Object(s) of interest + gaze transitions	Frequencies and patterns of attention shifts between OOIs
	Gaze-based networks	Aggregated	Object(s) of interest + gaze transitions	Structure of attention distribution between OOIs

Note. OOI = object of interest (in the classroom context, e.g., peer learners, the teacher, or instructional content on the board).

The commonly performed eye-tracking analysis of fixations (i.e., gaze resting at a particular place) and saccades (i.e., gaze shifting from one fixation to another) operates on the level of milliseconds and does not include contextual information. Fixations and saccades provide important insights into general information as they are typically interpreted such that

that longer fixation durations indicate increased cognitive processing, and longer saccades indicate increased search behavior (Goldberg & Kotval, 1999; Just & Carpenter, 1976; Salvucci & Goldberg, 2000). In a similar vein, pupillometry provides insights into information processing (Binda & Gamlin, 2017); greater pupil diameters are commonly associated with cognitive load (Appel et al., 2018; Coyne et al., 2017; Granholm et al., 1996; Jainta & Baccino, 2010; Souchet et al., 2021; Zu et al., 2020), information encoding and retrieval (Goldinger & Papesh, 2012; Otero et al., 2011; Privitera et al., 2010), and affective arousal (Bradley et al., 2008; Cebeci et al., 2019; Henderson et al., 2018; Maier & Grueschow, 2021; Partala & Surakka, 2003). However, similar to the aforementioned other physiological measures of brain or heart activity that do not include any contextual references, it is difficult to gain insights into the specifics of how individuals attend and respond to social information that is based on only fixations, saccades, and pupil diameters. In other words, eye movements alone may provide insights into cognitive and affective processes (i.e., eye-mind link; Just & Carpenter, 1976; Rayner, 1998; Reichle et al., 2012), but integrating information about what is looked at (and for how long) bears especially great potential for understanding individuals' need to process and gather information from these particular sources (Mudrick et al., 2019).

The concept of visual attention[14] (Bundesen, 1990; Carrasco, 2011; Chun et al., 2011; Folk, 2015; Lodge & Harrison, 2019; Tatler & Land, 2015) integrates eye movements with contextual information and therefore makes it possible to interpret eye-tracking data semantically. Visual attention builds on the so-called eye-mind link and furthermore assumes that eye movements are closely related to selective attentional processes (Findlay & Gilchrist, 2003; Hayhoe & Ballard, 2005; Tatler & Land, 2015). The concept of visual attention mostly goes back to Posner and Boies (1971), who found that visual focus on a certain stimulus was associated with more efficient processing of the respective information (see also Posner, 1988). Applying this to students' processing of social comparison information in the classroom, it can be assumed that students guide their attention in the classroom and focus on certain objects (e.g., their classmates) while ignoring others (e.g., the teacher). Consequently, what they look at can serve as an indication of what they pay attention to and compare themselves with.[15]

[14] Also referred to as "external" or "overt" visual or spatial attention (see, e.g., Chun et al., 2011; Folk, 2015).

[15] In the traditional eye-tracking literature, visual attention is typically analyzed with regard to certain predefined *areas* of interest (AOIs; see, e.g., Holmqvist et al., 2011). In the context of gaze data from an IVR environment, it is more common to refer to *objects* of interest (OOIs; see, e.g., Bozkir et al., 2021). Instead of identifying gazes in a certain area of interest on a two-dimensional stimulus as in most traditional eye-tracking set-ups, gaze measures from an IVR environment can use dedicated colliders to directly identify when a person's gaze hits upon certain objects of interest in the 3-D space. This simultaneously accounts for the dynamics and movements of virtual objects of interest as the aforementioned colliders move with the virtual objects, and gaze hits can be directly identified at any time.

Importantly, visual attention is suggested to be contextually cued (Chun, 2000), which means that, over time, humans adapt certain schemata that guide their visual attention in specific situations. The role of context has been found to play a particularly important role in social situations (e.g., Kenrick et al., 2003; Landau et al., 2010; Richardson & Gobel, 2015). As Richardson and Gobel (2015) argued: “Depending on the situation, social norms can both reduce social attention or increase social attention to other people” (p. 356). Generally, attention to social counterparts is suggested to be relatively high in classroom situations (Weinstein, 1991). However, information-processing models highlight the role of individual differences and the fact that some types of information attract more attention and are consequently more likely to be processed than others (e.g., Kenrick et al., 2003). This relates to different mechanisms of top-down (i.e., shaped by individual dispositions and willful control) versus bottom-up (i.e., shaped by situational cues and spontaneous reactions) processes of visual attention (see, e.g., Theeuwes et al., 2000). Notably, both top-down and bottom-up processes of visual attention are intertwined, particularly in social situations (Richardson & Gobel, 2015). Accordingly, with regard to social comparisons, it can be assumed that certain situational cues (e.g., particularly notable behaviors or statements) most likely draw attention to social comparison information more than others (Richardson & Gobel, 2015; Theeuwes et al., 2000). However, at the same time, it can be assumed that students’ active social comparison behavior is reflected in top-down attentional processes that are oriented toward social information and guided by individual characteristics, such as prior knowledge, willful plans, and current goals (Katsuki & Constantinidis, 2014).

Against this background, it seems additionally worthwhile not only to examine visual attention to single objects of interest (e.g., peer learners) but to identify patterns in how students gather information in the classroom. Such patterns are reflected in (consecutive) shifts in a person’s gaze between different objects of interest (e.g., from peer to peer or between the teacher and peers). A common method used to analyze such visual attention patterns and respective gaze shifts involves so-called scanpaths (Anderson et al., 2015; Kübler et al., 2017; Le Meur & Baccino, 2013), which aggregate transitions in a person’s gaze from one object to another over a certain time period (e.g., gaze directed toward Classmate A, then the teacher, and after the teacher toward Classmate B). Scanpaths are typically used to differentiate between different types of visual attention distribution and respective search patterns (Findlay & Gilchrist, 2003; Kaakinen, 2021), typically aided by machine-learning models that help to identify certain patterns in the underlying matrices (e.g., Kübler et al., 2017). Moreover, gaze transitions can be analyzed using network analysis (see, e.g., Guillon et al., 2015; Sadria et al., 2019; Schneider et al., 2013; Yazdan-Shahmorad et al., 2020). Whereas network analysis is based on the same data matrices as scanpaths, network analysis describes the distribution of visual attention with a number of features that can be calculated on the basis of the network structure using established methods from graph theory (Diestel, 2017; Erciyes, 2021) and network analysis (Chiesi, 2001; Curtin, 2018). With regard to students' social information processing in the classroom, this allows researchers to identify, for example, the focus of participants' gaze transitions (e.g., Do their gazes center on the teacher or rather the peer learners) or the connectedness of their gazes (e.g., Do they repeatedly look at different peers in a row or only gaze at single peers and then back to the teacher).

In sum, in order to gain insights into what I outlined as the black box of social comparisons and the respective processing of social information in the classroom (see Chapter 1.2.3), the potential of IVR is twofold: First, it provides a standardized environment that can be used to systematically examine how students perceive social comparison information (i.e., in terms of their cognitive response to and interpretation of it). Second, against the background that state-of-the-art IVR technology comes with integrated eye-trackers, IVR provides the potential to measure students' visual attention objectively and unobtrusively during their IVR experience as a reflection of their behavioral responses to the social comparison information. IVR makes it not just possible to collect the eye-tracking data but moreover provides high-quality data in terms of the standardized situation, including standardized lighting conditions, which are critical for the valid interpretation of eye-tracking data (Attard-Johnson et al., 2019; Kaakinen, 2021).

2

AIMS AND RESEARCH QUESTIONS

2 Aims and Research Questions

Against the background that students' academic self-concept plays a central role in students' learning and academic trajectories (Chapter 1.1.1), the previous chapters discussed the role of social comparisons as a major determinant of students' academic self-concepts. Despite a large body of research examining the effects of different reference groups in real-world school/class settings (see BFLPE, Chapter 1.2.1) and a considerable number of (primarily lab-based) experimental studies on the mechanisms of social comparisons (Chapter 1.2.2), I demonstrated that an in-depth understanding of the actual processes by which social comparisons are made in the classroom is still lacking ("the black box of social information processing in the classroom," Chapter 1.2.3). In the preceding Chapter 1.3, I thus outlined the potential of immersive virtual reality (IVR) classrooms as an experimental tool that can be used to gain insights into social comparisons in the classroom in a standardized and yet authentic manner. Building on this, the present dissertation's overarching research aim is to revisit social comparison research in the classroom and gain insights into underlying processes by using IVR as a research tool. Based on the model that I am proposing to be used to examine social comparisons in the classroom (see Figure 5, Chapter 1.2.3), the present dissertation has two subordinate objectives.

First, the present dissertation is aimed at achieving a theoretical advancement in terms of a more systematic and in-depth understanding of social comparisons in the classroom. In other words, the present dissertation is aimed at opening the black box of social information processing in the classroom and gaining insights into social comparisons as the mechanism that underlies the repeatedly found BFLPE pattern of results. To this end, the present dissertation is aimed at describing social comparison processes through covert (i.e., intrapersonal) and overt (i.e., interpersonal) behaviors that reflect students' cognitive and behavioral responses to social comparison information, respectively. In order to advance the theoretical understanding of social comparisons in the classroom, the present dissertation is therefore aimed at answering the following research questions:

1) How do students' cognitive responses to social comparison information (i.e., covert *intra*personal behavior associated with the processing of social information in the classroom) explain the BFLPE?
2) How do students' behavioral responses to social comparison information (i.e., overt *inter*personal behavior associated with the processing of social information in the classroom) explain the BFLPE?

Second, the present dissertation is aimed at achieving a methodological advancement to gain insights into social comparisons and the respective processing of social information in the classroom by using an IVR classroom. The potential of IVR environments as an experimental tool is evident as it allows researchers to authentically simulate a classroom scenario while simultaneously providing full experimental control (e.g., to manipulate the social comparison information provided by the virtual peer learners). Moreover, state-of-the-art IVR technology makes it possible to collect fine-grained process data reflecting students' overt social comparison behavior, specifically their visual attention to social comparison information. However, to date, IVR classrooms have not been used much in research in educational or social psychology, at least not to gain insights into classroom processes. Therefore, the present dissertation is aimed at answering the following question:

3) How can an IVR classroom be used as an experimental tool, specifically to gain insights into social comparisons and the corresponding processing of social information in the classroom?

I draw on three empirical studies to address the dissertation's aims and answer the aforementioned questions. In the following, I provide a brief description of the IVR classroom that was used as an experimental tool in the three studies and outline the studies in more detail.

Figure 8. Different performance levels indicated by the hand-raising behavior of virtual peer learners. The figure depicts the same situation during an IVR lesson with 20% of the students raising their hands (top left), 35% of the students raising their hands (top right), 65% of the students raising their hands (bottom left), and 80% of the students raising their hands (bottom right).

The IVR classroom used in the present dissertation was a fully preprogrammed simulation of a classroom situation. The IVR classroom simulation was designed for sixth-grade students to experience a 15-min lesson on basic principles of computational thinking in the IVR classroom. The IVR scenario was based on motion captures and audio recordings from a real classroom and consisted of a dialogue between the virtual teacher and virtual students; the virtual teacher explained the lesson material, asked questions, and presented a task for the virtual students to work on, and the virtual students answered the questions (see the full script of the IVR classroom scenario in the Appendix).[16] Most importantly with regard to the investigation of social comparison processes in the IVR classroom, the virtual peer learners' performance-related behavior was manipulated. Specifically, the proportion of virtual classmates who raised their hands to answer the virtual teacher's questions or to indicate that they knew the correct solution to a task during the IVR lesson was systematically varied on a between-subjects level (see Figure 8). The experimental manipulation was designed to systematically vary the social comparison information provided by the peer learners while all other factors (e.g., the teacher's feedback) were held constant.

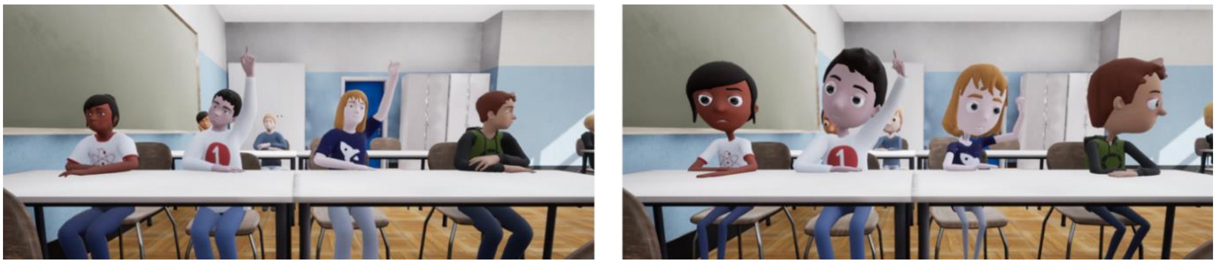


Figure 9. A realistic graphical representation of virtual learners (left) and a cartoon-styled graphical representation of virtual learners (right).

Aside from the experimental manipulation of the virtual peer learners' performance level (i.e., hand-raising), the IVR classroom included two additional variations. Students experienced the IVR classroom situation either from a position in the front of the classroom (second row) or from the back (last row) of the IVR classroom. Additionally, virtual avatars in the IVR classroom were visually represented in a realistic manner or as cartoons (see Figure 9). Both of these configurations were counterbalanced across the hand-raising conditions.

[16] A detailed description of the IVR classroom scenario is included in the Method sections of Studies 1 to 3 presented in Chapters 3 to 6.

The present dissertation's research questions were addressed in three empirical studies that can be located in the theoretical model that I propose to examine social comparisons in the classroom (see Figure 7).

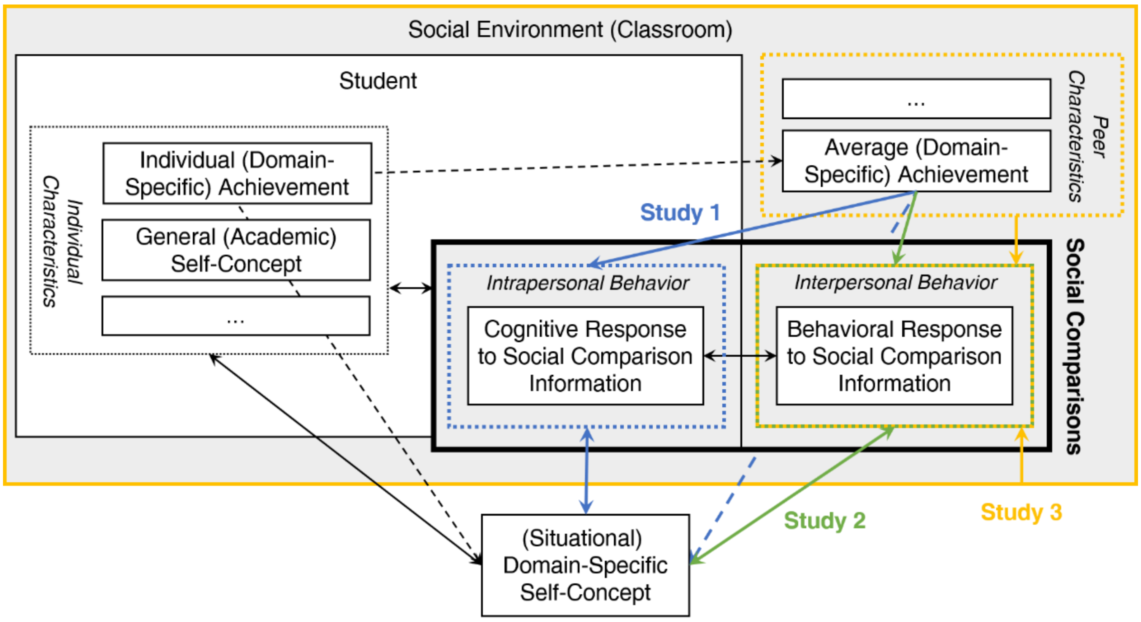


Figure 7. Locations of the empirical studies in the theoretical model used to examine social comparisons in the classroom. Only the central research question of each study is represented.

Study 1 (*Does a 15-minute exposure to strong classmates affect students' self-concept? An experimental test of the Big-Fish-Little-Pond effect using an immersive virtual reality classroom*) examined whether the typically found and empirically well-supported effects of social comparisons on learners' self-concepts (i.e., the BFLPE) could be reproduced in an experimental setting by using an IVR classroom in which classmates' performance-related behavior was manipulated (i.e., their hand-raising behavior). Aside from obtaining experimental evidence to corroborate the BFLPE, the study was aimed at investigating the respective underlying social comparison processes leading to the effects. More specifically, the study used students' self-reports of how they perceived their classmates' implicit performance-related behavior (i.e., the experimentally manipulated hand-raising behavior) in the classroom to see how students' perceptions of their classmates' implicit performance-related behavior explain interindividual differences in their self-concepts. In this vein, the study particularly targeted the first aim of the dissertation, to provide an answer to the question of how students' cognitive responses to social comparison information (i.e., their interpretation of social information provided in the classroom) explains the BFLPE.

Study 2 (*Do students actively seek comparisons with others? Using eye-movement data from a virtual reality classroom to uncover social information processing*) examined social comparison processes in the IVR classroom with regard to students' overt behavioral responses to social comparison information. Specifically, the study used students' visual attention to their virtual peer learners (i.e., How many peer learners a student looked at, how often, and for how long) as a proxy for active social comparisons initiated by the students. Based on this, the study examined (a) the extent to which students' behavioral responses to social comparison information are affected by the experimentally manipulated performance-related behavior of peers, and (b) how students' behavioral responses to social comparison information are related to interindividual differences in situational self-concept. In this vein, Study 2 focused on the first aim of the dissertation to provide an answer to the question of how students' behavioral responses to social comparison information (i.e., their interpersonal behavior in the sense of visual attention to their peer learners) explain the BFLPE.

Study 3 (*Configuring an immersive virtual reality classroom for educational research and practice: Implications from students' gaze-based attention networks*) examined how different configurations of an IVR classroom affect students' processing of respectively provided (social) information. Specifically, the study used students' gaze-based visual attention networks to identify how the configuration of the social environment and peer characteristics in the IVR classroom (i.e., locations of the seats of participating students, style with which the virtual avatars were represented, performance-related behavior of virtual peers) impact students' overt behavioral responses to the (social comparison) information provided in the IVR classroom. In addition, the study examined how students' gaze-based attention networks in the IVR classroom are related to their situational self-concept, interest, and performance after the IVR lesson. Study 3 thereby focused in particular on the second aim of the dissertation and moreover widens the focus from using the IVR classroom to gain insights into social comparison processes in the classroom to the use of an IVR classroom for educational psychology research (and practice) in general.

3

STUDY 1

Hasenbein, L., Trautwein, U., Hahn, J.-U., Soller, S., & Göllner, R. (2021). *Does a 15-minute exposure to strong classmates affect students' self-concept? An experimental test of the Big-Fish-Little-Pond effect using an immersive virtual reality classroom.* Manuscript to be submitted.

The following manuscript has not yet been accepted or published. The version displayed here might not exactly replicate the final version published in the journal. It is not the copy of record.

Abstract

Academic self-concept plays a central role in successful learning and is substantially shaped by social comparisons. Research on the so-called Big-Fish-Little-Pond effect (BFLPE) has yielded a highly robust and generalizable pattern of negative effects of higher class/school average achievement on students' self-concept when controlling for individual achievement. Most BFLPE research has used data from authentic learning environments, yielding high ecological validity, but most studies have not provided information about the causes behind or the mechanisms underlying the proposed effects. We used a fully immersive virtual reality (IVR) classroom to experimentally test the extent to which students recognized implicit performance-related classroom behavior as social comparison information and how these perceptions explained differences in students' self-concepts. Participants (N = 381 sixth-grade students) experienced an authentic and yet standardized IVR teaching situation with classmates with different hand-raising behaviors (20% vs. 35% vs. 65% vs. 80% high-performing). We investigated effects on socially-oriented, criterion-oriented, and dispositional domain-specific self-concepts. Hand-raising behavior had a statistically significant positive effect on students' perception of the class' performance level ($d_{20\%\text{vs.}65\%} = 0.60$; $d_{20\%\text{vs.}80\%} = 1.24$), indicating that students actively process implicit performance-related comparison information. In line with the BFLPE, results showed a negative effect of higher performing classmates on students' socially-oriented self-concept ($d_{20\%\text{vs.}80\%} = 0.30$). Students' perceptions of the class' performance level fully mediated the effect of hand-raising behavior on students' socially-oriented self-evaluations. Results provide new insights into the emergence of social comparison effects in the classroom and make a general contribution to the use of IVR in experimental research.

Does a 15-Minute Exposure to Strong Classmates Affect Students' Self-Concept? An Experimental Test of the Big-Fish-Little-Pond Effect Using an Immersive Virtual Reality Classroom

From an early age on, we evaluate ourselves: We want to know how competent we are in all sorts of things, and over time, every one of us develops an individual picture of how well we do in different areas of our lives (Eccles et al., 1989; Wrzus & Roberts, 2017). How we perceive our own abilities plays an important role in individual life trajectories: For instance, self-concept has been found to be closely related to self-esteem (e.g., Campbell, 1990; Greenwald et al., 2002; Swann et al., 2007) and well-being (e.g., Cross et al., 2003; Marsh et al., 2006). Moreover, academic self-concept has been found to be associated with individual achievement (e.g., Eccles & Wigfield, 2002; Marsh & Craven, 2006; Trautwein & Möller, 2016), achievement emotions (Pekrun et al., 2019), academic choices, and career aspirations (e.g., Marsh & Yeung, 1997; Nagengast & Marsh, 2012).

How self-concept emerges and how individual differences subsequently manifest themselves have been key interests of psychologists from different fields (i.e., differential, social, educational) for several decades. There are many sources of our self-evaluations (e.g., Brewer & Weber, 1994; Wolff et al., 2018), yet research has repeatedly highlighted the role of social comparisons for individual differences in self-concept (Buunk & Gibbons, 2007; Festinger, 1954; Suls et al., 2002). In other words, we often refer to others when assessing our own abilities and—even if these comparisons sometimes happen only subconsciously—how we eventually perceive ourselves greatly depends on our respective reference group. A substantial body of research situated in educational settings has found that this is also true in school (see Dijkstra et al., 2008 for a review): Students' evaluations of their own abilities are affected by their peers and the reference group they are surrounded by (e.g., classmates, school). The effects of certain reference groups on students' self-evaluations are encapsulated in the perhaps most prominent model of social comparison effects in classrooms, the so-called Big-Fish-Little-Pond effect (BFLPE; Marsh, 1987; Marsh, Seaton, et al., 2008). The BFLPE suggests that students constantly evaluate themselves and that students' academic self-concept suffers when they are placed in high-achieving environments where they are below the average performance level. Conversely, students evaluate their own abilities more positively in learning environments with a lower average performance level than their own. Unless students are "the big fish in the pond" (i.e., above the average performance level of their reference group), the BFLPE implies that they are forced to constantly compare their own achievement to more able peers and to consequently evaluate themselves as less competent. Previous research has extensively demonstrated the negative relationship between the average performance level of a class and students' self-concept when individual ability is controlled for (e.g., Seaton et al., 2009).

Attesting to social comparison effects in authentic classroom settings, BFLPE research typically uses achievement data and students' self-reports of their academic self-concept that are collected in real-world classrooms. However, given the restrictions on opportunities to implement far-reaching experimental manipulations in schools, this also means that BFLPE research typically has to rely on nonexperimental designs. Such correlational research is limited in the extent to which it can provide causal evidence for the proposed (social comparison) processes; moreover, except for a few exceptions (e.g., Huguet et al., 2009), the underlying mechanisms are typically deduced rather than observed or manipulated. Hence, despite an abundance of empirical support for the BFLPE, surprisingly little is known about how social comparisons in the classroom proceed and ultimately impact learners' self-concept, prompting some critics of the BFLPE (e.g., Dai & Rinn, 2008) to cast doubt on some of the processes that are assumed to underlie this effect.

To overcome this shortcoming, we used fully immersive virtual reality (IVR) as a methodology to combine the advantages of field research in real classrooms with the affordances of experimentally controllable settings. The study's objective is to contribute to a deeper and more systematic understanding of social comparison processes in classrooms and their associated effects on self-concept. We aimed to investigate the extent to which and how students perceive their classmates' classroom behavior and how these perceptions explain individual differences in students' self-concept. We examined (a) whether different proportions of high-achievers in a 15-min IVR classroom situation would lead to distinct perceptions of a class' performance level and (b) whether the typically found and empirically well-supported effects of social comparisons on learners' self-concept (i.e., the BFLPE) could be reproduced in an experimental setting as such. Moreover, we examined (c) how the perceived performance level of the class affected the relationship between the manipulated proportion of high-achieving students and the expected results on students' self-concept.

Self-Concept: Its Importance and How It is Shaped by Social Comparisons

One of the most critical determinants of successful learning and individuals' development in school and beyond is academic self-concept. Self-concept generally refers to ideas and evaluations of various aspects of one's own person (Marsh & Shavelson, 1985). Current self-concept models in educational psychology share the notion that self-concept is best described by a multifaceted and hierarchical structure (Brunner et al., 2010; Marsh & Shavelson, 1985; Shavelson et al., 1976). With respect to the different facets of self-concept, educational psychology has focused in particular on domain-specific academic self-concepts (e.g., Marsh & Shavelson, 1985), which play an important role in learning: High academic self-concept suggests high levels of confidence in one's own abilities (in a certain domain) and in one’s respective performance in school (Arens et al., 2019; Bong & Skaalvik, 2003; Eccles & Wigfield, 2002; Harter, 1986; Trautwein & Möller, 2016). Academic self-concept has been found to be an important predictor of the achievement of academic goals and successful education and career choices, even when other background variables are controlled for (Guay et al., 2003; Helmke & van Aken, 1995; Marsh & Craven, 2006; Marsh & Martin, 2011; Marsh & O'Mara, 2008; Valentine et al., 2004).

Regarding the hierarchical structure, theoretical conceptualizations speak of a more stable (i.e., dispositional, trait-like) self-concept on a higher level and an increasingly situation-specific and consequently less stable (i.e., state-like) self-concept on a lower level (Brunner et al., 2010; Marsh & Shavelson, 1985; Shavelson et al., 1976). Recently, the importance of situation-specific aspects of individuals’ self-beliefs has been re-emphasized (Eccles & Wigfield, 2020). Whereas other factors, including top-down processes, might play an additional role in shaping evaluations of the self in specific domains, it is assumed that students’ repeated experiences of success or failure in similar situations are particularly important for leading to changes in domain-specific self-concepts (Harter, 1986; Shavelson et al., 1976). This assumption is also in line with models from personality psychology that link short-term situational processes with long-term personality development (e.g., Wrzus & Roberts, 2017). Single experiences of success or failure and their respective bottom-up processes draw on a range of sources of information, including social (e.g., Festinger, 1954) and criterial information (e.g., Ferring & Filipp, 1996). Notably, typical classroom situations are characterized by a prevailing evaluative atmosphere and a large amount of social comparison information (Dijkstra et al., 2008; Levine, 1983).

What could the emergence of individual differences in learners’ self-concept look like in an actual learning situation with fellow students? Imagine two new students entering a classroom with about 20 unknown classmates for the first time. Imagine that this is the first lesson in a subject that just got introduced in their curriculum. Not yet having a clear picture of their own abilities in this subject, it is likely that the new students will refer to their peers in the classroom and see how the peers perform in this specific situation—possibly evaluating themselves as better or worse than others in this new subject on the basis of their relative standing in the class. In the remainder of this article, we refer to this emerging and primarily reference-group-based self-concept as *socially-oriented self-concept*. Over the course of the lesson, the students also work on tasks that they can or cannot solve—again, evaluating themselves as better or worse than others on the basis of their recurring experiences with the demands of the lesson. We refer to this as *criterion-oriented self-concept*. Eventually, the students will have had many lessons in the once new subject, they will have had repeated experiences of success or failure, and on the basis of this, they will have developed a relatively stable idea of their own abilities in the subject (domain-specific *dispositional self-concept*). In other words, students are assumed to develop beliefs about their own competence in specific situations and domains, which are initially shaped by recurring comparisons with the available reference group as well as by recurring comparisons with the performance standards or requirements of this specific situation. These beliefs are assumed to become more stable with repeated experiences throughout childhood and adolescence (Suls & Mullen, 1982; Wigfield et al., 2015).

Since first identified in Festinger's social comparison theory (Festinger, 1954), social comparisons have been widely acknowledged as a central aspect of human interaction (Buunk & Gibbons, 2007; Buunk & Mussweiler, 2001) and are particularly strongly linked to self-evaluations (Buunk et al., 2007; Cialdini & Richardson, 1980; Helgeson & Mickelson, 1995). The most prominent and empirically well-supported pattern of results in the context of social comparison effects on learners' self-concept in real-world classrooms is the Big-Fish-Little-Pond effect (BFLPE; Marsh, 1987; Marsh, Seaton, et al., 2008). The BFLPE suggests that higher ability schools or classes have a negative impact on academic self-concept. According to the BFLPE, equally able students differ in their self-concept when they attend classes with different performance levels. BFLPE research has suggested that differences in self-concept result from different learning environments and comparisons with the respective reference group that a student is surrounded by. Students compare themselves with either their more high-achieving peers (upward) or their lower achieving counterparts (downward). The BFPLE typically refers to students perceiving an upward comparison as a contrast (i.e., assessing their own abilities as different from the comparison targets). Consequently, a lasting upward comparison with a higher performing reference group then results in lower self-concept (Diener & Fujita, 1997; Marsh, Trautwein, et al., 2008; Mussweiler et al., 2004; Mussweiler & Strack, 2000). The BFLPE has been shown to be very robust and valid across culturally and economically different countries (Marsh & Hau, 2003; Marsh et al., 2000; Marsh et al., 2020; Seaton et al., 2009). Moreover, BFLPE research is not restricted to detrimental associations between upward comparisons and learners' academic self-concept, but it also suggests a lasting impact on educational pathways, for instance, by being related to career aspirations (e.g., Davis, 1966; Göllner et al., 2018; Nagengast & Marsh, 2012; von Keyserlingk et al., 2020).

Open Questions in Traditional BFLPE Research: Causality of Effects and Underlying Mechanisms?

There is a large body of existing research supporting social comparison effects in terms of the BFLPE. However, BFLPE research typically uses the average performance level of schools/classes and the respective individuals to examine reference-group effects (Dicke et al., 2018; Marsh, Seaton, et al., 2008). It primarily addresses the results of social comparison effects in the classroom by means of regression analyses that are correlational in nature. Hence, despite the abundance of empirical support for the BFLPE, central questions about the causality of the proposed effects and the underlying mechanisms remain open.

A first issue is that, with respect to the causality of the proposed effect, experimental support for the BFLPE is scarce (Zell & Alicke, 2009, 2010). Aiming to identify causal relationships between social comparison information and the resulting differences in self-concept, experimental research has usually worked with lab-based settings, which clearly differ from authentic classroom environments. In the typical experimental design that is used to investigate social comparison effects, participants are seated alone or in a small group in a lab room and are asked to work on a task; subsequently, they are provided with information about their performance compared with a (mostly fictitious) reference (e.g., another student or other participants). Such studies have supported the role of social comparison information for individual differences in self-concept and have allowed causal conclusions to be drawn. However, it is not clear whether they examined the same phenomenon as described in the majority of BFLPE studies in authentic classroom environments. A first issue here is that, by providing social comparison information by offering manipulated performance feedback (e.g., Pyszczynski et al., 1985; Wheeler, 1966), the effects of the social comparisons that are being tested are artificially induced rather than naturally observed. A second issue is that comparison choice studies explicitly instruct students to make a comparison with a specific comparison target of their choice (e.g., Blanton et al., 1999; Dumas et al., 2005; Huguet et al., 2001; Seaton et al., 2008; Suls & Wheeler, 2000), and therefore, students' attention is drawn specifically toward social comparisons that might not occur in the same way in more authentic learning environments. A third issue is that, in experimental designs that introduce participants to a given (mostly fictitious) comparison target (e.g., Mussweiler, 2003; Mussweiler et al., 2004;

Mussweiler & Strack, 2000; Neugebauer et al., 2016), not only are the social comparisons forced, but the respective comparison information is purposefully provided by the researchers in a controlled manner. Not surprisingly, depending on the experimental design and the specificities of the manipulation, experimental research on social comparisons provides different results. On the one hand, there is some experimental evidence for the assumptions expressed in the BFLPE when manipulated social comparison feedback is given to students (e.g., Zell & Alicke, 2009). On the other hand, in contrast to what the BFLPE implies, some social psychological experiments have found positive effects of upward comparisons on self-perception (Blanton et al., 1999; Blascovich et al., 1999), particularly when highlighting similarities between the participants and the higher achieving comparison targets (Mussweiler, 2003; Mussweiler et al., 2004). Taken together, experimental findings have suggested that distinct reference groups and respective social comparisons are in fact the cause of the BFLPE. However, existing experimental approaches (i.e., manipulated performance feedback, forced comparisons with a fictitious or explicit comparison target) are not able to adequately capture the complexity or reflect the nature of social comparisons in classrooms (Möller & Köller, 2001). Therefore, the extent to which social comparisons are responsible for the observed effects on self-concept in a real-world classroom setting remains unclear.

A second issue in traditional BFLPE research is that it has provided a lot of evidence for the consequences of social comparisons (i.e., effects on self-concept; Trautwein & Möller, 2016), but information about the underlying cognitive and affective processes of social comparisons is scarce (Buunk & Mussweiler, 2001). How exactly do social comparisons in classrooms proceed, and how pervasive are they? Under which circumstances do learners compare themselves with their classmates and what do they refer to when they do? Depending on contextual and individual factors, social comparisons happen more or less deliberately (Mussweiler, 2001; Wheeler & Suls, 2005). In general, the classroom has been referred to as a "total environment" (Diener & Fujita, 1997, p. 350) that (a) presents students with a naturally limited amount of available comparison information and (b) makes it hard to disregard these social influences. Students can presumably hardly do anything other than make use of the ubiquitous social comparison information in the classroom. In other words, students naturally and continuously compare themselves with their classmates in order to assess their own standing. Research on the BFLPE has highlighted the central role of students' perceived standing relative to their classmates (Huguet et al., 2009). However, BFLPE research has hardly ever focused on actual classroom behavior and the role of students' individual perceptions thereof (Trautwein & Lüdtke, 2005). The BFLPE suggests that social comparisons happen rather implicitly with a generalized other (i.e., the average performance level), whereas social psychological research works with explicit social comparison information that stems from a (mostly predefined) specific other (i.e., comparisons with specific school- and classmates; Marsh, Trautwein, et al., 2008). Consequently, declaring that social comparisons are the mechanism that underlies the BFLPE is only an indirect conclusion rather than an explicit examination thereof. Notably, when individual grades have been controlled for, the BFLPE has been considerably less pronounced (e.g., Trautwein et al., 2006). Such findings point to the potentially important role of teacher feedback and grading practices for students to develop different levels of self-concept (e.g., Lüdtke et al., 2005; Trautwein & Lüdtke, 2005); in fact, such results prompt the question of whether teacher feedback—and not students' active social comparison processes—are perhaps the driving factor that underlies the BFLPE. Taken together, it remains unclear to what extent students realize how good their classmates are in a regular classroom situation and to what extent rather implicit social comparison information is related to individual differences in self-concept or whether students need explicit performance feedback to make comparisons with.

A third issue in BFLPE research is that, assuming that students have a certain perception of their classmates' performance level, BFLPE research has not yet answered the question of how exactly these perceptions impact self-concept. How quickly do the negative effects of high-achieving reference groups on learners' self-concept occur? BFLPE research to date has mostly worked with large-scale data from real learning contexts, focusing on learners' general domain-specific self-concepts. Considering that specific contexts and comparisons with present reference groups substantially shape self-concept (e.g., Becker & Neumann, 2016, 2018) and attesting to the situation-sensitivity of competence beliefs (Eccles & Wigfield, 2020), it seems worthwhile to go beyond only investigating social comparison effects on more global criterion-oriented and dispositional self-concepts by also considering situation-specific perceptions of the self. In other words, going back to our example from the beginning about two new students entering a class for the first time: We might not see immediate reference-group effects on the students' criterion-oriented and especially their dispositional self-concepts in the respective domain after one lesson—more precisely, we would even hope that at least the latter would not be the case, considering the far-reaching associations of self-concept with further academic trajectories. Initially, we might be most likely to observe social comparison effects on situation-specific self-evaluations (e.g., socially-oriented self-concept), assuming that individual differences in self-concept emerge in the specific situations in which students first compare themselves with the respective reference group. Empirical evidence from intervention studies on learners' self-beliefs has suggested that effects on self-concept can occur even after relatively short interventions (e.g., Brisson et al., 2017). Whereas there is sound evidence that domain-specific self-concept can easily be affected by interventions (see O'Mara et al., 2006, for a meta-analytic overview), it is unclear (a) how these findings generalize to a regular classroom situation without any specifically reinforced activities to enhance learners' self-concept and (b) how quickly effects on self-concept can be observed there. In particular, when learners face a relatively new domain, situation-specific self-concept measures gain importance because learners have presumably not yet developed a stable (i.e., dispositional and more trait-like) evaluation of their own ability in the respective domain (Suls & Mullen, 1982; Wigfield et al., 2015).

The aforementioned shortcomings of experimental research on the BFLPE and its associated social comparisons are widely acknowledged. However, these challenges have been difficult to overcome in the absence of a methodological approach that allows for both experimental control and an authentic classroom situation. To address questions about causality and to uncover the mechanisms that underlie the BFLPE, there is a need for an ecologically valid research tool that permits the isolation and systematic variation of relevant variables (i.e., actual performance-related classroom behavior) while simultaneously preserving the authenticity and realism of a classroom situation. Immersive virtual reality allows for exactly these things: realistic simulations and experimental designs (Blascovich et al., 2002; Parsons, 2015).

Immersive Virtual Reality as an Experimental Tool to Establish Causality and Uncover Underlying Processes

Immersive virtual reality (IVR) presents a promising avenue for experimental research as it allows researchers to integrate experimental control and authenticity in their research designs. This combines the advantages of field research with the affordances of lab studies and makes it possible to establish causal relationships and gain more in-depth insights into their underlying processes. In general, IVR offers a computer-generated simulated environment that allows for realistic perceptions and seemingly real interactions in an artificial virtual world (e.g., through head-mounted displays; Blascovich et al., 2002). Highly immersive VR systems are able to place users into surroundings that give the illusion of reality, which can consequently be vividly experienced (Cummings & Bailenson, 2016; Makransky & Lilleholt, 2018). Particularly in recent years, fast-paced technical developments in the field of software and hardware development have led to extremely realistic simulations (e.g., Slater & Sanchez-

Vives, 2016). The application of IVR began in the gaming sector and was quickly adapted by the military and health sectors, particularly for practicing the handling of situations that can hardly or cannot be simulated in real life (e.g., Blascovich et al., 2002; Carl et al., 2019; Parsons & Mitchell, 2002; Richards, 2017). Similarly, educational psychologists have started to use IVR as a tool for teacher training (Dieker et al., 2007; Lugrin et al., 2016). Beyond its application for training purposes, more and more disciplines have begun to recognize and use the enormous potential of virtual reality as a methodological tool (e.g., Lanier et al., 2019).

In order to use IVR as an ecologically valid experimental tool, users must feel like they are physically and mentally immersed in the virtual environment (Bailenson et al., 2008; Blascovich et al., 2002). The corresponding impression of IVR environments as “places visited rather than as images seen” (Slater & Wilbur, 1997, p. 4) is described as *presence*. Presence is determined by the degree of *immersion* (i.e., shutting out the outside world) and *realism* (i.e., accurate representations of objects, events, and people) in an IVR environment (Lombard et al., 2009; Slater & Wilbur, 1997). In general, the use of VR as a research tool has proven to be suitable and promising, especially in younger subjects: Children are more responsive to IVR environments than adults, both cognitively and behaviorally (Bailey & Bailenson, 2017). They perceive the simulations as more real and feel a higher presence, which makes them act more spontaneously while thinking less about the world outside the IVR (Bailenson et al., 2008; Bailey & Bailenson, 2017).

IVR is particularly interesting as it offers an experimental approach to systematic and in-depth examinations of basic pedagogical-psychological theories (Bailenson et al., 2008; Kizilcec et al., 2015). IVRs have successfully been used as an experimental tool for bridging the gap between fully controllable designs and ecologically valid and authentic environments (Bailenson et al., 2008; Blume et al., 2019). Most important for the present study, there is a recent IVR study in the field of economics that investigated social comparisons in the workplace (Bönsch et al., 2017; Gürerk et al., 2019). The authors were able to show that the simulation of social comparisons and associated effects in IVR environments worked for the adult participants. Results did not consider self-concept as an outcome but showed that the presence of a more powerful virtual co-worker compared with a weaker counterpart had an effect on participants’ productivity, with a particularly strong effect on participants with low initial performance.

The Present Study

In the present study, we examined how individual differences in students' academic self-concepts emerge and can be traced back to social comparison processes in the classroom. More specifically, we investigated whether and to what extent students use implicit performance-related information in the classroom (i.e., classmates' behavior rather than explicit performance feedback) as a source of social comparisons and how these processes explain individual differences in students' self-concepts. To gain systematic insights into the BFLPE and social comparisons as its underlying mechanism, we used a research design that allowed for the strict experimental control of social comparison information and yet preserved the authenticity of a real classroom situation. In doing so, the present study is—to our knowledge—the first experimental investigation of the BFLPE that was not restrained by the typical limitations that lab studies or research in authentic learning environments have to face.

We used a fully immersive VR classroom environment in which students learned the basic principles of computational thinking, and we implemented different scenarios in the IVR to systematically vary students' performance levels in terms of actual classroom behavior. More specifically, we used students' hand-raising behavior as an indicator of performance (e.g., Böheim et al., 2020) and manipulated the proportions of students in the class who raised their hands to answer the teachers' questions or otherwise indicated that they knew the correct solution to a task. The IVR setting made it possible for every participant to experience the exact same lesson on the topic of computational thinking where the only difference was the performance of virtual classmates. Based on the experimental manipulation, the study aimed to test the typically found reference-group effects (i.e., the BFLPE) on individuals' self-concept in a controlled and authentic environment. Further extending previous BFLPE research, we investigated not only effects on criterion-oriented and dispositional domain-specific self-concept but particularly effects on more situation-specific self-evaluations. We examined *socially-oriented* self-concept (a situation-specific measure referring to the reference group that was in place) and *criterion-oriented* self-concept (referring to situation-specific task criteria and performance standards) as well as *dispositional* self-concept in the domain of computational thinking. We aimed to use the IVR setting to gain insights into the emergence of effects of the reference group on self-concept, which we expected to begin with situation-specific effects on socially-oriented self-concept. We furthermore examined whether a 15-min exposure to more or less high-performing virtual classmates would (already) have an impact on criterion-oriented and dispositional self-concept. Therefore, we investigated whether and to what extent learners actively compared themselves with their classmates (rather than relying on explicit performance feedback from their teacher). We examined how students perceived implicit performance-related social comparison information (i.e., the hand-raising behavior of their virtual classmates) and how this perception was related to effects on the different self-concepts.

In a first step, we asked whether and to what extent students processed implicit performance-related social comparison information (i.e., hand-raising behavior) and used this information as an indicator of classmates' performance. We examined the extent to which the hand-raising behavior was predictive of the reported number of registered hand-raising students. Moreover, we investigated whether students obtained different perceptions of the class' performance level based solely on the manipulated hand-raising behavior as an implicit performance indicator. We expected that the different proportions of high-performing students would positively predict (a) the reported number of registered hand-raising students and (b) the perceived performance level of the class. We hypothesized:

> The more virtual students raise their hands to answer the teachers' questions or indicate that they know the correct solution to a task, (a) the higher the reported number of registered hand-raising students will be and (b) the higher the perceived performance level of the class will be (Hypothesis 1).

In a second step, we asked about the impact of classmates' hand-raising behavior on individual learners' self-concepts. Considering the situation-sensitive nature of self-concept, we investigated *socially-oriented* self-concept as well as *criterion-oriented* self-concept and *dispositional* self-concept in the domain of computational thinking as outcome measures. In line with the typical results of BFLPE research, we expected a negative effect of hand-raising behavior on learners' self-concepts. We hypothesized:

> The higher the proportion of virtual students raising their hands (i.e., showing high-performing behavior), the lower the self-concepts of participating students will be (Hypothesis 2). We expected the effect to be particularly pronounced for *socially-oriented* self-concept. We furthermore explored the effect of the 15-min-long exposure to high-achieving classmates on *criterion-oriented* and *dispositional* self-concept.

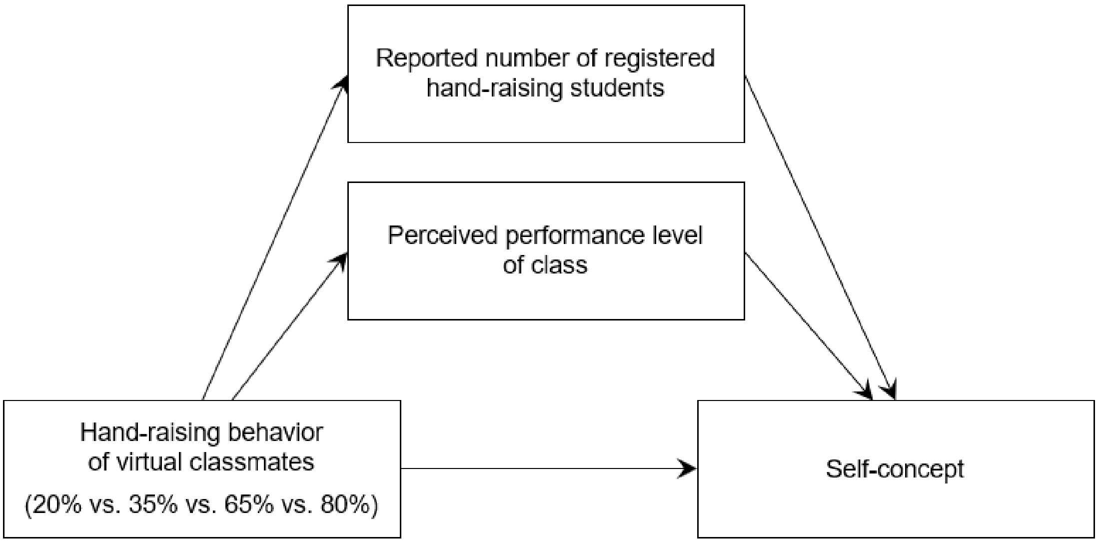
Finally, provided that an effect of the hand-raising conditions on one of the self-concepts was found, we aimed to gain more insights into the mechanism underlying the BFLPE. More specifically, we examined whether the observed effect of the different hand-raising behaviors on students' self-concept would—as suggested by the BFLPE—be explained by students' actual social comparison processes. We argue that the experimental conditions should lead to differences in (a) the reported number of registered hand-raising students and (b) the perceptions of the class' overall performance level (see Hypothesis 1) and that these perceptions in turn should affect learners' self-concept. Hence, we expected that the reported number of registered hand-raising students and the perceived performance level would fully mediate the relationship between the proportion of high-achieving classmates and the effects on students' self-concept. We hypothesized:

> The higher the proportion of virtual students who raise their hands and show a high-performing profile during the IVR lesson, (a) the higher the reported number of registered hand-raising students and (b) the higher the perceived performance level, which in turn will negatively predict students' self-concept (Hypothesis 3). In line with Hypothesis 2, we expected this effect to be particularly pronounced for *socially-oriented* self-concept as an outcome.

Figure 1 shows the hypothesized model.

Figure 1

Hypothesized Mediation Model



Note. The experimental conditions (i.e., hand-raising behavior of virtual classmates on four levels) are the independent variables, the reported number of registered hand-raising students and the perceived performance level of the class are potential mediators, and self-concept is the dependent variable. Self-concept stands for three distinct self-concept measures for assessing socially-oriented self-concept, criterion-oriented self-concept, and dispositional self-concept in the domain of computational thinking.

Method

The study and data collection were approved by regional educational authorities and the ethics committee of the University of Tübingen who confirmed that the procedures were in line with ethical standards of research on human subjects (date of approval: 11/25/2019, file number: A2.5.4-106_aa).

Sample

Data were collected in the beginning of 2020, and the recruited sample consisted of $N = 381$ students in Grade 6. Data from 28 participants had to be excluded due to technical issues during data collection (i.e., mostly the unexpected crashing of HMDs and computers or audio issues in the middle of the IVR experience). The cleaned sample consisted of $N = 353$ students ($M_{Age} = 11.52$ years, $SD_{Age} = 0.55$; 46.7% girls) from a total of 25 classes at 14 different schools. With regard to the age of participants, sixth-graders were considered particularly suitable because effects of social comparisons can be observed beginning at 8 to 10 years of age (Dijkstra et al., 2008), when children are also particularly responsive to IVR environments (Bailenson et al., 2008; Bailey & Bailenson, 2017).

An a priori power analysis was computed to determine the required sample size, considering existing findings from experimental studies (Möller & Köller, 2001, Study 1: $d = 0.85$, Study 3: $d = 1.37$; Wolff et al., 2018, Study 1: $d = 0.73$). Considering the fact that the use of explicit performance feedback in these studies presumably overestimates the expected effects of varying reporting behavior in the present study, small to medium effects ($f = .20$) were assumed, and a necessary sample size of $N = 90$ students for each of the four conditions was determined for respective analyses of variance (for two-tailed tests with a .05 alpha level and a minimum power of .90).

Procedure

Sixth grade students were recruited from local academic track schools via e-mails and invitation letters. After obtaining written informed consent from both the students and their parents or legal guardians, all students who indicated interest were admitted into the study. The participating students were tested in groups of up to 10 with all test sessions taking place in a quiet room at the participants' school (see Figure 2 for an impression).

For each of the 10 students in a group, head-mounted displays (HMDs; HTC Vive Pro Eye) were set up prior to the test session. The set-up included selecting one of the experimental conditions, which the researchers had randomly assigned to each of the 10 seats in the testing room in advance (see the details of the IVR classroom's configuration and manipulation below). Random number generations at an individual level were used to randomly distribute the experimental conditions within and across test groups. Students were free to choose any seat when they entered the testing room without knowing about the different experimental conditions. The procedure was identical for all test sessions, as the experimental conditions differed only with respect to the specific characteristics that were manipulated in the virtual classroom situation.

Figure 2

Participating Students Tested in Groups of up to 10 (Pictures: Gabriele Loges)

In total, each test session took approximately 45 min, including all instructions and preparations. The experiment consisted of three main parts: First, the participants completed a paper-based pretest, which included demographics (i.e., age, gender) and relevant personality characteristics (i.e., social orientation, self-concept of intelligence) as well as students' learning background (i.e., grades, interest in and prior experience with the lecture topic and IVR). Second, participants experienced the IVR lesson. Up to four research assistants were present at all times to supervise the participants and help them put on the HMDs after they completed the pretest. The IVR lesson was introduced as a learning experience (i.e., without any reference to possible social comparisons with virtual classmates), and all students began the IVR lesson at the same time. Third, as soon as the participants finished the virtual learning experience, they completed a paper-based posttest questionnaire followed by a debriefing. The posttest included scales for measuring the participants' self-concepts and their experience with the IVR lesson (i.e., perceptions of their virtual classmates and the virtual class' performance level).

The Immersive VR Classroom

Contents of the IVR Lesson

The IVR lesson’s contents were adapted from tested and evaluated materials from a course that was designed to teach kids basic computational thinking skills (titled “Understanding how computers think”). Computational thinking generally describes the ability to sequence a problem or task into substeps, to formulate solution steps, and to use a computer for this purpose (Weintrop et al., 2016). Due to its central importance, computational thinking is regarded as one of the central key competencies of the 21st century, which should be introduced into formal education at an early stage (e.g., Grover & Pea, 2013). However, there are currently only a few concepts that are usually taught to implement computational thinking in the curricula of primary and lower secondary schools. Against this background, computational thinking was considered particularly suitable for the purpose of the study as it could be assumed that the participating students had little to no prior knowledge and corresponding learning experience with this subject matter. Consequently, social comparison effects could be investigated largely independently of students’ previous experiences, thus offering a way to look at the genesis of differences in self-concept in the field of computational thinking.

In the 15-min IVR classroom experience, the students learn about the meaning of coding and sequences and loops as basic computational concepts. At the beginning of the lesson, the teacher gives a short introduction to the topic and asks a number of open questions such as "Who can explain what coding means?" "Who has heard the word sequence before?" or "Who knows an example of something in everyday life that works like a loop?" This is followed by two exercises in which the use of the concepts "sequence" and "loop" is tested (adapted from the Computational Thinking test by Román-González et al., 2017). The students have to choose the correct answer from four options, and after each task, the teacher checks the class' performance by going through the answer options and asking who thinks each of the options is the correct answer. Eventually, the lesson sequence concludes with a brief summary by the teacher.

Configuration and Manipulation of the IVR Classroom

The whole IVR learning experience was situated in a simulated classroom. Figure 3 shows the design of the virtual classroom (for a preview stream of the events in the virtual classroom, see https://osf.io/jb8vq/?view_only=ce786eb760604c1ca9f388a8515f5d91). The IVR lesson was a fully preprogrammed simulation of a typical teaching situation that used audio recordings and motion captures stemming from a real classroom to ensure that the pace and content of the virtual students' answers as well as their movements would be calibrated to reflect those typical of a sixth grader. Graphical representations of the virtual classmates and the teacher were designed by considering the Uncanny Valley effect (Mori et al., 2012) and aiming to capture an appropriate degree of (behavioral) realism (Bailenson et al., 2004; Guadagno et al., 2007).[17]

Figure 3

Virtual Classroom Situation with Different Proportions of Hand-Raising Students

Note. The top shows a situation with 20% of the virtual classmates raising their hands, whereas the bottom image comes from the 80% condition. For a preview stream of the IVR lesson, see https://osf.io/jb8vq/?view_only=ce786eb760604c1ca9f388a8515f5d91.

Participating students were told they would join another class of sixth graders in the IVR classroom as soon as they put on their HMDs. Participants experienced the classroom situation from the perspective of a student sitting in the virtual classroom surrounded by 24 virtual classmates, with the tables arranged in the typical parallel-rows set-up that can be found

[17] IVR environments provide a promising avenue to address the lack of replication studies in social and educational psychology (e.g., Blascovich et al., 2002). Researchers who are interested in using the IVR classroom used in the present study to replicate our findings are asked to contact the corresponding author for further details.

in most German secondary schools. Participants were placed at a desk that was identical to their position in the IVR classroom. Thus, the real world was used to provide haptic feedback congruent to the virtual world. Before the lesson in the virtual classroom began, the teacher entered the room and announced that she would be back in a minute, giving participants the opportunity to adjust to the IVR classroom environment. Throughout the lesson, the teacher stood at the front of the virtual classroom and worked with a blackboard and a video screen. The IVR simulation consisted of explanations by the teacher, dialogue between the teacher and the virtual students, time to work on exercises, and a discussion of solutions afterward. Participants were asked not to walk around in the virtual classroom and to remain seated and quiet, but they could engage in any other activities to explore the virtual environment (e.g., look around, raise their hands), and they were instructed to behave as they would in a normal classroom situation.

When the teacher interacted with the virtual students, the experimental groups differed in terms of the proportion of students who raised their hands in response to the teachers' questions or who indicated that they knew the correct solution to a task. The students' answers were always correct, which was accordingly communicated by the teacher. The proportion of virtual classmates who responded to the teachers' questions and showed high-performing participation in the lesson was manipulated on four levels with 20% vs. 35% vs. 65% vs. 80% hand-raising students (see Figure 3 for an impression of the two extreme conditions). We chose these four levels to ensure that we had (a) an effective study design with a limited number of conditions and required participants so that we would have sufficient power and (b) a differentiated picture of when aversive versus positive effects appeared. Hence, there was a relatively fine-grained difference between 20% and 35% as well as between 65% and 80%, whereas there was a larger difference between 35% and 65% to ensure differentiated grading and yet unambiguous information about whether the percentage of students who were high-achieving was below versus above average. Participants were randomly assigned to one of the experimental conditions. [18]

[18] In addition to the manipulation of virtual classmates' hand-raising behavior, there were two other variables that were manipulated in the present IVR experience, but they were not relevant to the current study. The locations of participating students' seats in the IVR classroom and the virtual classmates' and teacher's graphical representations were systematically varied on two levels each. Participants were randomly allocated to a seat in either the second or fourth row of the virtual classroom. Virtual classmates and the teacher were presented in either a more cartoon-like manner or more realistically. Participants' allocation to the different configuration conditions was counterbalanced also with respect to the main manipulation conditions of different hand-raising behaviors.

Participants' Experience in the IVR Classroom: Presence and Realism

The IVR environment was designed to ensure an authentic classroom experience for participants. We checked participants' experienced presence and perceived realism via self-reports. Participants were asked to rate their experienced level of presence in the IVR classroom with nine items (e.g., "I felt like I was sitting in the virtual classroom" or "I felt like the teacher in the virtual classroom really addressed me"). The items were based on common conceptualizations of presence by Schubert et al. (2001) and (Lombard et al., 2009). Perceived realism was assessed with six items (e.g., "What I experienced in the virtual classroom could also happen in a real classroom" or "The students in the virtual classroom behaved similarly to real classmates") that were developed specifically to assess students' perceptions of the IVR classroom situation. Both experienced presence and perceived realism were rated on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*) and had acceptable Cronbach's alpha values of .77 and .78, respectively.

Notably, the self-reports indicated an overall authentic experience of the IVR environment. Reported mean levels of experienced presence and perceived realism ranged from 2.82 to 2.97 ($0.52 < SDs < 0.62$) in all configuration conditions. The different conditions had no statistically significant effects on participants' experienced level of presence and their perceived realism of the IVR classroom (all $ps > .05$).

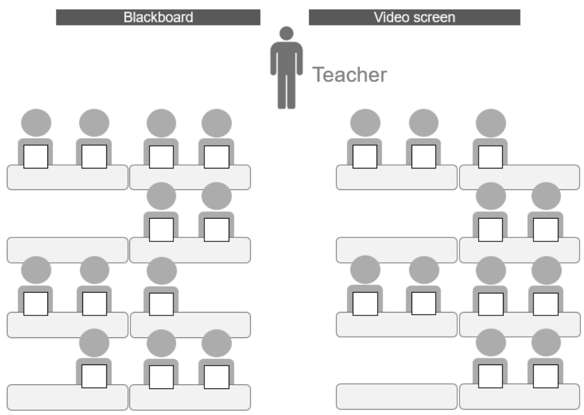
Measures

Perceived Performance Level

The posttest questionnaire administered after the IVR experience included two measures of participants' perceptions of their classmates' performance-related behavior. First, participants had to indicate how many classmates responded to the teacher's questions and raised their hands to indicate that they knew the correct answers. The reported number of registered hand-raising students was assessed with a multiple-choice question in which participants were presented with the seating plan of the IVR classroom and had to mark all the students who raised their hands (see Figure 4 for the response format). Second, participants were asked to rate the perceived performance level of the virtual class via five items (e.g., "There were many good students in the class" or "The class would do quite well on a test on the topic of the lesson"; see Appendix A for all items). These items reflected the systematically varied characteristics of the VR classrooms and were developed specifically to assess students' perception of the VR classroom. The five items were rated on the a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*), and the scale had a Cronbach's α value of .87.

Figure 4

Question Format to Assess the Reported Number of Students Who Raised Their Hands



Note. Participants were instructed as follows: "What is your estimate of how many of your virtual classmates correctly responded to the teacher's questions and indicated that they knew the right solutions? The picture below shows the virtual classroom you were just in. Mark all the students who responded to the teacher's questions and raised their hands to give the correct solutions in the picture below. If you don't remember the exact students, just mark the number of students."

Self-Concept

Learners' self-concept in the domain of computational thinking was assessed at the beginning of the posttest questionnaire. In order to obtain a differentiated picture of effects on individuals' self-concept, we administered three distinct self-concept scales with four to six items each: socially-oriented self-concept (SC1), criterion-oriented self-concept (SC2), and dispositional self-concept (SC3) in the domain of computational thinking. All self-concept scales used a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*), and the items were based on the commonly used wording by Schwanzer et al. (2005) and adapted for situation- and domain-specificity (see Appendix A for all items).

Socially-oriented self-concept (SC1) assessed situational self-concept, tailored to the situation in the VR and focusing in particular on social comparisons with the virtual classmates. The four-item scale consisted of four items (e.g., "Compared with the others, I was really good at giving the robots the right commands") of which two were reverse-scored and recoded accordingly.

Criterion-oriented self-concept (SC2) assessed the more normative dimension of situational self-concept via four items. Unlike the items for assessing socially-oriented self-concept (SC1), the items on the criterion-oriented self-concept scale were formulated with respect to the learning materials and tasks without an explicit reference to the social comparison targets in the virtual classroom (e.g., “I could have easily explained the solutions to the robot tasks”). Two of the items were reverse-scored and recoded accordingly.

Finally, the third self-concept scale (SC3) measured *dispositional self-concept in the domain of computational thinking* with six items (e.g., “I am good at breaking a complicated problem down into smaller steps” or “I am usually quick to understand topics related to technology, robots, computers, programming, etc.”). The items covered the core competencies associated with computational thinking (Grover & Pea, 2013; Román-González et al., 2017; Weintrop et al., 2016). Three of the items on the SC3 scale were reverse-scored and recoded accordingly.

All self-concept scales had acceptable Cronbach’s α coefficients of .71 (SC1), .77 (SC2), and .73 (SC3).

Covariates

The pretest questionnaire included demographics and the dispositional personality traits that are considered relevant in the context of social comparisons, such as learners’ self-concept of intelligence and their social orientation. Self-concept of intelligence was measured with four items (e.g., “I often think I'm not as smart as the others”) from the scale by Schwanzer et al. (2005). Social orientation was assessed with seven items (e.g., “I pay close attention to how I do things compared with my classmates”) adapted from Gibbons and Buunk (1999). Both scales were rated on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*) and had acceptable Cronbach’s α values of .72 and .74, respectively. The pretest also asked participants about any prior experience with the lesson’s content. Prior experience with computational thinking was assessed via a dichotomous variable (i.e., “Have you ever attended a course on programming [for example as a working group in your school or in extracurricular activities]?”). Finally, the pretest included a measure of participants’ IVR experience, asking them about their prior IVR use (i.e., “Have you ever used virtual reality [VR] glasses?”) that participants had to rate as 0 (*never*), 1 (*once*), or 2 (*more often*).

Statistical Analyses

In all analyses, hand-raising behavior as the experimental manipulation on four levels was included in all regression analyses via three dummy variables. Due to practical reasons during data collection, the testing groups always consisted of students who belonged to one school. To account for this, we controlled for cluster effects by using a school variable in all analyses (number of clusters $N = 14$). All models were estimated in Mplus 8.2, using full information maximum likelihood estimation for missing values (Muthén & Muthén, 1998-2017).[19] We posted all data and data analysis scripts on the Open Science Framework under the following link: https://osf.io/jb8vq/?view_only=ce786eb760604c1ca9f388a8515f5d91.

To examine whether classmates' different hand-raising behaviors predicted the perceived performance level of the class and had an effect on students' self-concepts, we computed multiple linear regression analyses with the experimental conditions (i.e., different proportions of hand-raising virtual classmates) as the independent variable and the perceived performance level and self-concept as the dependent variables. In the next step, we included the perceived performance level in the regression model as an additional predictor (next to hand-raising behavior) to test whether it mediated the effect of hand-raising behavior on self-concept as the dependent variable. To test the indirect effect, we used 10,000 bootstrapped samples and 95% confidence intervals (see recommendations by MacKinnon et al., 2002; Preacher & Kelley, 2011). Finally, following suggestions by Mayer et al. (2014), who argued that potential confounds should also be considered in strictly randomized research designs, we added a number of background variables to the regression model as a robustness check of the examined regression models.

Based on strong theoretical and empirical evidence, all hypotheses for the perception of the manipulated hand-raising behavior (i.e., number of registered hand-raising students and perceived performance level of the class) and regarding the most proximal self-concept measure of socially-oriented self-concept were directional and were thus tested with one-tailed tests. All remaining hypotheses on the effects on criterion-oriented self-concept and dispositional self-concept were tested with two-tailed tests as the present study aimed to explore whether and to what extent a 15-min lesson would impact these outcomes. We used a critical *p*-value and confidence intervals set at an alpha level of .05 for all hypothesis tests, and we report standardized regression coefficients for all regression analyses. To account for the fact that we computed a number of significance tests, we applied the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995, 2000) to avoid the accumulation of Type I errors.

[19] All analyses were calculated a second time by additionally including the different configuration conditions (i.e., seat position [back vs. front] and graphical representation [cartoon vs. more realistic] of virtual characters) as covariates. The results were very similar, suggesting that seat position and graphical representation had no impact on the relations between the hand-raising behavior of virtual classmates, the perceived performance level of the class, and individuals' self-concept. For reasons of simplicity and clarity, the results reported in this paper do not include the configuration conditions as covariates, but detailed results are included in the supplementary material (see Tables S2–S7; supplemental materials).

Results

Table 1 presents descriptive statistics for the basic sample characteristic and all the covariates after participants were randomized to one of the conditions. There were no statistically significant between-group differences with respect to any of the variables measured before the students experienced the IVR classroom situation.

Do Different Variations of Hand-Raising Behavior Lead to Distinct Perceptions of the Class' Performance Level?

The first research question asked whether and to what extent students process implicit performance-related social comparison information (i.e., hand-raising behavior, rather than explicit performance feedback) and use this information as an indicator of their classmates' performance. To address this, we tested (a) whether students recognized their classmates' hand-raising (i.e., whether they actively processed the available social comparison information) and (b) whether students obtained different perceptions of the class' performance level on the basis of the manipulated hand-raising behavior (i.e., whether they used the available social comparison information as an indicator of performance).

Speaking to the number of registered hand-raising students, as intended, descriptive statistics (see Table 2) showed a continuously increasing mean value across Conditions 1 to 4. The number of registered hand-raising students did not reflect the exact proportion of hand-raising students, which was underestimated in all but the first (20%) condition. However, the results of multiple linear regressions indicated that the experimental condition (i.e., the proportion of hand-raising students) statistically significantly predicted the number of registered hand-raising students. As presented in Table 3, there was no statistically significant effect of Condition 2 (35% hand-raising students; $M = 7.68$, $SD = 3.68$) compared with Condition 1 (20% hand-raising students; $M = 6.70$, $SD = 3.74$), but all other conditions led to statistically significantly different numbers of registered hand-raising students (Condition 3 with 65% hand-raising students, $M = 10.12$, $SD = 4.85$; Condition 4 with 80% hand-raising students, $M = 13.58$, $SD = 5.45$).

Regarding the perceptions of different hand-raising behavior, there was a statistically significant positive correlation between the number of registered hand-raising students and the perceived performance level of the class, $r(333) = .30$, $p < .001$. We consequently investigated whether the proportion of hand-raising students (i.e., the experimental manipulation of the number of high-achievers in a class) predicted the perceived performance level of the class. As Table 3 shows, the hand-raising conditions had a statistically significant positive effect on the perceived performance level. Conditions 1 (20% hand-raising students) and 2 (35% hand-raising students) did not result in statistically significantly different perceptions of the class' performance level ($b = 0.06$, $SE = 0.05$, $p = .099$). However, compared with the first condition (20% hand-raising students), the perceived performance level was statistically significantly higher for Condition 3 with 65% hand-raising students ($b = 0.25$, $SE = 0.04$, $p < .001$, $d = 0.60$) as well for Condition 4 with 80% high-achieving students ($b = 0.45$, $SE = 0.07$, $p < .001$, $d = 1.24$). Descriptive statistics are presented in Table 2.

Table 1

Descriptive Statistics after Randomization to one of the Conditions

Variable	20% Hand-raising (N = 92)	35% Hand-raising (N = 86)	65% Hand-raising (N = 85)	80% Hand-raising (N = 90)	Statistics
Age	11.43(0.52)	11.51(0.55)	11.55(.63)	11.56(0.52)	$F(3,349) = 1.04, p = .377$
Gender					
Female	43	43	39	40	$\chi^2(3) = 0.583, p = .900$
Male	49	43	46	50	
Grades [a]					
Mathematics	2.60(0.87)	2.62(0.83)	2.70(0.91)	2.57(0.87)	$F(3,331) = 0.34, p = .795$
German	2.47(0.62)	2.52(0.67)	2.55(0.76)	2.45(0.79)	$F(3,332) = 0.31, p = .815$
BNT	2.28(0.64)	2.31(0.71)	2.28(0.72)	2.31(0.70)	$F(3,331) = 0.05, p = .986$
Prior CT experience [b]					
No	63	66	62	70	$\chi^2(9) = 10.65, p = .300$
Yes	28	19	22	18	
n/a	1	1	1	2	
Prior IVR experience [c]					
None	43	27	39	33	$\chi^2(6) = 7.640, p = .266$
Once	33	38	27	29	
Often	16	20	19	24	
n/a	0	1	0	4	
Self-concept of intelligence [d]	3.06(0.58)	2.95(0.60)	3.14(0.58)	3.06(0.62)	$F(3,349) = 1.42, p = .237$
Social orientation [d]	2.49(0.52)	2.45(0.52)	2.48(0.52)	2.40(0.55)	$F(3,346) = 0.55, p = .648$

Note. Mean values and standard deviations M(SD) are shown for continuous variables, categorical variables are shown in absolute numbers; One-way between-subjects ANOVAs were calculated to compare the mean values in the conditions; Chi-square tests (two-tailed) were used to compare the distributions of the categorical descriptive variables. CT = Computational Thinking; IVR = Immersive Virtual Reality.

[a] Grades were on a scale from 1–6 with lower numbers indicating better achievement; [b] Prior CT experience was assessed via a dichotomous variable asking whether one had previously attended a course on programming; [c] Prior IVR experience was assessed via one item asking about whether one had previously used IVR glasses; [d] Self-concept of intelligence and social orientation are denoted by the mean of four items measured on a 4-point rating scale with higher values indicating higher levels.

Table 2

Descriptive Statistics for the Number of Registered Hand-Raising Students, Perceived Performance Level, and Mean Self-Concept Values in the Different Hand-Raising Conditions

	Registered hand-raising students			Perceived class performance level		Socially-oriented self-concept (SC1)		Criterion-oriented self-concept (SC2)		Dispositional self-concept (SC3)	
	N	*M (SD)*	Actual	*N*	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*
20% Hand-raising	88	6.70 (3.74)	5	89	2.87 (0.51)	92	3.46 (0.53)	92	3.55 (0.57)	92	3.17 (0.48)
35% Hand-raising	85	7.68 (3.68)	9	85	2.96 (0.59)	86	3.44 (0.53)	86	3.50 (0.57)	86	3.09 (0.51)
65% Hand-raising	81	10.12 (4.85)	16	82	3.21 (0.55)	85	3.41 (0.48)	85	3.61 (0.46)	83	3.20 (0.45)
80% Hand-raising	84	13.58 (5.45)	20	88	3.46 (0.43)	90	3.31 (0.61)	90	3.55 (0.47)	90	3.16 (0.44)

Note. The possible answer range for the registered hand-raising students was between 0 and 24 (see question format in Figure 4). Actual = the actual number of hand-raising students (i.e., the true value of how many students raised their hands in the respective conditions).

Table 3

Standardized Regression Coefficients for the Effect of Different Hand-Raising Conditions on the Number of Registered Hand-Raising Students and the Perceived Performance Level of the Class

	Number of registered hand-raising students					Perceived performance level of class				
	b	*SE*	*t*	*p*	95% CI	*b*	*SE*	*t*	*p*	95% CI
(Intercept)	0.01	0.02	0.36	.721	[-0.05, 0.05]	0.00	0.06	0.01	.999	[-0.11, 0.11]
20% vs. 35% Hand-raising	0.08	0.05	1.49	.069	[-0.03, 0.19]	0.06	0.05	1.29	.099	[-0.03, 0.16]
20% vs. 65% Hand-raising	0.28	0.07	4.19	**<.001**	[0.15, 0.41]	0.25	0.04	6.22	**<.001**	[0.17, 0.33]
20% vs. 80% Hand-raising	0.58	0.05	10.63	**<.001**	[0.47, 0.68]	0.45	0.07	6.78	**<.001**	[0.32, 0.57]
35% vs. 65% Hand-raising	0.20	0.08	2.70	**<.001**	[0.06, 0.35]	0.19	0.04	5.06	**<.001**	[0.12, 0.26]
35% vs. 80% Hand-raising	0.50	0.05	9.79	**<.001**	[0.40, 0.60]	0.38	0.06	6.18	**<.001**	[0.26, 0.50]
65% vs. 80% Hand-raising	0.29	0.06	4.38	**<.001**	[0.16, 0.42]	0.19	0.05	3.84	**<.001**	[0.09, 0.29]

Note. The Benjamini-Hochberg correction was used to control the false-positive rate in the face of multiple tests (Benjamini & Hochberg, 1995). Statistically significant results ($p < .05$) after the correction are presented in bold. CI = Confidence interval.

Do the Variations in Hand-Raising Behavior Affect Learners' Self-Concept?

The second research question asked about the impact of classmates' hand-raising behavior on individual learners' self-concepts. The effects on self-concepts were tested separately for the three self-concept measures. Descriptive statistics (see Table 2) showed that the mean values for socially-oriented self-concept in the different conditions continuously decreased as the number of hand-raising students increased from Condition 1 (20% hand-raising students) to Condition 4 (80% hand-raising students).

Table 4

Standardized Regression Coefficients for the Effects of Different Hand-Raising Conditions on Socially-oriented and Criterion-oriented as well as Dispositional Self-Concept in the Domain of Computational Thinking

	b	*SE*	*t*	*p*	95% CI
Model 1: Socially-oriented self-concept (SC 1)					
(Intercept)	0.00	0.10	0.00	1.000	[-0.20, 0.20]
20% vs. 35% Hand-raising	-0.02	0.06	-0.28	.390	[-0.12, 0.09]
20% vs. 65% Hand-raising	-0.04	0.06	-0.67	.252	[-0.16, 0.08]
20% vs. 80% Hand-raising	-0.12	0.05	-2.18	**.015**	[-0.22, -0.01]
Model 2: Criterion-oriented self-concept (SC 2)					
(Intercept)	0.00	0.08	0.00	1.000	[-0.16, 0.16]
20% vs. 35% Hand-raising	-0.04	0.03	-1.22	.223	[-0.09, 0.02]
20% vs. 65% Hand-raising	0.05	0.05	0.97	.333	[-0.05, 0.15]
20% vs. 80% Hand-raising	0.00	0.07	0.02	.981	[-0.13, 0.13]
Model 3: Dispositional self-concept (SC 3)					
(Intercept)	0.00	0.07	0.01	.994	[-0.14, 0.14]
20% vs. 35% Hand-raising	-0.07	0.04	-1.76	.078	[-0.16, 0.01]
20% vs. 65% Hand-raising	0.03	0.08	0.33	.739	[-0.14, 0.19]
20% vs. 80% Hand-raising	-0.01	0.07	-0.13	.894	[-0.14, 0.12]

Note. The Benjamini-Hochberg correction was used to control the false-positive rate in the face of multiple tests (Benjamini & Hochberg, 1995). Statistically significant results ($p < .05$) after the correction are presented in bold. CI = Confidence interval.

As presented in Table 4, Conditions 1 (20% hand-raising students) and 2 (35% hand-raising students) did not result in any statistically significant differences for socially-oriented self-concept (SC1; $b = -0.02$, $SE = 0.06$, $p = .390$). Similarly, Condition 3 (65% hand-raising students) did not predict any differences in socially-oriented self-concept compared with Condition 1 (20% hand-raising students; $b = -0.04$, $SE = 0.06$, $p = .252$). However, there was a statistically significant difference in the two extreme hand-raising conditions when predicting socially-oriented self-concept: The 80% condition had a statistically significant negative effect on the mean level of socially-oriented self-concept (SC1) compared with the 20% condition (b = -0.12, SE = 0.05, p = .015, d = 0.30). As Table 4 shows, the hand-raising conditions were not associated with statistically significant differences in the criterion-oriented self-concept (SC2) and the dispositional self-concept (SC3) measures (20% vs. 80% hand-raising students; b = 0.00, SE = 0.07, p = .981 and b = -0.01, SE = 0.07, p = .894 for SC2 and SC3, respectively).

Do Different Perceptions of the Class' Performance Levels Affect Learners' Self-Concept?

The third research question asked whether perceptions of the hand-raising behavior (i.e., the number of registered students and the perceived performance level of the class) had an effect on learners' self-concepts and if so, how students' perceptions of the different hand-raising conditions affected the relationship between the hand-raising conditions and learners' self-concept. First, we tested for whether the different perceptions of the hand-raising conditions affected learners' self-concepts. Second, when we found a statistically significant effect of perceptions on self-concept, we tested whether these perceptions mediated the relationship between the proportion of high-achieving classmates and the effects on students' self-concept.

The number of registered hand-raising students did not predict any statistically significant differences in socially-oriented self-concept (SC1; b = -0.05, SE = 0.04, p = .122). Moreover, results revealed no statistically significant associations between the number of registered hand-raising students and criterion-oriented self-concept (SC2; b = 0.07, SE = 0.05, p = .193) or dispositional self-concept (SC3; b = 0.06, SE = 0.07, p = .423). Similarly, the perception of the hand-raising conditions (i.e., perceived performance level of the class) did not lead to any statistically significant differences in criterion-oriented self-concept (SC2; b = -0.53, SE = 0.05, p = .360) or dispositional self-concept (SC3; b = -0.03, SE = 0.04, p = .426). However, the perceived performance level was significantly negatively related to socially-oriented self-concept (SC1; b = -0.19, SE = 0.05, p < .001), showing that students who perceived the class' performance level as higher reported lower levels of socially-oriented self-concept.

On the basis of this pattern of results and the statistically significant effect of the hand-raising conditions on socially-oriented self-concept (see Research Question 2), we added the perceived performance level to the regression models as a mediator with the hand-raising conditions as the independent variable and socially-oriented self-concept (SC1) as the dependent variable. As can be seen in Table 5, including the perceived performance level in the regression model revealed a statistically significant negative association between the perceived performance level and socially-oriented self-concept (SC1; b = -0.18, SE = 0.04, p < .001). In addition, including the perceived performance level substantially reduced the direct effect of the manipulation (particularly the effect of the extreme conditions with 20% vs. 80% hand-raising) on socially-oriented self-concept (SC1) so that it became statistically nonsignificant (b = -0.04, SE = 0.06, p = .267). Results showed that the perceived performance level fully mediated the relationship between the different hand-raising behaviors of virtual classmates (20% vs. 80%) and socially-oriented self-concept (indirect effect with 10,000 bootstrapped samples, b = -0.05, SE = 0.02, p < .001, 95% CI [-0.08, -0.02]). This finding is in line with the results that were expected on the basis of the BFLPE, suggesting that when a class has a higher perceived performance level (in this case manipulated by the hand-raising behavior of classmates), this leads to lower evaluations of learners' own abilities compared with their classmates.

As a robustness check, a number of background variables (gender, grades in mathematics, German, and science, participants' prior experience with computational thinking, previous experience with IVR, self-concept of intelligence, and social orientation) were added to the regression models. Adding the covariates did not change the overall pattern of results for any of the regression models. The effects of the hand-raising conditions on the number of registered hand-raising students and on the perceived performance level (Research Question 1) remained statistically significant when the background variables were added to the model. Details on the respective statistics can be found in the supplemental materials. Most importantly, as presented in Tables 5 and 6, the effect of the different hand-raising behaviors (20% vs. 80%) on socially-oriented self-concept (Research Question 2) and the mediating effect of the perceived performance level (Research Question 3) remained statistically significant when we controlled for the background variables (b = -0.11, SE = 0.05, p = .008 and b = -0.07, SE = 0.02, p < .001 for the indirect effect, respectively).

Table 5

Standardized Regression Coefficients for Effects of the Different Hand-Raising Conditions on Socially-oriented Self-Concept Including Mediator and Background Variables

	b	*SE*	*t*	*p*	95% CI
Model 1: Hand-raising conditions on socially-oriented self-concept					
(Intercept)	0.00	0.10	0.00	1.000	[-0.20, 0.20]
20% vs. 35% Hand-raising	-0.02	0.06	-0.28	.390	[-0.12; 0.09]
20% vs. 65% Hand-raising	-0.04	0.06	-0.67	.252	[-0.16; 0.08]
20% vs. 80% Hand-raising	-0.12	0.05	-2.18	**.015**	[-0.22; -0.01]
Model 2: Including mediator perceived performance level					
(Intercept)	0.00	0.10	-0.01	.997	[-0.20, 0.20]
20% vs. 35% Hand-raising	0.00	0.06	-0.06	.477	[-0.12; 0.11]
20% vs. 65% Hand-raising	0.01	0.06	0.07	.472	[-0.12; 0.13]
20% vs. 80% Hand-raising	-0.04	0.06	-0.62	.267	[-0.15; 0.08]
Perceived performance	-0.18	0.04	-4.62	**<.001**	[-0.26, -0.11]
Model 3: Including mediator and background variables					
(Intercept)	0.00	0.09	0.01	.995	[-0.18, 0.19]
20% vs. 35% Hand-raising	0.00	0.06	0.07	.471	[-0.10; 0.11]
20% vs. 65% Hand-raising	-0.01	0.05	-0.10	.462	[-0.13; 0.12]
20% vs. 80% Hand-raising	-0.04	0.03	-0.91	.183	[-0.13; 0.05]
Perceived performance	-0.15	0.03	-5.02	**.002**	[-0.21; -0.09]
Gender [a]	0.05	0.06	0.77	.444	[-0.08, 0.17]
Math grade [b]	-0.15	0.04	-3.72	**<.001**	[-0.26; -0.07]
German grade [b]	0.16	0.04	4.33	**<.001**	[0.09; 0.23]
Science grade [b]	0.08	0.06	1.47	.141	[-0.03; 0.19]
Prior CT experience [c]	0.06	0.06	1.02	.310	[-0.06, 0.18]
Prior IVR experience [d]	0.02	0.07	0.33	.742	[-0.11; 0.16]
Self-concept intelligence [e]	0.11	0.08	1.41	.158	[-0.04; 0.26]
Social orientation [e]	0.04	0.05	0.86	.390	[-0.05; 0.13]

Note. The Benjamini-Hochberg correction was used to control the false-positive rate in the face of multiple tests (Benjamini & Hochberg, 1995). Statistically significant results ($p < .05$) after the correction are presented in bold. CI = Confidence interval; CT = Computational thinking; IVR = Immersive virtual reality.

[a] Gender female = 1, male = 2; [b] Grades were on a scale from 1–6 with lower numbers indicating better achievement; [c] Prior CT experience no = 0, yes = 1; [d] Prior IVR experience none = 0, once = 1, often = 2; [e] Assessed on a 4-point rating scale with higher values indicating higher levels.

Table 6

Standardized Regression Coefficients for the Effects of Different Hand-Raising Conditions on Socially-oriented and Criterion-oriented as well as Dispositional Self-Concept in the Domain of Computational Thinking, Including Background Variables

	b	*SE*	*t*	*p*	95% CI
Model 1: Socially-oriented self-concept (SC1)					
(Intercept)	0.00	0.09	0.01	.992	[-0.18, 0.19]
20% vs. 35% Hand-raising	-0.01	0.05	-0.08	.470	[-0.11, 0.10]
20% vs. 65% Hand-raising	-0.05	0.06	-0.72	.235	[-0.17, 0.08]
20% vs. 80% Hand-raising	-0.11	0.05	-2.40	**.008**	[-0.20, -0.02]
Gender [a]	0.06	0.07	0.88	.379	[-0.07, 0.19]
Math grade [b]	-0.15	0.05	-3.27	**.001**	[-0.24, -0.06]
German grade [b]	0.16	0.04	4.25	**<.001**	[0.08, 0.23]
Science grade [b]	0.10	0.05	1.90	.058	[0.00, 0.21]
Prior CT experience [c]	0.07	0.07	0.97	.319	[-0.06, 0.20]
Prior IVR experience [d]	0.02	0.07	0.30	.764	[-0.11, 0.15]
Self-concept intelligence [e]	0.12	0.08	1.61	.108	[-0.03, 0.28]
Social orientation [e]	0.03	0.04	0.80	.422	[-0.05, 0.12]
Model 2: Criterion-oriented self-concept (SC2)					
(Intercept)	0.00	0.07	0.02	.986	[-0.14, 0.14]
20% vs. 35% Hand-raising	-0.01	0.04	-0.32	.750	[-0.09, 0.06]
20% vs. 65% Hand-raising	0.06	0.04	1.27	.205	[-0.03, 0.14]
20% vs. 80% Hand-raising	0.02	0.07	0.26	.792	[-0.11, 0.14]
Gender [a]	-0.01	0.06	-0.12	.908	[-0.13, 0.12]
Math grade [b]	-0.18	0.07	-2.60	**.009**	[-0.31, -0.04]
German grade [b]	0.06	0.06	0.95	.340	[-0.07, 0.19]
Science grade [b]	0.14	0.08	1.74	.082	[-0.02, 0.30]
Prior CT experience [c]	0.11	0.06	1.61	.106	[-0.02, 0.23]
Prior IVR experience [d]	-0.03	0.04	-0.67	.503	[-0.11, 0.05]
Self-concept intelligence [e]	0.18	0.07	2.59	**.010**	[0.04, 0.31]
Social orientation [e]	0.05	0.05	0.93	.353	[-0.05, 0.15]
Model 3: Dispositional self-concept (SC3)					
(Intercept)	0.00	0.06	-0.02	.986	[-0.11, 0.11]
20% vs. 35% Hand-raising	-0.03	0.05	-0.72	.475	[-0.13, 0.06]
20% vs. 65% Hand-raising	0.02	0.08	0.25	.804	[-0.14, 0.17]
20% vs. 80% Hand-raising	0.01	0.06	0.18	.856	[-0.11, 0.13]
Gender [a]	0.04	0.05	0.84	.399	[-0.06, 0.14]
Math grade [b]	-0.13	0.05	-2.60	**.009**	[-0.23, -0.03]
German grade [b]	0.07	0.05	1.37	.170	[-0.03, 0.17]
Science grade [b]	0.10	0.07	1.34	.179	[-0.05, 0.24]
Prior CT experience [c]	0.19	0.04	4.88	**<.001**	[0.11, 0.27]
Prior IVR experience [d]	0.01	0.04	0.34	.735	[-0.07, 0.10]
Self-concept intelligence [e]	0.35	0.09	4.12	**<.001**	[0.18, 0.52]
Social orientation [e]	0.01	0.03	0.20	.838	[-0.06, 0.07]

Note. The Benjamini-Hochberg correction was used to control the false-positive rate in the face of multiple tests (Benjamini & Hochberg, 1995). Statistically significant results ($p < .05$) after the correction are presented in bold. CI = Confidence interval; CT = Computational thinking; IVR = Immersive virtual reality.
[a] Gender female = 1, male = 2; [b] Grades were on a scale from 1–6 with lower numbers indicating better achievement; [c] Prior CT experience no = 0, yes = 1; [d] Prior IVR experience none = 0, once = 1, often = 2; [e] Assessed on a 4-point rating scale with higher values indicating higher levels.

Discussion

The present study examined how individual differences in students' academic self-concepts emerge and how these differences can be explained by students' use of implicit performance-related information in their classmates' behavior for social comparisons. By systematically examining the BFLPE, the present study aimed to provide causal evidence that social comparison processes in classrooms are the underlying mechanism that leads to differential effects on students' academic self-concepts. Using a fully immersive virtual reality classroom, the study examined (a) whether students take notice of classmates' hand-raising behavior that is indicative of their achievement level and perceive it as an actual indicator of the class' performance level. Moreover, the study examined (b) the impact of classmates' hand-raising behavior (i.e., the proportion of high-achieving classmates) on individual learners' socially-oriented, criterion-oriented, and dispositional self-concepts, and (c) the role of the perceived performance level as a mediator of the effects of different classroom behaviors on students' aforementioned self-concepts.

Results supported the expectation that (a) hand-raising behavior predicted the perceived performance level of the class, and the more students raised their hands, the higher learners' perception of the class' performance level. In line with the expected effects of the manipulation, the findings revealed that participants recognized classmates' hand-raising behavior, and in line with the experimental conditions, they perceived it as an indicator of performance. Moreover, the results provided support for the hypothesis that (b) the differently manipulated classroom behaviors affected learners' situation-specific self-concept such that the higher the proportion of high-performing students was, the lower the socially-oriented self-concept of participating students. More specifically, the expected negative effect occurred only for *socially-oriented* self-concept and between the extreme conditions (i.e., 20% vs. 80% hand-raising) but not for *criterion-oriented* self-concept and *dispositional* self-concept or for more moderate standards of comparison. The effect is rather small yet fully in line with predictions that were based on the BFLPE. Furthermore, despite the small effects of the experimental conditions on socially-oriented self-concept, the mediating role of the reported number of registered hand-raising students and the perceived performance level of the class were also investigated. Whereas the reported number of registered hand-raising students had no statistically significant (indirect) effect on socially-oriented self-concept, results supported the assumption that (c) the perceived performance level mediated the relationship between the proportion of high-achieving classmates and the effects on students' socially-oriented self-concept: The more students raised their hands, the higher the perceived performance level in the class, and this perception in turn negatively predicted students' socially-oriented self-concept.

Corroborating Evidence for the BFLPE from an Authentic Experimental Design

Social comparisons have been considered the major underlying cause of differential effects on students' self-concept (Buunk & Gibbons, 2007; Festinger, 1954; Trautwein & Möller, 2016). A large number of studies have provided evidence that upward comparisons with higher achieving peers lead to negative effects on students' self-concept when individual achievement is controlled for—a finding prominently known as the BFLPE (Marsh, 1987; Marsh & Hau, 2003; Marsh et al., 2000; Marsh et al., 2020; Seaton et al., 2009). However, an explicit investigation of the direct role of social comparison processes for the BFLPE has been missing so far because large-scale research in school settings makes truly randomized designs impractical and ethically difficult to realize. Thus, previous BFLPE studies have mostly relied on descriptive and correlational approaches. Existing experimental social comparison studies allow for causal conclusions but typically cannot reflect the complexity of social comparisons in real-world settings (e.g., Möller & Köller, 2001) where social comparison information is often not explicitly given but rather implicitly provided and needs to be discovered by individual learners. To our knowledge, the present IVR study is the first to overcome these existing shortcomings in BFLPE research as it combined a strictly experimental design with social comparison information that stemmed from an "authentic" classroom situation. In this vein, the present study specifically contributes to research on the BFLPE as it used actual classroom behavior as social comparison information to investigate social comparison processes and associated effects among students in an authentic yet controllable way.

This new approach yielded a number of highly relevant findings. Perhaps most importantly, the study attests to the ubiquity of social comparison processes in students. Although researchers have long argued that social comparison processes are highly pertinent in everyday life (e.g., Buunk & Gibbons, 2007; Buunk & Mussweiler, 2001; Festinger, 1954), speaking to educational contexts, there has still been a question about the relative importance of internally processed, implicit social comparison information (e.g., the observable academic behavior of fellow students) versus explicitly presented social comparison information (e.g., direct teacher feedback and school grades; Lüdtke et al., 2005; Trautwein & Lüdtke, 2005). The present study cannot ultimately clarify the role of explicit comparison information as used in prior studies, but it established clear evidence for social comparison processes among students on the basis of implicit comparison information that is speculated to be one of the main drivers of the BFLPE. In fact, the study showed that students notice what is happening in the classroom and infer relevant cues for social comparisons from it. Our results showed that virtual classmates' hand-raising behavior strongly influenced the perception of the class in the IVR. We found that the experimental variations in classmates' hand-raising behavior corresponded to (a) the pure number of registered hand-raising students and (b) the perceived performance level of the class. However, only the perceived performance level was found to predict differences in socially-oriented self-concept. In other words, solely recognizing that more or fewer classmates showed a certain performance-related behavior did not affect learners' self-evaluation, but what mattered was the respective perception of this behavior as a performance indicator. Similar to the argument made by Huguet et al. (2009), this suggests that students had to actively process the available social comparison information, and above all, their perceptions thereby explained the effects on their self-evaluation. In the current study, this was all the more impressive given that the students were in a novel learning situation that required them to be attentive to the topic of the lesson and because the social comparison information was processed in a learning situation that barely exceeded 15 min.

Second, we found experimental support for the BFLPE. Whereas Dai and Rinn (2008) called into question whether social-contextual influences are the major reason for the BFLPE, the present study showed that by varying only classmates' behavior, the typical BFLPE results could be reproduced, and when the peer learners had higher performances, the students' socially-oriented self-concepts were lower. In doing so, the study presents an important step in the direction suggested by Collins (2000), who emphasized the need for more naturalistic studies that can account for individually shaped perceptions of social comparison information.

Third, we found support for the underlying social comparison mechanism as assumed in BFLPE research. More specifically, the experimental manipulation of the average achievement of fellow students statistically significantly impacted participants' perceptions of the (experimentally manipulated) class' overall performance level, and this perception in turn predicted differences in socially-oriented self-concept. The overall effects were rather small, but they allow important insights into a better understanding of the mechanisms that underlie the BFLPE. The findings provide evidence for the causal relationship between social surroundings and self-concept. Moreover, the results showed how easily the social environment impacts socially-oriented self-concept as the different perceptions of a class' performance level and respective effects on socially-oriented self-concept were based solely on a manipulation of hand-raising behavior and occurred after only 15 min of experiencing the classroom situation.

Notably, as an additional important finding, we found effects on only the most proximal self-concept measure (i.e., socially-oriented self-concept) and between the extreme standards of comparison (i.e., 20% vs. 80% high-performing classmates), whereas criterion-oriented self-concept and dispositional self-concept in the domain of computational thinking were not affected at all. The lack of an effect on dispositional self-concept is not surprising considering the results of a recent meta-analysis (O'Mara et al., 2006) that suggested that effects of self-concept interventions are mostly observable on a domain- or situation-specific level rather than general self-evaluations. In fact, considering the importance of academic self-concept and its long-term effects for individual academic trajectories (Marsh & Martin, 2011; Marsh & O'Mara, 2008; Valentine et al., 2004), it would not have been desirable for a 15-min classroom experience to have an observable effect on dispositional self-concept in a specific domain. However, based on the notion of a multifaceted self-concept and the assumption that more long-term and enduring self-perceptions are substantially shaped by single situations (e.g., Harter, 1986; Marsh & Shavelson, 1985; Suls & Mullen, 1982; Wrzus & Roberts, 2017), it can be assumed that repeatedly evoking effects on socially-oriented self-concept would eventually lead to effects on criterion-oriented self-evaluations and dispositional self-concept. In other words, the present study enabled us to determine how differences in students' self-concepts emerge as a result of social comparisons in a classroom situation. The observed differences might result in more stable differences in (criterion-oriented and dispositional) self-concept when students repeatedly experienced a classroom situation like the one in the present experiment.

Limitations and Future Directions

The present study made use of the potential of immersive virtual reality as an experimental tool for investigating social psychological processes in a standardized and yet authentic classroom situation. Ever since first proposed and tested (Bailenson et al., 2008; Blascovich et al., 2002), this potential has only scarcely been used by researchers (e.g., Blume et al., 2019; Kizilcec et al., 2015). Notably, experimental studies in general come along with many decisions that researchers need to make, and these degrees of freedom are further increased when running an IVR study. We believe IVR is a promising avenue through which to gain more in-depth systematic insights not only into social comparisons but also into other processes (in the classroom). When aiming to replicate and extend the present study's findings, researchers should consider that the IVR used in the present study was primarily designed to provide a reasonable degree of social information to ensure an authentic classroom situation with natural social comparison processes. Whereas the results indicated that this was accomplished in general, the present study also has some limitations that might be a starting point for additional future research.

First, for practical as well as economic reasons, the IVR lesson in the present study lasted for only 15 min. This is in line with suggestions from IVR usability research, particularly for first-time IVR users. However, it limits the time of the experiment and students' experiences in the classroom situation. The results of the present study suggest that 15 min were sufficient for learners to engage in the IVR scenario and to recognize all relevant cues to obtain the desired impression of the classroom situation. However, we found effects on socially-oriented self-perceptions but not on criterion-oriented self-concept or dispositional self-concept. Future research should investigate this distinction more closely in longer and repeated learning sequences. It would be interesting to see whether the effects are more pronounced after longer and/or repeated experiences as in the present study's IVR classroom.

Second, the only indicator of performance used in the present study was students' hand-raising behavior. Whereas we found that hand-raising behavior was strongly linked to the perceived performance level of the class, one needs to consider that in real classroom settings, there is a very salient evaluative atmosphere that is shaped by known peers, constant performance feedback, and events that go beyond single teacher-student interactions (Dijkstra et al., 2008; Levine, 1983; Wheeler & Suls, 2005). The rather small sizes of the effects of hand-raising behavior on socially-oriented self-concept might be due to the fact that the evaluative atmosphere in the IVR was realistic but still limited and not as salient as the atmosphere that would be found in a real classroom/school setting. Future studies should additionally examine other behavioral cues that additionally affect learners' perceptions of a class' performance level and thus lead to more pronounced differences in individuals' self-concept (e.g., the quality of students' answers or their attention-related behavior). Moreover, taking up the selective accessibility hypothesis, which suggests that individual characteristics influence the perception and processing of social comparison information (Mussweiler, 2003), the IVR classroom presents an ideal experimental design to systematically and yet authentically investigate the influence of respective individual moderating variables on the BFLPE. In addition, adding adaptive social interactivity between the virtual teacher and participating students might provide even more salient opportunities to engage in the classroom situation (e.g., Heidicker et al., 2017; Howard, 2018).

Third, the IVR lesson was designed specifically for academic track students. We selected the learning material on the topic of computational thinking as something that is not included in the curriculum of academic track schools before Grade 7, and our findings consequently apply only to our sample of students in Grade 6, who we expected to be rather unacquainted with the subject matter. However, we cannot completely rule out the possibility that some academic track students might already be familiar with some of the topics from extracurricular activities in Grade 6. For future research, it would be worthwhile to replicate the present study with less experienced students from other class levels or school types and/or even more demanding material in the IVR lesson to examine whether the observed effects would be more pronounced under these circumstances.

Conclusion

The present study used immersive virtual reality as a novel approach to test the BFLPE and to investigate associated social comparison processes. The standardized yet authentic IVR setting allowed us to provide evidence for the causality of the BFPLE and yielded important insights into the mechanisms that underlie the effect. The results indicate how ubiquitous social comparisons in the classroom are and highlight the major role of students' perceptions of their classmates when explaining differences in self-evaluations. Moreover, our findings showed how easy it is for the social environment to impact learners' socially-oriented self-concept and thus emphasize the necessity to consider the situation-specificity of self-concept when examining effects on self-evaluations. Beyond this, not only do the results of the present study provide new insights into the emergence of social comparison effects in the classroom, but they make a general contribution to the use of virtual reality in educational and social psychological research. By replicating the empirically well-supported BFLPE, the results of the present study provide support for the feasibility and validity of conducting experimental studies in an IVR classroom and thus provide the grounds for establishing IVR as a promising tool for experimental studies in educational and social psychological research. Future research should extend the use of the technical affordances of this technology.

References

Arens, A. K., Schmidt, I., & Preckel, F. (2019). Longitudinal relations among self-concept, intrinsic value, and attainment value across secondary school years in three academic domains. *Journal of Educational Psychology, 111*(4), 663–684. https://doi.org/10.1037/edu0000313

Bailenson, J., Aharoni, E., Beall, A. C., Guadagno, R. E., Dimov, A., & Blascovich, J. (2004). Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. *Proceedings of the 7th Annual International Workshop on Presence*, 1864–1105.

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences*, *17*(1), 102–141. https://doi.org/10.1080/10508400701793141

Bailey, J. O., & Bailenson, J. (2017). Immersive virtual reality and the developing child. In F. C. Blumberg & P. J. Brooks (Eds.), *Cognitive Development in Digital Contexts* (pp. 181–200). Academic Press. https://doi.org/10.1016/b978-0-12-809481-5.00009-2

Becker, M., & Neumann, M. (2016). Context-related changes in academic self concept development: On the long-term persistence of Big-Fish-Little-Pond effects. *Learning and Instruction, 45*, 31–39. https://doi.org/10.1016/j.learninstruc.2016.06.003

Becker, M., & Neumann, M. (2018). Longitudinal Big-Fish-Little-Pond effects on academic self-concept development during the transition from elementary to secondary schooling. *Journal of Educational Psychology, 110*(6), 882–897. https://doi.org/10.1037/edu0000233

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics, 25*(1), 60–83. https://doi.org/10.3102/10769986025001060

Blanton, H., Gibbons, F. X., Buunk, A. P., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology*, *76*(3), 420–430. https://doi.org/10.1037/0022-3514.76.3.420

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Blascovich, J., Mendes, W. B., Hunter, S. B., & Salomon, K. (1999). Social" facilitation" as challenge and threat. *Journal of Personality and Social Psychology*, *77*(1), 68–77. https://doi.org/10.1037/0022-3514.77.1.68

Blume, F., Göllner, R., Moeller, K., Dresler, T., Ehlis, A.-C., & Gawrilow, C. (2019). Do students learn better when seated close to the teacher? A virtual classroom study considering individual levels of inattention and hyperactivity-impulsivity. *Learning and Instruction, 61*, 138–147. https://doi.org/10.1016/j.learninstruc.2018.10.004

Böheim, R., Urdan, T., Knogler, M., & Seidel, T. (2020). Student hand-raising as an indicator of behavioral engagement and its role in classroom learning. *Contemporary Educational Psychology, 62*, Article 101894. https://doi.org/10.1016/j.cedpsych.2020.101894

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Bönsch, A., Wendt, J., Overath, H., Gurerk, O., Harbring, C., Grund, C., Kittsteiner, T., & Kuhlen, T. W. (2017). Peers at work: Economic real-effort experiments in the presence of virtual co-workers. *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 301–302. https://doi.org/10.1109/VR.2017.7892296

Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology, 66*(2), 268–275. https://doi.org/10.1037/0022-3514.66.2.268

Brisson, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal, 54*(6), 1048–1078. https://doi.org/10.3102/0002831217716084

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology, 102*(4), 964–981. https://doi.org/10.1037/a0019644

Buunk, A. P., Cohen-Schotanus, J., & van Nek, R. H. (2007). Why and how people engage in social comparison while learning social skills in groups. *Group Dynamics, 11*(3), 140–152. https://doi.org/10.1037/1089-2699.11.3.140

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes, 102*(1), 3–21. https://doi.org/10.1016/j.obhdp.2006.09.007

Buunk, A. P., & Mussweiler, T. (2001). New directions in social comparison theory. *European Journal of Social Psychology, 31*(5), 467–475. https://doi.org/10.1002/ejsp.77

Campbell, J. (1990). Self-esteem and clarity of the self-concept. *Journal of Personality and Social Psychology, 59*(3), 538–549. https://doi.org/10.1037/0022-3514.59.3.538

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., Asmundson, G. J. G., Carlbring, P., & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal ofAnxiety Disorders, 61*, 27–36. https://doi.org/10.1016/j.janxdis.2018.08.003

Cialdini, R. B., & Richardson, K. D. (1980). Two indirect tactics of image management: Basking and blasting. *Journal of Personality and Social Psychology, 39*(3), 406–415. https://doi.org/10.1037/0022-3514.39.3.406

Collins, R. (2000). Among the better ones: Upward assimilation in social comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison* (pp. 159–171). Springer. https://doi.org/10.1007/978-1-4615-4237-7_9

Cross, S. E., Gore, J. S., & Morris, M. L. (2003, Nov). The relational-interdependent self-construal, self-concept consistency, and well-being. *Journal of Personality and Social Psychology, 85*(5), 933–944. https://doi.org/10.1037/0022-3514.85.5.933

Cummings, J. J., & Bailenson, J. (2016). How immersive is enough? A meta-analysis of the affect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Dai, D. Y., & Rinn, A. N. (2008). The Big-Fish-Little-Pond effect: What do we know and where do we go from here? *Educational Psychology Review*, *20*(3), 283–317. https://doi.org/10.1007/s10648-008-9071-x

Davis, J. A. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology, 22*(1), 17–31. https://doi.org/10.1086/224257

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology*, *110*(8), 1112–1126. https://doi.org/10.1037/edu0000259

Dieker, L., Hynes, M., Stapleton, C., & Hughes, C. (2007). Virtual classrooms: STAR simulator. *New Learning Technology SALT, 4*, 1–22.

Diener, E., & Fujita, F. (1997). Social comparisons and subjective well-being. In A. P. Buunk & F. X. Gibbons (Eds.), *Health, coping, and well-being: Perspectives from social comparison theory* (pp. 329–358). Lawrence Erlbaum.

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research, 78*(4), 828–879. https://doi.org/10.3102/0034654308321210

Dumas, F., Huguet, P., Monteil, J. M., Rastoul, C., & Nezlek, J. (2005). Social comparison in the classroom: Is there a tendency to compare upward in elementary school? *Current Research in Social Psychology, 10*(12), 166–187.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eccles, J. S., Wigfield, A., Flanagan, C. A., Miller, C., Reuman, D. A., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality*, *57*(2), 283–310. https://doi.org/10.1111/j.1467-6494.1989.tb00484.x

Ferring, D., & Filipp, S.-H. (1996). Messung des Selbstwertgefühls: Befunde zu Reliabilität, Validität und Stabilität der Rosenberg-Skala [Measurement of self-esteem: Findings on reliability, validity, and stability of the Rosenberg Scale]. *Diagnostica, 42*, 284–292.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Gibbons, F. X., & Buunk, A. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76*(1), 129–142. https://doi.org/10.1037/0022-3514.76.1.129

Göllner, R., Damian, R. I., Nagengast, B., Roberts, B. W., & Trautwein, U. (2018). It's not only who you are but who you are with: High school composition and individuals' attainment over the life course. *Psychological Science, 29*(11), 1785–1796. https://doi.org/10.1177/0956797618794454

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*(1), 3–25. https://doi.org/10.1037/0033-295x.109.1.3

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the Field. *Educational Researcher, 42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Guadagno, R. E., Blascovich, J., Bailenson, J., & McCall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology, 10*(1), 1–22. https://doi.org/10.108/15213260701300865

Gürerk, Ö., Bönsch, A., Kittsteiner, T., & Staffeldt, A. (2019). Virtual humans as co-workers: A novel methodology to study peer effects. *Journal of Behavioral and Experimental Economics, 78*, 17–29. https://doi.org/10.1016/j.socec.2018.11.003

Harter, S. (1986). Cognitive-developmental processes in the integration of concepts about emotions and the self. *Social Cognition, 4*(2), 119–151. https://doi.org/10.1521/soco.1986.4.2.119

Heidicker, P., Langbehn, E., & Steinicke, F. (2017). Influence of avatar appearance on presence in social VR. *IEEE Symposium on 3D User Interfaces*. https://doi.org/10.1109/3DUI.2017.7893357

Helgeson, V. S., & Mickelson, K. D. (1995). Motives for social comparison. *Personality and Social Psychology Bulletin, 21*(11), 1200–1209. https://doi.org/10.1177/01461672952111008

Howard, M. C. (2018). Virtual reality interventions for personal development: A meta-analysis of hardware and software. *Human-Computer Interaction, 34*(3), 205–239. https://doi.org/10.1080/07370024.2018.1469408

Huguet, P., Dumas, F., Marsh, H. W., Wheeler, L., Seaton, M., Nezlek, J., Suls, J., & Régner, I. (2009). Clarifying the role of social comparison in the Big-Fish-Little-Pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97*(1), 156–170. https://doi.org/10.1037/a0015558

Huguet, P., Dumas, F., Monteil, J. M., & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its benficial impact on performance. *European Journal of Social Psychology, 31*, 557–578. https://doi.org/10.1002/ejsp.81

Kizilcec, R. F., Bailenson, J., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*(3), 724–739. https://doi.org/10.1037/edu0000013

Lanier, M. K., Tech, V., Waddell, T. F., Tamul, D. J., Ivory, J. D., & Przybylski, A. K. (2019). Virtual reality check: Statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments. *Computers in Human Behavior, 100*, 70–78. https://doi.org/10.1016/j.chb.2019.06.015

Levine, J. M. (1983). Social comparison and education. In J. M. Levine & M. C. Wang (Eds.), *Teacher and student perceptions: Implications for learning* (pp. 29–55). Lawrence Erlbaum.

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). Measuring presence: The temple presence inventory. *Proceedings of the 12th Annual International Workshop on Presence*, 1–15.

Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the Big-Fish-Little-Pond effect. *Contemporary Educational Psychology, 30*(3), 263–285. https://doi.org/10.1016/j.cedpsych.2004.10.002

Lugrin, J.-L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT, 3*, Article 26. https://doi.org/10.3389/fict.2016.00026

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*(1), Article 83. https://doi.org/10.1037/1082-989X.7.1.83

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development, 66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Marsh, H. W. (1987). The Big-Fish-Little-Pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. *Perspectives on Psychological Science, 1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Hau, K. T. (2003). Big-Fish-Little-Pond effect on academic self-concept. A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist, 58*(5), 364–376. https://doi.org/10.1037/0003-066x.58.5.364

Marsh, H. W., Kong, C. K., & Hau, K. T. (2000). Longitudinal multilevel models of the Big-Fish-Little-Pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, *78*(2), 337–349. https://doi.org/10.1037//0022-3514.78.2.337

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*, 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*(4), 542–552. https://doi.org/10.1177/0146167207312313

Marsh, H. W., Parker, P. D., Guo, J., Basarkod, G., Niepel, C., & Van Zanden, B. (2020). Illusory gender-equality paradox, math self-concept, and frame-of-reference effects: New integrative explanations for multiple paradoxes. *Journal of Personality and Social Psychology*, *121*(1), 168–183. https://doi.org/10.1037/pspp0000306

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The Big-Fish-Little-Pond effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*(3), 319–350. https://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and Big-Fish-Little-Pond effect on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*(3), 510–524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: construct validation and relations to well-being and achievement. *Journal of Personality, 74*(2), 403–456. https://doi.org/10.1111/j.1467-6494.2005.00380.x

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal, 34*(4), 691–720. https://doi.org/10.3102/00028312034004691

Mayer, A., Thoemmes, F., Rose, N., Steyer, R., & West, S. G. (2014). Theory and analysis of total, direct, and indirect causal effects. *Multivariate Behavioral Research, 49*(5), 425–442. https://doi.org/10.1080/00273171.2014.931797

Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology, 93*(4), 826–835. https://doi.org/10.1037/0022-0663.93.4.826

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Mussweiler, T. (2001). Focus of comparison as a determinant of assimilation versus contrast in social comparison. *Personality and Social Psychology Bulletin, 27*(1), 38–47. https://doi.org/10.1177/0146167201271004

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*(3), 472–489. https://doi.org/10.1037/0033-295X.110.3.472

Mussweiler, T., Rüter, K., & Epstude, K. (2004). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology, 87*(6), 832–844. https://doi.org/10.1037/0022-3514.87.6.832

Mussweiler, T., & Strack, F. (2000). The "relative self": Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology, 79*(1), 23–38. https://doi.org/10.1037/0022-3514.79.1.23

Muthén, L. K., & Muthén, B. O. (1998-2017). *MPlus user's guide* (8th ed.). Muthén & Muthén. https://doi.org/10.1111/j.1600-0447.2011.01711.x

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology, 104*(4), 1033–1053. https://doi.org/10.1037/a0027697

Neugebauer, J., Ray, D. G., & Sassenberg, K. (2016). When being worse helps: The influence of upward social comparisons and knowledge awareness on learner engagement and learning in peer-to-peer knowledge exchange. *Learning and Instruction, 44*, 41–52. https://doi.org/10.1016/j.learninstruc.2016.02.007

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41*(3), 181–206. https://doi.org/10.1207/s15326985ep4103

Parsons, S., & Mitchell, P. (2002). The potential of virtual reality in social skills training for people with autistic spectrum disorders. *Journal of Intellectual Disability Research, 46*(5), 430–443. https://doi.org/10.1046/j.1365-2788.2002.00425.x

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*, 660, Article 660. https://doi.org/10.3389/fnhum.2015.00660

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. *Journal of Personality and Social Psychology, 117*(1), 166–185. https://doi.org/10.1037/pspp0000230

Preacher, K. J., & Kelley, K. (2011, Jun). Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93–115. https://doi.org/10.1037/a0022658

Pyszczynski, T., Greenberg, J., & LaPrelle, J. (1985). Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. *Journal of Experimental Social Psychology, 29*(1), 195–211. https://doi.org/10.1016/0022-1031(85)90015-0

Richards, J. (2017). Infrastructures for immersive media in the classroom. In D. Liu, C. Dede, R. Huang, & J. Richards (Eds.), *Virtual, augmented, and mixed realities in education* (pp. 89–104). Springer. https://doi.org/10.1007/978-981-10-5490-7_6

Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior, 72*, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence, 10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica, 51*(4), 183–194. https://doi.org/10.1026/0012-1924.51.4.183

Seaton, M., Marsh, H., Dumas, F., Huguet, P., Monteil, J. M., Regner, I., Blanton, H., Buunk, A. P., Gibbons, F. X., Kuyper, H., Suls, J. M., & Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the Big-Fish-Little-Pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology, 47*, 73–103. https://doi.org/10.1348/014466607X202309

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the Big-Fish-Little-Pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology, 101*(2), 403–419. https://doi.org/10.1037/a0013838

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441. https://doi.org/10.2307/1170010

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3*, 1–47, Article 74. https://doi.org/10.3389/frobt.2016.00074

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence, 6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science, 11*(5), 159–163. https://doi.org/10.1111/1467-8721.00191

Suls, J., & Wheeler, L. (2000). A selective history of classic and neo-social comparison theory. In J. M. Suls & L. Wheeler (Eds.), *Handbook of Social Comparison* (pp. 3–19). Springer. https://doi.org/10.1007/978-1-4615-4237-7_1

Suls, J. M., & Mullen, B. (1982). From the cradle to the grave: Comparison of self-evaluation across the lifespan. In J. M. Suls & A. G. Greenwald (Eds.), *Psychological Perspectives on the Self* (Vol. 2, pp. 97–125). Lawrence Erlbaum.

Swann, W. B., Jr., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist, 62*(2), 84–94. https://doi.org/10.1037/0003-066X.62.2.84

Trautwein, U., & Lüdtke, O. (2005). The Big-Fish-Little-Pond effect: Future research questions and educational implications. *Zeitschrift für Pädagogische Psychologie, 19*(3), 137–140. https://doi.org/10.1024/1010-0652.19.3.137

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. Roberts (Eds.), *Psychosocial skills and school systems in the twenty-fi rst century: Theory, research, and applications* (pp. 187–214). Springer. https://doi.org/10.1007/978-3-319-28606-8_8

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

von Keyserlingk, L., Becker, M., Jansen, M., & Maaz, K. (2020). Leaving the pond—Choosing an ocean: Effects of student composition on STEM major choices at university. *Journal of Educational Psychology, 112*(4), 751–764. https://doi.org/10.1037/edu0000378

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology, 25*(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology, 1*, 27–31. https://doi.org/10.1016/0022-1031(66)90062-X

Wheeler, L., & Suls, J. (2005). Social comparison and self-evaluations of competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 566–578). Guilford Press.

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of achievement motivation and engagement. In R. M. Lerner (Ed.), *Handbook of child psychology and developmental science* (7th ed., pp. 657–700). Wiley. https://doi.org/10.1002/9781118963418.childpsy316

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology, 110*(7), 1005–1025. https://doi.org/10.1037/edu0000248

Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. *Personality and Social Psychology Review, 21*(3), 253–277. https://doi.org/10.1177/1088868316652279

Zell, E., & Alicke, M. D. (2009). Contextual neglect, self-evaluation, and the frog-pond effect. *Journal of Personality and Social Psychology, 97*(3), 467–482. https://doi.org/10.1037/a0015453

Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: evidence and explanations. *Personality and Social Psychology Review, 14*(4), 368–384. https://doi.org/10.1177/1088868310366144

Appendix A

Items Used to Assess Students' Socially-Oriented, Criterion-Oriented, and Dispositional Self-Concept in the Domain of Computational Thinking and Their Perception of the Class' Performance Level

Scale	Items
Socially-oriented self-concept (SC1)	1. I could not solve the robot tasks as easily as the other students in the virtual classroom. 2. Compared with the others, I was really good at giving the right commands to the robots. 3. I could solve the robot tasks faster than the others. 4. It was harder for me to understand the robot tasks than for the other students.
Criterion-oriented self-concept (SC2)	1. I had some problems solving the robot tasks. 2. It was not a problem for me to give the right commands to the robots. 3. I could have easily explained the solutions to the robot tasks. 4. It took me some time to understand the robot tasks.
Dispositional self-concept (SC3)	1. No matter how hard I try, I am not good at solving puzzles. 2. I usually quickly understand topics related to technology, robots, computers, programming, etc. 3. I am good at breaking a complicated problem down into smaller steps. 4. Most of the time it is hard for me to imagine how individual parts of a problem are connected. 5. It is easy for me to solve a complicated task step by step. 6. I simply have no talent for logical thinking.
Perception of the class' performance level	1. There were many good students in the class. 2. The lesson was easy for most of the students in the class. 3. The class would do quite well on a test on the topic of the lesson. 4. The virtual class was pretty good. 5. Most of the class did not really understand the topic.

Supplemental Materials

First, the supplemental materials include additional detailed statistics for the robustness check (see Table S1). Second, all analyses reported in the present study were calculated a second time by additionally including the different configuration conditions of the IVR environment (i.e., seat position [back vs. front] and graphical representation [cartoon vs. more realistic] of virtual characters) as covariates. The supplemental materials (see Table S2–S7) includes the respective statistics.

Table of Contents

Table S1

Standardized Regression Coefficients for the Effect of Hand-Raising Conditions on the Number of Registered Hand-Raising Students and the Perceived Performance Level of the Class, Including Background Variables

	Number of registered hand-raising students					Perceived performance level of class				
	b	*SE*	*t*	*p*	95% CI	*b*	*SE*	*t*	*p*	95% CI
(Intercept)	0.01	0.02	0.25	.801	[-0.04, 0.05]	0.00	0.06	-0.03	.973	[-0.12, 0.12]
20% vs. 35% Hand-raising	0.08	0.06	1.33	.092	[-0.04, 0.21]	0.05	0.05	0.99	.162	[-0.05, 0.15]
20% vs. 65% Hand-raising	0.27***	0.07	3.67	<.001	[0.13, 0.42]	0.26***	0.05	5.38	<.001	[0.16, 0.35]
20% vs. 80% Hand-raising	0.57***	0.05	10.57	<.001	[0.47, 0.68]	0.45***	0.07	6.32	<.001	[0.31, 0.58]
35% vs. 65% Hand-raising	0.19*	0.09	2.21	.014	[0.02, 0.36]	0.20***	0.04	4.96	<.001	[0.12, 0.29]
35% vs. 80% Hand-raising	0.49***	0.05	9.76	<.001	[0.39, 0.59]	0.39***	0.06	6.86	<.001	[0.28, 0.51]
65% vs. 80% Hand-raising	0.30***	0.07	4.31	<.001	[0.16, 0.43]	0.19***	0.05	3.71	<.001	[0.09, 0.28]
Gender [a]	0.03	0.07	0.44	.663	[-0.11, 0.18]	-0.07	0.04	-1.65	.098	[-0.16, 0.01]
Math grade [b]	0.01	0.07	0.21	.831	[-0.12, 0.14]	0.02	0.07	0.25	.802	[-0.12, 0.16]
German grade [b]	0.00	0.06	-0.06	.956	[-0.11, 0.11]	0.03	0.10	0.33	.742	[-0.17, 0.24]
Science grade [b]	-0.04	0.04	-0.95	.343	[-0.12, 0.04]	-0.14**	0.05	-3.08	.002	[-0.23, -0.05]
Prior CT experience [c]	0.00	0.05	0.05	.958	[-0.09, 0.10]	-0.03	0.05	-0.53	.597	[-0.13, 0.08]
Prior IVR experience [d]	0.05	0.06	0.80	.423	[-0.07, 0.16]	0.02	0.05	0.43	.667	[-0.08, 0.12]
Self-concept intelligence [e]	0.09	0.07	1.34	.179	[-0.04, 0.21]	-0.11	0.06	-1.78	.076	[-0.23, 0.01]
Social orientation [e]	0.03	0.05	0.65	.518	[-0.07, 0.13]	0.04	0.04	0.82	.412	[-0.05, 0.12]

Note. CI = Confidence interval; CT = Computational thinking; IVR = Immersive virtual reality.
[a] Gender female = 1, male = 2; [b] Grades were on a scale from 1–6 with lower numbers indicating better achievement; [c] Prior CT experience no = 0, yes = 1; [d] Prior IVR experience none = 0, once = 1, often = 2; [e] Assessed on a 4-point rating scale with higher values indicating higher levels.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table S2

Number of Registered Hand-Raising Students in the Hand-Raising Conditions for Seat Position (Front vs. Back) and the Graphical Representation of Virtual Characters (Cartoon vs. More Realistic)

	Front				Back				Actual
	Cartoon		Realistic		Cartoon		Realistic		
	N	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	*N*	*M (SD)*	
20% Hand-raising	15	4.40 (2.06)	21	6.19 (3.50)	32	7.78 (3.47)	20	7.25 (4.64)	5
35% Hand-raising	18	7.28 (2.91)	19	8.21 (5.87)	25	7.72 (2.65)	23	7.52 (2.98)	9
65% Hand-raising	18	7.89 (4.20)	18	9.94 (4.40)	23	11.83 (4.75)	22	10.32 (5.33)	16
80% Hand-raising	15	16.73 (5.71)	19	12.11 (5.25)	33	12.79 (5.09)	17	14.00 (5.44)	20

Note. The possible answer range was between 0 and 24 (see question format in Figure 4, main text). Actual = the actual number of hand-raising students (i.e., the true value of how many students raised their hands in the respective conditions).

Table S3

Standardized Regression Coefficients for the Effect of the Hand-Raising Conditions on Perceived Performance Level, Controlling for IVR Configuration

	Number of registered hand-raising students					Perceived performance level of class				
	b	*SE*	*t*	*p*	95% CI	*b*	*SE*	*t*	*p*	95% CI
(Intercept)	0.01	0.02	0.31	.754	[-0.04, 0.05]	0.01	0.05	0.03	.980	[-0.14, 0.10]
20% vs. 35% Hand-raising	0.08	0.05	1.54	.062	[-0.02, 0.19]	0.07	0.05	1.47	.071	[-0.02, 0.15]
20% vs. 65% Hand-raising	0.28***	0.07	4.22	<.001	[0.15, 0.42]	0.26***	0.04	7.28	<.001	[0.19, 0.33]
20% vs. 80% Hand-raising	0.58***	0.06	10.53	<.001	[0.47, 0.69]	0.45***	0.07	6.67	<.001	[0.32, 0.58]
35% vs. 65% Hand-raising	0.20**	0.08	2.67	.004	[0.05, 0.35]	0.19***	0.04	5.27	<.001	[0.12, 0.26]
35% vs. 80% Hand-raising	0.49***	0.05	9.39	<.001	[0.39, 0.60]	0.38***	0.07	5.92	<.001	[0.26, 0.51]
65% vs. 80% Hand-raising	0.29***	0.07	4.12	<.001	[0.15, 0.18]	3.55***	0.74	4.81	<.001	[0.09, 0.29]
Seat position [a]	0.08	0.05	1.48	.138	[-0.03, 0.18]	0.16**	0.06	2.84	.004	[0.05, 0.27]
Graphical representation [b]	-0.01	0.04	-0.17	.864	[-0.08, 0.07]	0.07*	0.03	2.18	.030	[0.01, 0.14]

Note. CI = Confidence interval.

[a] Seat position front = 1, back = 2; [b] Graphical representation cartoon = 1, realistic = 2.

* *p* < .05. ** *p* < .01. *** *p* < .001.

Table S4

Standardized Regression Coefficients for the Effect of the Hand-Raising Conditions on Perceived Performance Level, Including Background Variables and Controlling for IVR Configuration

	Number of registered hand-raising students					Perceived performance level of class				
	b	*SE*	*t*	*p*	95% CI	*b*	*SE*	*t*	*p*	95% CI
(Intercept)	0.01	0.02	0.22	.828	[-0.04, 0.05]	0.00	0.06	-0.07	.947	[-0.12, 0.11]
20% vs. 35% Hand-raising	0.08	0.06	1.38	.084	[-0.04, 0.21]	0.06	0.05	1.22	.113	[-0.03, 0.15]
20% vs. 65% Hand-raising	0.27***	0.07	3.68	<.001	[0.13, 0.42]	0.26***	0.04	6.41	<.001	[0.18, 0.34]
20% vs. 80% Hand-raising	0.57***	0.05	10.54	<.001	[0.47, 0.68]	0.45***	0.07	6.34	<.001	[0.31, 0.59]
35% vs. 65% Hand-raising	0.19*	0.09	2.18	.015	[0.02, 0.36]	0.20***	0.04	5.22	<.001	[0.13, 0.28]
35% vs. 80% Hand-raising	0.49***	0.05	9.56	<.001	[0.39, 0.59]	0.39***	0.06	6.48	<.001	[0.28, 0.51]
65% vs. 80% Hand-raising	0.30***	0.07	4.08	<.001	[0.15, 0.44]	0.19***	0.05	3.54	<.001	[0.08, 0.26]
Seat position [a]	0.08	0.05	1.51	.130	[-0.02, 0.19]	0.16**	0.06	2.81	.005	[0.05, 0.26]
Graphical representation [b]	0.01	0.04	0.23	.819	[-0.07, 0.09]	0.05	0.05	1.14	.256	[-0.04, 0.14]
Gender [c]	0.03	0.08	0.39	.696	[-0.12, 0.18]	-0.07	0.04	-1.67	.095	[-0.16, 0.01]
Math grade [d]	0.01	0.07	0.14	.886	[-0.13, 0.15]	0.02	0.07	0.25	.802	[-0.12, 0.16]
German grade [d]	0.00	0.06	-0.07	.946	[-0.12, 0.11]	0.03	0.10	0.25	.802	[-0.17, 0.22]
Science grade [d]	-0.03	0.04	-0.69	.492	[-0.12, 0.06]	-0.12**	0.04	-3.06	.002	[-0.20, -0.04]
Prior CT experience [e]	0.00	0.05	0.05	.963	[-0.09, 0.10]	-0.03	0.05	-0.53	.597	[-0.12, 0.07]
Prior IVR experience [f]	0.05	0.06	0.79	.429	[-0.07, 0.16]	0.02	0.05	0.42	.673	[-0.08, 0.12]
Self-concept intelligence [g]	0.09	0.07	1.36	.173	[-0.04, 0.23]	-0.10	0.06	-1.55	.122	[-0.22, 0.03]
Social orientation [g]	0.04	0.05	0.76	.450	[-0.06, 0.14]	0.05	0.04	1.11	.269	[-0.04, 0.13]

Note. CI = Confidence interval; CT = Computational thinking; IVR = Immersive virtual reality.
[a] Seat position front = 1, back = 2; [b] Graphical representation cartoon = 1, realistic = 2; [c] Gender female = 1, male = 2; [d] Grades on a scale from 1–6 with lower numbers indicating better achievement; [e] Prior CT experience no = 0, yes = 1; [f] Prior IVR experience none = 0, once = 1, often = 2; [g] Assessed on a 4-point rating scale with higher values indicating higher levels.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table S5

Standardized Regression Coefficients for the Effects of the Hand-Raising Conditions on Self-Concept, Controlling for IVR Configuration

	b	*SE*	*t*	*p*	95% CI
Model 1: Socially-oriented self-concept (SC1)					
(Intercept)	0.00	0.10	0.00	1.000	[-0.19, 0.19]
20% vs. 35% Hand-raising	-0.02	0.05	-0.36	.359	[-0.12, 0.08]
20% vs. 65% Hand-raising	-0.05	0.06	-0.83	.205	[-0.16, 0.06]
20% vs. 80% Hand-raising	-0.11*	0.05	-2.30	.011	[-0.21, -0.02]
Seat position [a]	-0.12*	0.06	-2.00	.045	[-0.23, 0.00]
Graphical representation [b]	0.00	0.07	0.06	.949	[-0.13, 0.14]
Model 2: Criterion-oriented self-concept (SC2)					
(Intercept)	0.00	0.08	0.00	1.000	[-0.16, 0.16]
20% vs. 35% Hand-raising	-0 04	0.03	-1.18	.239	[-0.10, 0.03]
20% vs. 65% Hand-raising	0.05	0.06	0.91	.361	[-0.06, 0.16]
20% vs. 80% Hand-raising	0.00	0.07	-0.02	.987	[-0.13, -0.13]
Seat position [a]	-0.08	0.04	-1.78	.075	[-0.16, 0.01]
Graphical representation [b]	-0.09	0.05	-1.69	.091	[-0.20, 0.02]
Model 3: Dispositional self-concept (SC3)					
(Intercept)	0.00	0.07	0.01	.995	[-0.14, 0.14]
20% vs. 35% Hand-raising	-0.07	0.04	-1.80	.072	[-0.16, 0.01]
20% vs. 65% Hand-raising	0.03	0.08	0.33	.742	[-0.13, 0.19]
20% vs. 80% Hand-raising	-0.01	0.07	-0.13	.896	[-0.14, 0.12]
Seat position [a]	-0.02	0.04	-0.61	.542	[-0.09, 0.05]
Graphical representation [b]	0.00	0.05	-0.08	.933	[-0.11, 0.10]

Note. CI = Confidence interval.
[a] Seat position front = 1, back = 2; [b] Graphical representation cartoon = 1, realistic = 2.
* $p < .05$.

Table S6

Standardized Regression Coefficients for the Effects of the Hand-Raising Conditions on Self-Concept, Controlling for IVR Configuration and Including Background Variables

	b	*SE*	*t*	*p*	95% CI
Model 1: Socially-oriented self-concept (SC 1)					
(Intercept)	0.00	0.09	0.01	.992	[-0.17, 0.17]
20% vs. 35% Hand-raising	-0.01	0.05	-0.15	.439	[-0.10, 0.09]
20% vs. 65% Hand-raising	-0.05	0.06	-0.90	.186	[-0.16, 0.06]
20% vs. 80% Hand-raising	-0.11**	0.04	-2.52	.006	[-0.19, -0.02]
Seat position [a]	-0.10	0.05	-1.91	.057	[-0.21, 0.00]
Graphical representation [b]	0.03	0.07	0.38	.707	[-0.11, 0.17]
Gender [c]	-0.07	0.07	0.88	.379	[-0.08, 0.21]
Math grade [d]	-0.14**	0.05	-3.00	.003	[-0.24, -0.05]
German grade [d]	0.15***	0.04	3.97	<.001	[0.08, 0.23]
Science grade [d]	0.09*	0.05	1.97	.049	[0.00, 0.19]
Prior CT experience [e]	0.07	0.06	1.09	.278	[-0.05, 0.19]
Prior IVR experience [f]	0.02	0.07	0.32	.748	[-0.11, 0.15]
Self-concept intelligence [g]	0.12	0.08	1.51	.131	[-0.04, 0.28]
Social orientation [g]	0.03	0.04	0.60	.547	[-0.06, 0.11]
Model 2: Criterion-oriented self-concept (SC 2)					
(Intercept)	0.00	0.07	0.02	.985	[-0.14, 0.14]
20% vs. 35% Hand-raising	-0.01	0.04	-0.34	.737	[-0.09, 0.06]
20% vs. 65% Hand-raising	0.06	0.05	1.19	.233	[-0.04, 0.14]
20% vs. 80% Hand-raising	0.02	0.06	0.24	.814	[-0.11, -0.14]
Seat position [a]	-0.06	0.04	-1.38	.167	[-0.14, 0.03]
Graphical representation [b]	-0.07	0.06	-1.34	.181	[-0.18, 0.04]
Gender [c]	-0.01	0.07	-0.20	.841	[-0.14, 0.12]
Math grade [d]	-0.18*	0.07	-2.46	.014	[-0.32, -0.04]
German grade [d]	0.07	0.06	1.02	.306	[-0.06, 0.19]
Science grade [d]	0.13	0.08	1.73	.084	[-0.02, 0.29]
Prior CT experience [e]	0.10	0.07	1.58	.115	[-0.03, 0.23]
Prior IVR experience [f]	-0.03	0.04	-0.73	.468	[-0.11, 0.05]
Self-concept intelligence [g]	0.17**	0.06	2.63	.009	[0.04, 0.29]
Social orientation [g]	0.04	0.05	0.82	.411	[-0.06, 0.15]

(continued)

	b	*SE*	*t*	*p*	95% CI
Model 3: Dispositional self-concept (SC 3)					
(Intercept)	0.00	0.06	-0.02	.982	[-0.11, 0.11]
20% vs. 35% Hand-raising	-0.03	0.05	-0.74	.462	[-0.12, 0.06]
20% vs. 65% Hand-raising	0.02	0.08	0.22	.824	[-0.14, 0.17]
20% vs. 80% Hand-raising	0.01	0.06	0.21	.831	[-0.11, 0.13]
Seat position [a]	-0.01	0.03	-0.17	.868	[-0.07, 0.06]
Graphical representation [b]	0.05	0.04	1.08	.280	[-0.04, 0.13]
Gender [c]	0.05	0.05	0.95	.342	[-0.05, 0.15]
Math grade [d]	-0.13*	0.05	-2.58	.010	[-0.23, -0.03]
German grade [d]	0.07	0.05	1.32	.188	[-0.03, 0.17]
Science grade [d]	0.10	0.07	1.33	.185	[-0.05, 0.24]
Prior CT experience [e]	0.19***	0.03	5.00	<.001	[0.12, 0.27]
Prior IVR experience [f]	0.02	0.04	0.36	.719	[-0.07, 0.10]
Self-concept intelligence [g]	0.35***	0.09	4.08	<.001	[0.18, 0.52]
Social orientation [g]	0.01	0.03	0.20	.845	[-0.06, 0.07]

Note. CI = Confidence interval; CT = Computational thinking; IVR = Immersive virtual reality. [a] Seat position front = 1, back = 2; [b] Graphical representation cartoon = 1, realistic = 2; [c] Gender female = 1, male = 2; [d] Grades on a scale from 1–6 with lower numbers indicating better achievement; [e] Prior CT experience no = 0, yes = 1; [f] Prior IVR experience none = 0, once = 1, often = 2; [g] Assessed on a 4-point rating scale with higher values indicating higher levels.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table S7

Standardized Regression Coefficients for the Effects of the Hand-Raising Conditions on Socially-oriented Self-concept including the Mediator, Controlling for IVR Configuration and Including Background Variables

	b	*SE*	*t*	*p*	95% CI
Model 1: Hand-raising conditions on socially-oriented self-concept					
(Intercept)	0.00	0.10	0.00	1.000	[-0.19, 0.19]
20% vs. 35% Hand-raising	-0.02	0.05	-0.36	.359	[-0.12, 0.08]
20% vs. 65% Hand-raising	-0.05	0.06	-0.83	.205	[-0.16, 0.06]
20% vs. 80% Hand-raising	-0.11*	0.05	-2.30	.011	[-0.21, -0.02]
Seat position [a]	-0.12*	0.06	-2.00	.045	[-0.23, 0.00]
Graphical representation [b]	0.00	0.07	0.06	.949	[-0.13, 0.14]
Model 2: Including mediator perceived performance level					
(Intercept)	0.00	0.10	0.00	.997	[-0.19, 0.19]
20% vs. 35% Hand-raising	-0.01	0.06	-0.13	.449	[-0.12, 0.10]
20% vs. 65% Hand-raising	0.00	0.06	-0.06	.478	[-0.12, 0.11]
20% vs. 80% Hand-raising	-0.04	0.05	-0.76	.225	[-0.15, 0.06]
Perceived performance	-0.17***	0.04	-4.29	<.001	[-0.24, -0.09]
Seat position [a]	-0.09	0.06	-1.46	.146	[-0.21, 0.03]
Graphical representation [b]	0.02	0.07	0.23	.821	[-0.12, 0.15]
Model 3: Including mediator and background variables					
(Intercept)	0.00	0.10	0.00	.997	[-0.17, 0.17]
20% vs. 35% Hand-raising	0.00	0.05	0.02	.494	[-0.10, 0.10]
20% vs. 65% Hand-raising	-0.01	0.06	-0.24	.405	[-0.13, 0.10]
20% vs. 80% Hand-raising	-0.05	0.04	-1.02	.156	[-0.13, 0.04]
Perceived performance	-0.14***	0.03	-4.48	<.001	[-0.20, -0.08]
Seat position [a]	-0.08	0.06	-1.38	.167	[-0.20, 0.03]
Graphical representation [b]	0.03	0.07	0.46	.643	[-0.11, 0.18]
Gender [c]	0.06	0.07	0.78	.435	[-0.08, 0.19]
Math grade [d]	-0.14**	0.04	-3.39	.001	[-0.22, -0.06]
German grade [d]	0.16***	0.04	4.19	<.001	[0.08, 0.23]
Science grade [d]	0.08	0.05	1.54	.124	[-0.02, 0.17]
Prior CT experience [e]	0.06	0.06	1.11	.268	[-0.05, 0.18]
Prior IVR experience [f]	0.02	0.07	0.35	.729	[-0.11, 0.16]
Self-concept intelligence [g]	0.11	0.08	1.37	.172	[-0.05, 0.26]
Social orientation [g]	0.03	0.05	0.71	.476	[-0.06, 0.12]

Note. CI = Confidence Interval; CT = Computational Thinking; IVR = Immersive Virtual Reality. [a] Seat position front = 1, back = 2; [b] Graphical representation cartoon = 1, realistic = 2; [c] Gender female = 1, male = 2; [d] Grades on a scale from 1–6 with lower numbers indicating better achievement; [e] Prior CT experience no = 0, yes = 1; [f] Prior IVR experience none = 0, once = 1, often = 2; [g] Assessed on a 4-point rating scale with higher values indicating higher levels.
* $p < .05$. ** $p < .01$. *** $p < .001$.

4

STUDY 2

Hasenbein, L., Stark, P., Trautwein, U., Gao, H., Kasneci, E., & Göllner, R. (2021). *Do students actively seek comparisons with others? Using eye-movement data from a virtual reality classroom to uncover social information processing.* Manuscript submitted for publication.

The following manuscript has not yet been accepted or published. The version displayed here might not exactly replicate the final version published in the journal. It is not the copy of record.

Abstract

Higher-achieving peers have repeatedly been found to negatively impact students' evaluations of their own academic abilities. Building on social comparison theory, this pattern is assumed to result from students comparing themselves to their classmates; however, based on existing research, it remains unclear how exactly students make use of social comparison information in the classroom. To examine to what extent students (N = 353 sixth graders) actively attend and respond to different levels of peers' achievement-related behaviour, we used eye-tracking data from an immersive virtual reality classroom. We experimentally varied virtual classmates' achievement-related behaviour (i.e., their hand-raising) during instruction and found that students actively processed this social comparison information (as indicated by the number of peer learners looked at and mean pupil diameter). Students who attended more to social comparison information (as indicated by more frequent and longer gazes at peer learners) exhibited less favourable self-evaluations.

Do Students Actively Seek Comparisons with Others? Using Eye Movement Data from a Virtual Reality Classroom to Uncover Social Information Processing

Social comparisons are a central aspect of human nature. How we perceive and evaluate ourselves (e.g., how competent we think we are in a specific domain) is substantially shaped by who we compare ourselves with[1,2]. Understanding social comparison processes in-depth is crucial, considering that the beliefs we hold about ourselves and our abilities have far-reaching consequences for individual life trajectories. In particular, a high academic self-concept—meaning high levels of confidence in one's own abilities and performance at school (e.g., "I am good at mathematics")[3]—is considered a critical determinant of successful learning and a fundamental prerequisite for achieving one’s academic goals and successful education and career choices[4-7].

It is thus no wonder that the determinants of students’ academic self-concept—in other words, questions of what leads to individual differences in beliefs about one’s own abilities—are among the most studied phenomena in social and educational psychology[8,9]. One of the most prominent findings in the educational context, the so-called Big-Fish-Little-Pond Effect (BFLPE)[10], suggests that students in a high-performing class evaluate their own academic abilities as worse compared to equally able peers in a class with a lower average performance level. Thirty-five years after the original BFLPE study[10] was published in 1987, searching for “Big-Fish-Little-Pond Effect” yields well over 4,500 results on Google Scholar. The BFLPE has been confirmed for different grade levels, school types, and in different countries all over the world[11-13]. BFLPE research typically sets a class’s performance level into relation with individual students’ performance; if classmates’ performance is statistically significantly related to individual students’ academic self-concept, it is concluded that social comparisons have occurred (e.g., lower-achieving students evaluate their own academic abilities as worse due to comparisons with higher-achieving peers)[11,12,14].

A question that has been rather tangential to BFLPE research so far, but which is central to understanding its underlying processes, pertains to the nature of the social comparisons that take place: Do students *actively compare* themselves with their peers during instruction? Or are interindividual differences in academic self-concept the consequence of other factors, such as achievement-related remarks from peers (e.g., “Look at them, they’re not getting it once again”) and/or teacher feedback and grading practices[15,16]?

In order to answer this question and advance research on social comparisons in the classroom, the present study brings together three strands of research, which we explain in more detail in the following sections. First, we outline the status quo on BFLPE research and identify the role of active social comparisons as a central open question. Second, we draw on eye-tracking research and describe how eye movements can be used to obtain indicators of social comparison processes. Third, we illustrate how immersive virtual reality (IVR) provides the optimal experimental setting for such research, allowing for a realistic and authentic simulation of classroom social scenarios while at the same time making it possible to work and collect fine-grained process data in a standardized and controlled environment.

In a typical classroom situation, peer learners are considered the most important source of information and the primary reference group for social comparisons[9,17]. The large body of BFLPE research is based on correlational analyses of large-scale data from real-world learning contexts, which, although it provides compelling evidence[11-14,18], leaves the actual causes and underlying mechanisms an open question[19]: Is it in fact active social comparisons that lead to the negative effect of higher-achieving peers on lower-achieving students' academic self-concept? In order to answer this question, researchers would need to gain insights into social comparison processes at the moment they occur—something that is difficult to achieve in real-world classrooms given all of their complex dynamics and simultaneous happenings during instruction[20]. In order to achieve standardized conditions that allow such fine-grained insights into social comparison processes and can help to answer questions of causality, social psychological studies typically take a strictly experimental approach and therefore tend to be situated in lab settings[2,21-23]. Although such experimental studies provide valuable insights into the basic mechanisms underlying social comparisons, it is unclear to what extent these findings apply to an actual—much more complex and dynamic—classroom situation. Typically, these experiments provide manipulated performance-related social comparison information (i.e., researchers explicitly tell participants that they belong to the lowest vs. highest achieving group)[2,24] or instruct participants to compare themselves to a specific (fictitious) comparison target (i.e., participants do not need to actively search for social comparison information)[25-28]. As a result of these experimental designs, such studies can hardly answer the question of whether students actively engage in social comparisons themselves or are rather passively affected by comparisons stemming from their peers, teacher or grades.

Notably, all of the aforementioned research relied upon students' self-reports to gain insights into social comparison processes and resulting differences in academic self-evaluations. Hence, their findings ultimately rely on students' introspective (and possibly biased) statements, which are likely to differ in the extent to which they correspond to actual behaviour[29]. Eye-tracking has great potential with respect to gaining a more in-depth and unbiased understanding of the processes underlying social comparisons[30-32]. Basic eye movement data and eye-related features can be used to identify different mechanisms of decoding and integrating social comparison information. First of all, students need to orient themselves in a classroom situation and actively attend to the social comparison information in order to notice it. Eye-tracking data makes it possible to extract information about the object a student is looking at[33,34] (we explain all technical details about how to use ray-casting[35] to identify the object of gaze in an IVR environment in the Methods section). Given that humans are able to guide their attention in the world and selectively focus on relevant objects while ignoring others, it can be assumed that moving one's eyes to a relevant location in space is an indication that one is paying attention to the information contained in the object of one's gaze (i.e., overt spatial or so-called visual attention)[36,37]. With regards to processing social comparison information in the classroom, looking at higher numbers of peer learners would be an indication of students noticing corresponding social comparison information. Secondly, beyond simply noticing it, students need to actively engage in social comparisons and therefore process the social comparison information. Hence, not only looking at a higher number of peer learners but also the frequency of gazing and gaze duration at them provide valuable insights[38]. Based on what is known about fixations (i.e., the time a gaze rests at a particular place[39]), more frequent and longer visual attention (in the sense of a higher number and longer durations of fixations) are a sign of deeper cognitive processing of the gazed information[40-42]. In other words, the more often and the longer students look at their classmates, the more they presumably process the provided social comparison information (i.e., consider it when making comparisons). Lastly, pupillometry provides further insights into students' responses to the social comparison information. Pupil diameter is a commonly examined eye-related feature[43,44], which has been associated with cognitive load[45-49], information encoding and retrieval[50-52] and affective arousal[53-57]. In this vein, greater pupil diameter indicates that students are more concerned with the (social comparison) information they process, both in terms of cognitive processing (e.g., because the information is new and might deviate from their usual experiences) and in terms of an affective reaction to it (e.g., because students relate the information to themselves and use it to compare their own abilities).

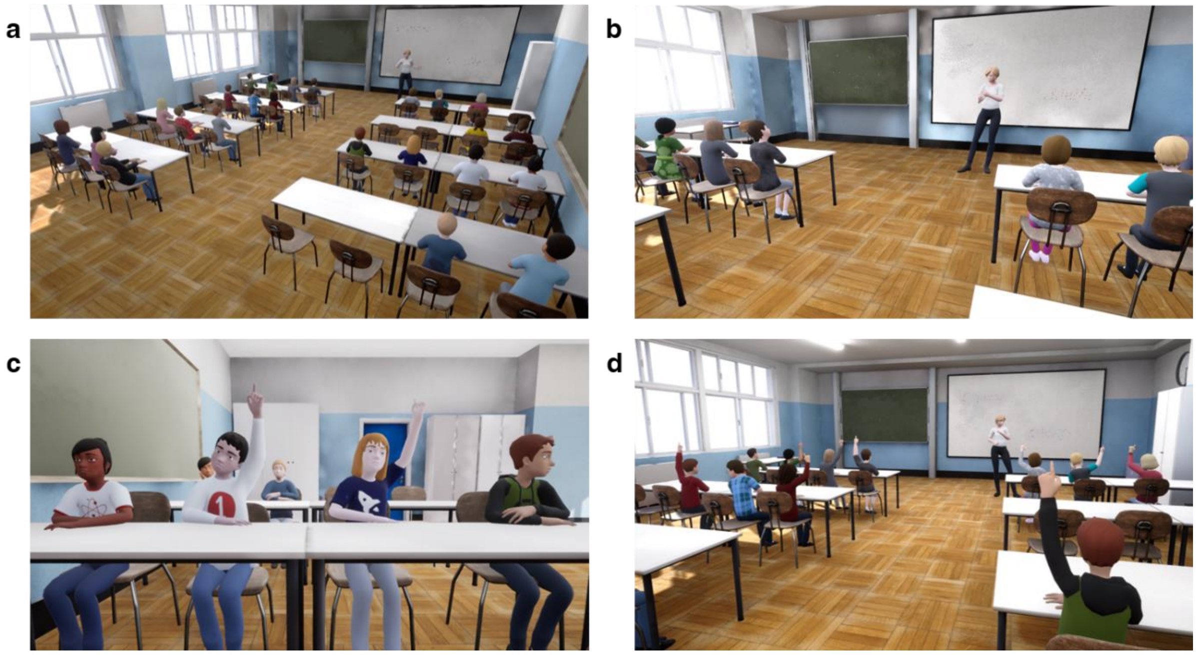


Fig. 1 | Virtual classroom situation with different peer behaviours. The images show **a**, a bird's eye view of the IVR classroom, and **b**, the view of a student in a sitting position in the second row from the front (both in a situation without any hand-raising by peers), **c**, a close-up of virtual peer learners' hand-raising, and **d**, a situation with 80% hand-raising peer learners from the perspective of a student sitting in the back.

We used an immersive virtual reality (IVR) classroom in our study (see Fig. 1 for an impression) to gain insights into social comparison processes during instruction. Recently, motivated by continuous advances in IVR technology, increasing numbers of researchers have come to acknowledge the methodological affordances of IVR as an experimental tool[58-63]. IVRs—once programmed—provide cost- and time-efficient, highly reproducible testing settings with maximum control of confounding and manipulated variables while simultaneously providing an authentic experience (for examples in the classroom context, see refs. [64-69]). Advocates of IVR as an experimental tool highlight evidence that users' behaviour in IVRs is similar to real-life behaviour[70-75]. Children in particular have been found to experience high levels of immersion and an exhaustive sense of presence in IVR environments[76,77]. Here, presence refers to (a) a spatial perception of actually *being* in the virtual environment[77-79], and (b) a social perception of *being with another* in the virtual environment and a respective response to and/or interaction with virtual actors[79-81]. In addition, IVRs not only provide authentic and yet experimentally controlled research set-ups, modern head-mounted displays (HMDs) with integrated eye-tracking devices simultaneously make it possible to collect eye-tracking data non-intrusively and under standardized (lighting) conditions[82]. Consequently, IVRs make it easy to examine behavioural data such as pupillometry or visual attention as a complement to commonly used self-report measures[29,83,84].

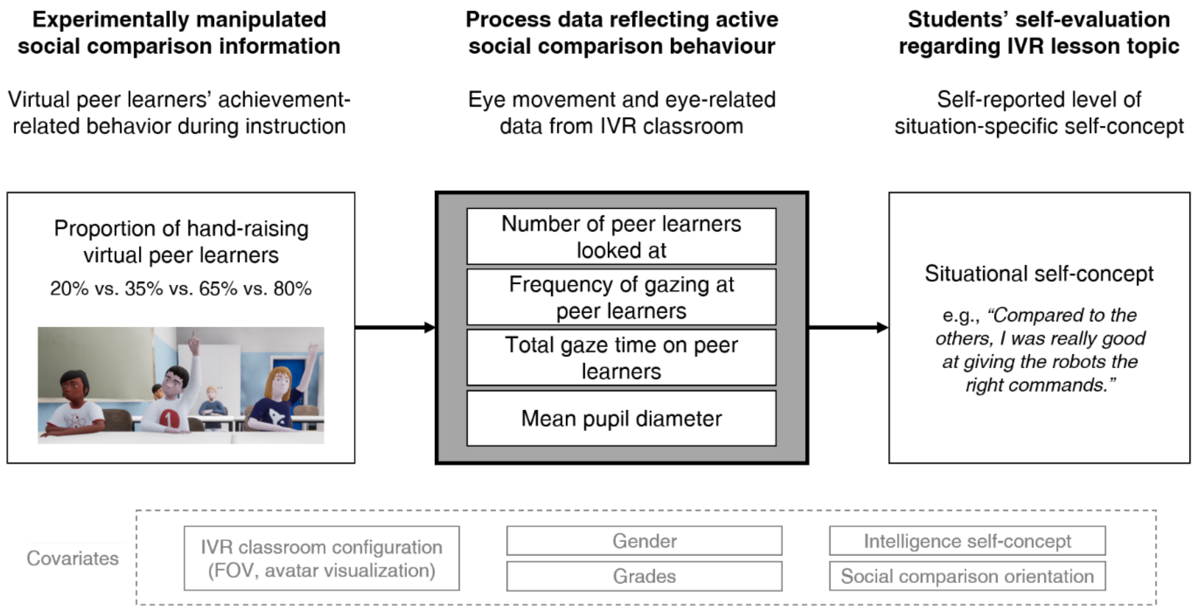


Fig. 2 | Summary of the theoretical structural model and experimental design. We used an IVR classroom with experimental variation of virtual peer learners' performance level (i.e., the proportion of students raising their hands; shown on the left) to systematically investigate social comparisons in the classroom. Using eye-tracking data from the IVR classroom, we selected eye movement features as behavioural indicators of social comparisons (i.e., visual attention to peer learners and mean pupil diameter; depicted in the middle) to determine to what extent students actively engage in social comparisons in the classroom. Lastly, we examined how these behavioural indicators of active social comparisons are related to students' self-evaluations (i.e., situational self-concept; described on the right). We included a number of covariates in the model to account for potential influencing factors in the social comparison context and in the specific IVR classroom configuration. Detailed information on the covariates is included in the Appendix. FOV = field of view.

Taken together, this study pursues two aims in order to advance research on social comparisons in the classroom (see Fig. 2): Firstly, to meet the need for an authentic and yet experimentally controllable set-up, we use an IVR classroom in which we manipulated social comparison information (i.e., virtual peer learners' achievement-related behaviour) to systematically examine social comparisons and respective effects on students' academic self-evaluations. We implemented four different performance levels for the virtual peer learners in the IVR classroom by systematically varying the proportion of virtual peer learners who actively participated and raised their hands to indicate that they knew the correct answer. More specifically, in the four conditions, 20% vs. 35% vs. 65% vs. 80% of the virtual peer learners exhibited high-achieving hand-raising behaviour (see Fig. 1). The participating students were told they would experience a simulation of a real-world classroom scenario in the IVR. Secondly, in order to gain insights into the underlying social comparison processes, we use eye-tracking data from the IVR classroom to examine (a) to what extent students actively engage in social comparisons, and therefore attend and respond to peer learners' achievement-related behaviour, and (b) how students' behavioural responses to the provided social comparison information (i.e., indicators of active social comparisons) are related to differences in situational self-concept. We used four eye-tracking features as indicators of active social comparison behaviour: The *number of peer learners looked at* (i.e., the extent to which students noticed and actively attended to social comparison information), the *frequency of gazing at peer learners* (i.e., how often students' visual attention shifted to social comparison information over the course of the lesson), the *total gaze time on peer learners* (i.e., how long students spent processing the social comparison information), and students' *mean pupil diameter* (i.e., reflecting cognitive and affective arousal associated with information processing).

Results

We first examined to what extent virtual peer learners' achievement-related behaviour (i.e. the experimental variation of hand-raising behaviour) affects how students attend and respond to this social comparison information provided in the IVR classroom. We investigated how the different hand-raising conditions (20% vs. 35% vs. 65% vs. 80% of students raising their hands and therefore engaging in high-achieving behaviour) impact students' visual attention to virtual peer learners (i.e., the number of peer learners looked at, the frequency of gazing at peers and the total gaze time on them) as well as students' mean pupil diameter (as an indicator of overall arousal associated with processing activities).

More visual attention to peer learners and greater pupil diameter in extreme hand-raising conditions. Whereas we expected higher levels of visual attention to the virtual classmates and increased pupil diameter as the proportion of peer learners raising their hands increased, we found the highest values for all four variables in the extreme conditions with 20% and 80% of students raising their hands (see Fig. 3a-d). This finding indicates that students were particularly likely to actively process their peers' implicit achievement-related (i.e., hand-raising) behaviour when the respective social comparison information could be clearly interpreted and a clear minority/majority of peer learners engaged in high-achieving behaviour.

We analysed the effects of the four hand-raising conditions on the four indicators of social comparisons in more detail via multiple regression models. We included a number of covariates in the models to account for (a) individual differences in the social comparison context (e.g., gender, individual competence beliefs) as well as (b) potential effects of the IVR classroom configuration (field of view, avatar visualization style) on the processing of social comparison information in the IVR classroom. Detailed information on all of the covariates is included in the Method section and in the Supplementary Appendices 1-5.

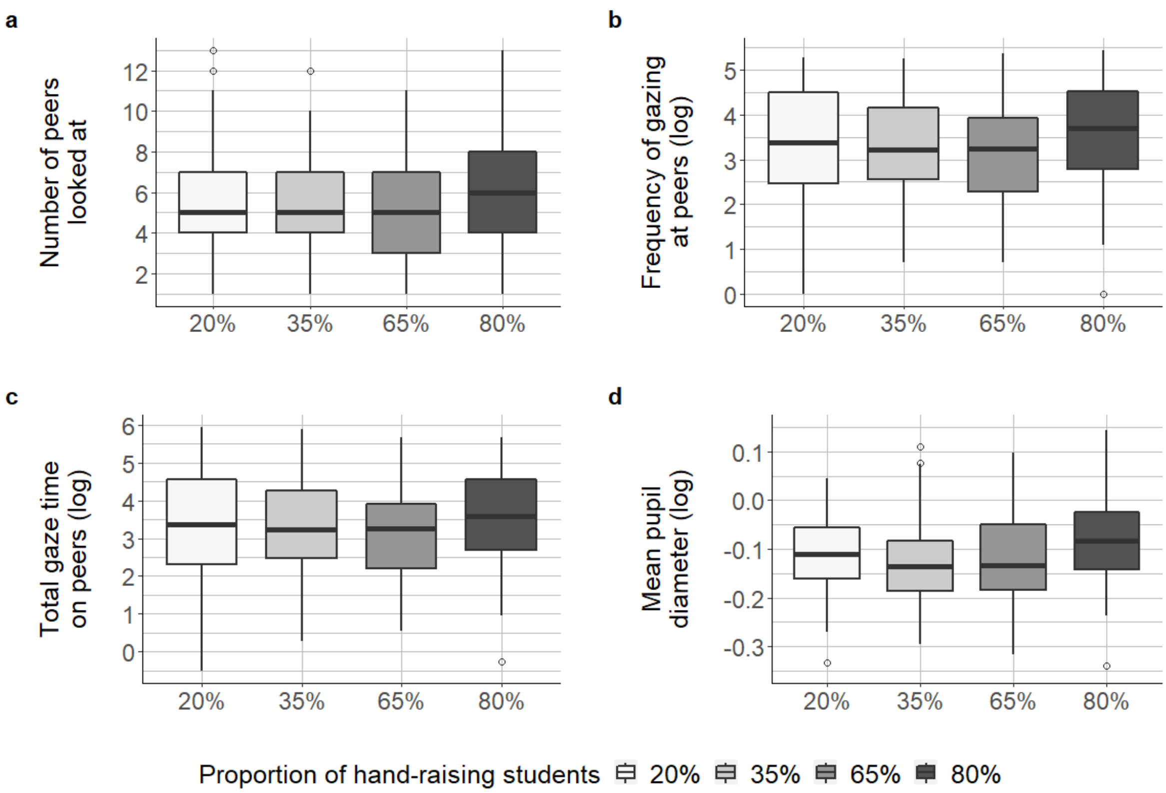


Fig. 3 | Boxplots for the frequency of gazing at peers, total gaze time on peers, and mean pupil diameter in the different hand-raising conditions. The proportion of hand-raising students refers to the experimental variation of peer learners' performance level via their hand-raising behaviour. Participants (N = 353) were randomly assigned to one of the four hand-raising conditions, with 20% (n = 92), 35% (n = 86), 65% (n = 85) and 80% (n = 90) of peer learners raising their hands, respectively. **a**, Number of peers looked at was highest for 20% (M = 5.56, SD = 2.81) and 80% (M = 6.10, SD = 2.64), with the highest number of peers looked at for 80% hand-raising and the mean values for 35% (M = 5.48, SD = 2.88) and 65% (M = 5.05, SD = 2.65) only slightly lower than for 20%. **b**, Frequency of gazing at peers (log) was highest for 20% (M = 3.38, SD = 1.19) and 80% (M = 3.53, SD = 1.10) hand-raising, with slightly lower frequencies for 35% (M = 3.23, SD = 1.15) and 65% (M = 3.09, SD = 1.11). **c**, Total gaze time on peers (log) was highest for 20% (M = 3.41, SD = 1.35) and 80% (M = 3.57, SD = 1.20) hand-raising, and slightly lower for 35% (M = 3.28, SD = 1.29) and 65% (M = 3.09, SD = 1.18). **d**, Mean pupil diameter (log) was highest for 20% (M = -0.12, SD = 0.12) and 80% (M = -0.05, SD = 0.14) hand-raising, with the mean value for the 80% condition considerably higher than for the 20% condition as well as the even lower 35% (M = -0.12, SD = 0.12) and 65% (M = -0.10, SD = 0.14) hand-raising conditions.

Proportion of hand-raising peer learners affects how many peer learners are looked at but not for how often or how long. We found different results for the number of peer learners looked at compared to the frequency and total time they were looked at, indicating that these eye movement features reflect different processing activities. With regard to the number of peer learners looked at (Fig. 3a, statistics for the full regression model in Supplementary Appendix 1), we found no significant difference between the 20% compared to 35% (b = 0.00, SE = 0.07, t = 0.02, P = 0.994) and 20% compared to 65% (b = -0.08, SE = 0.07, t = -1.04, P = 0.300) hand-raising conditions, but a significant difference between the 20% and 80% hand-raising conditions (b = 0.09, SE = 0.04, t = 2.06, P = 0.040, d = 0.20).

Notably, the number of peer learners looked at was descriptively lower in the 65% hand-raising condition compared to the 20% condition, but significantly higher in the 80% condition. In contrast, whereas the descriptive pattern of results was similar for the frequency of gazing at peer learners (Fig. 3b) and the total gaze time on peer learners (Fig. 3c), we found no statistically significant differences in how often (b = -0.03, SE = 0.03, t = -0.78, P = 0.434, and b = -0.09, SE = 0.08, t = -1.08, P = 0.279, and b = 0.05, SE = 0.03, t = 1.91, P = 0.056, for 20% compared to 35%, 65% and 80% hand-raising, respectively) and how long (b = -0.02, SE = 0.04, t = -0.40, P = 0.690, and b = -0.09, SE = 0.08, t = -1.02, P = 0.308, and b = 0.05, SE = 0.03, t = 1.70, P = 0.089, for 20% compared to 35%, 65% and 80% hand-raising, respectively) participants looked at the virtual peer learners in the different hand-raising conditions (see statistics for the full regression models in Supplementary Appendices 2 and 3). These findings indicate that the hand-raising conditions affected the extent to which students noticed and actively attended to their peers' behaviour (i.e., the *number of peers looked at*), but less so the intensity and time students spent processing the social comparison information (i.e., the *frequency of gazing at peer learners* and the *total gaze time on peer learners*).

Students show greater pupil diameter for extreme hand-raising levels of peer learners, particularly when the clear majority is high-achieving. With regard to the effects on *mean pupil diameter* (Fig. 3d), the differences between 20% and 35% hand-raising (b = -0.04, SE = 0.05, t = -0.78, P = 0.433) as well as between 20% and 65% hand-raising (b = 0.05, SE = 0.06, t = 0.98, P = 0.328) were not significant (see statistics for the full regression model in Supplementary Appendix 4). However, the 80% hand-raising condition led to a statistically significantly higher mean pupil diameter (b = 0.20, SE = 0.08, t = 2.45, P = 0.014, d = 0.52) compared to the 20% hand-raising condition, suggesting that a majority of high-achieving peers (i.e., 80% of students raising their hands) led to a considerably higher level of mental effort and arousal for participants when processing this information compared to when a minority of peers exhibited high-achieving behaviour.

Eye-tracking features as indicators of social comparison information show relations to situational self-concept. In addition to the effect of the experimental hand-raising conditions on students' visual attention and pupil diameter, we examined how these behavioural indicators of active social comparisons relate to differences in students' situational self-concept. In line with the theoretical assumptions underlying the BFLPE (i.e., social comparisons in the classroom lead to differences in individual academic self-concept), we expected students' situational self-concept to be related to visual attention to peer learners (i.e., number of peers looked at, frequency of gazing at peers and total gaze time on peers) as well as to associated mental effort and arousal (indicated by the mean pupil diameter). The results revealed the expected relations for all three indicators of visual attention to peer learners: We found a statistically significant negative effect on students' self-evaluations for the *number of peer learners looked at* (b = -0.13, SE = 0.05, t = -2.56, P = 0.010), the *frequency of gazing at peer learners* (b = -0.11, SE = 0.05, t = -2.06, P = 0.040), and the *total gaze time on peer learners* (b = -0.10, SE = 0.05, t = -2.27, P = 0.023); the *mean pupil diameter* was not related to differences in participants' situational self-concept (b = -0.07, SE = 0.04, t = -1.65, P = 0.100). Detailed statistics for the full regression models are provided in Supplementary Appendix 5. These findings suggest that the active processing of social comparison information—indicated by visual attention to peer learners—was in fact related to students' self-evaluations.

Lastly, we examined whether students' eye movements can explain the impact of classmates' hand-raising behaviour on individual learners' situational self-concepts. The results revealed that only the two 'extreme' hand-raising conditions were predictive of students' situational self-concept: In line with the BFLPE, the 80% hand-raising condition (i.e., higher-achieving peer learners) led to a statistically significantly lower situational self-concept than the 20% hand-raising condition (b = -0.12, SE = 0.04, t = -2.60, P = 0.009, d = 0.25). Notably, this negative effect of the experimental hand-raising conditions on situational self-concept (i.e., the 80% hand-raising condition resulting in a statistically significantly lower self-concept compared to the 20% hand-raising condition) remained statistically significant for all three models, with students' situational self-concept being predicted by the eye movement features, specifically the number of peer learners looked at (b = -0.10, SE = 0.04, t = -2.42, P = 0.015), the frequency of gazing at peer learners (b = -0.11, SE = 0.05, t = -2.46, P = 0.014), the total gaze time on peer learners (b = -0.11, SE = 0.04, t = -2.50, P = 0.012), and the mean pupil diameter (b = -0.10, SE = 0.05, t = -2.17, P = 0.030). Full statistics for the regression models are provided in the Supplementary Appendix 5. Hence, our results indicate two types of effects on students' self-concept: (a) a general psychological effect of classmates' achievement-related behaviour in the classroom (i.e., the experimentally manipulated social comparison information affected students' self-evaluation) and (b) a differential effect of interindividually different processing of social comparison information (i.e., students' active social comparison behaviour—indicated by their visual attention to peer learners and mean pupil diameter—was related to their situational self-concept). We will discuss these findings in more detail below.

Discussion

Consequences of social comparisons in real-world classrooms are well-known (see, e.g., research on the BFLPE[11-14]). However, how exactly such social comparisons proceed—i.e., how students make use of social (comparison) information in the classroom—has so far remained unclear. Therefore, to answer the question of to what extent students in fact actively engage in social comparisons with their peer learners, we used an IVR classroom as a standardized yet authentic research setting with experimental variation of peer learners' achievement-related behaviour (i.e., different proportions of peers who raised their hands). Moreover, we used eye-tracking data (i.e., students' visual attention towards and pupillary response to their peers) to examine (a) to what extent students actively engage in social comparisons, and therefore attend and respond to peer learners' achievement-related behaviour, and (b) how students' eye movements, as indicators of social comparison processes, are related to differences in situational self-concept. We found that (a) the different levels of hand-raising behaviour had an effect on students' visual attention towards their peer learners and their mean pupil diameter and that (b) markers of students' visual attention to their peers were related to students' situational self-concept. Fig. 4 gives an overview of the results, which we discuss in more detail below.

Overall, our findings provide experimental evidence that students actively engage in social comparisons in the classroom. Whereas previous research has consistently highlighted the importance of social comparisons for students' self-evaluations, it has not answered the question of whether students do in fact *actively* compare themselves to their peer learners or whether they are rather affected by comparisons based on grades or comments by their peers or teachers. Our results based on behavioural indicators of social comparisons indicate that participants did actively attend and respond to their virtual peers' hand-raising behaviour in the IVR classroom. Moreover, our findings provide evidence for both general psychological effects of the experimental conditions (i.e., peer learners' achievement-related behaviour) as well as differential effects of social information processing.

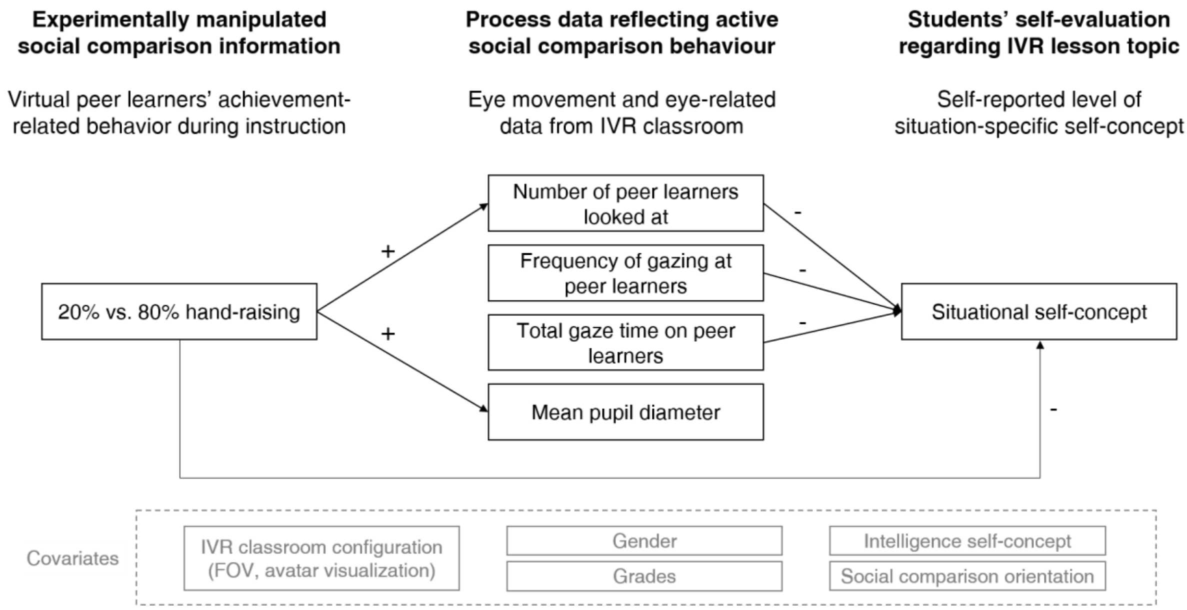


Fig. 4 | Overview of the revealed effects summarized in one structural model. A summary of all statistically significant relations from the different statistical models are depicted (see detailed statistics including the covariates in Supplementary Appendices 1–5). FOV = field of view.

Firstly, with regard to general psychological effects of the experimentally manipulated hand-raising behaviour of peer learners, we found that visual attention to peer learners (i.e., how many different virtual classmates were looked at as well as how often and for how long virtual classmates were looked at in general) and mean pupil diameter were greatest in the extreme hand-raising conditions of 20% and 80%. The differences were not always statistically significant, yet these descriptive statistics indicate that students' behavioural responses to their peer learners' hand-raising behaviour do not simply reflect the amount of activity happening in the classroom (i.e., more hand-raising) but rather the amount of social information processed. In other words, very low or very high performance by peer learners (i.e., a large minority or majority raising their hands) seems to provide more social comparison information to students compared to more moderate levels of peers' hand-raising (i.e., 35% and 65% of students raising their hands). This is in line with psychological research on social comparisons suggesting that contrastive social comparison effects (i.e., negative self-evaluations in response to a high-performing reference group) are more likely when the comparison information is more extreme and unambiguous[85,86]. Taking a closer look at the 80% compared to the 20% condition, we did in fact find the expected pattern of more hand-raising behaviour leading to a greater number of peer learners looked at and greater pupil diameter. Based on these results, it can be argued that students notice and respond more to extreme and clearly interpretable social comparison information, which can in turn be considered a prerequisite for respective responses, reflected in effects on self-evaluations. It is thus not surprising that we found significant direct effects of the hand-raising conditions on students' situational self-evaluations when comparing the extreme conditions with a clear minority/majority of classmates raising their hands.

Secondly, with regard to differential effects of social information processing, our results suggest that actively engaging in social comparisons can be observed at two levels that are differentially influenced by situational and interindividual differences. Whereas the *number of peer learners looked at* and *mean pupil diameter* differed significantly between the hand-raising conditions, the *frequency of gazing* and the *total gaze time on* peer learners were not affected by the proportion of hand-raising classmates. Notably, the number of peer learners looked at and mean pupil diameter are argued to reflect a simple 'noticing' of social comparison information and a respective affective arousal[44,55] in response to social comparison information (particularly in light of the fact that the content of the present IVR lesson was not particularly difficult for students; see details in the Method section). Taking notice of more peer learners and increased affective arousal might be due to the fact that the provided social comparison information (especially the 20% and 80% hand-raising conditions) is particularly conspicuous. For instance, pupil diameter has repeatedly been found to be higher when people viewed emotionally pleasant or unpleasant information[47,49-52], with particularly negative or threatening stimuli leading to increased pupillary responses[56,87,88]. In turn, the frequency of gazing at peer learners and the time spent looking at peer learners indicate a deeper (and possibly more wilful) level of processing for this information[40-42]. Considering the effects of the different hand-raising conditions on the number of peers looked at and the mean pupil diameter, but not the frequency of gazing or total gaze time, we argue that situation-specific social comparison information is noticed and responded to by students, but the extent to which students process this information seems to differ depending on factors unrelated to the situation (such as individual competence beliefs or social orientation; see Supplemental Appendices 1-4).

With regards to effects on situational self-concept, the results revealed no relation to students' mean pupil diameter during instruction, but significant relations to all three indicators of visual attention to peer learners (i.e., number of peers looked at, frequency of gazing at peers, and total gaze time on peers); the higher the number of peers looked at and the more often or the longer students looked at their virtual classmates on average, the lower their situational self-concept. In other words, regardless of whether students' visual attention was (at least partially) driven by their peers' hand-raising behaviour (i.e., the number of peers looked at) or by their individual need to process social information[89], students who actively attended more to their peers' performance—presumably, because they really cared about or were otherwise affected by it—did evaluate themselves as worse. While individual differences in social information processing tend to be inherent to the social psychological understanding of social comparison processes[86,89,90], there is a debate within educational psychological research regarding potential moderators of the BFLPE[91-94]. In order to extend the present study's findings, we argue that individual characteristics are worth investigating further in future research as influencing factors of social comparisons and their effects.

Turning to future research perspectives based on our findings, we would like to highlight that eye-tracking data from IVR environments provides a promising avenue for gaining insights into processes like social comparisons in complex and dynamic environments. In the present study, we selected and aggregated four markers allowing us to gain insights into the processing of social comparison information in a classroom (i.e., the number of peers looked at, frequency of gazing at peer learners, total gaze time on peer learners, and mean pupil diameter). In order to extend the insights of the present study, we suggest that future studies consider additional eye movement features and how these behavioural indicators develop over time. For instance, it would be interesting to see whether there are certain peer learners that students' visual attention keeps returning to or whether students follow the gaze of the teacher and increasingly look at (and consequently compare themselves to) students that are the teacher's focus of attention.

In conclusion, we were able to extend existing research on the BFLPE and provide experimental support for the role of active social comparisons during instruction by using eye movement data from an IVR classroom. In line with the claim that new technologies allow researchers to bridge the gap between "experimental and methodological rigorousness on the one hand, and the complexity and uncontrollable nature of an authentic classroom full of pupils"[95], we see IVR as a tool allowing us to advance research on social comparisons and similar phenomena in classrooms and beyond. We believe that the approach presented in this study provides an important foundation for future work to extend these insights and apply them to other topics as well.

Methods

This research complies with ethical standards of research with human subjects, confirmed by the ethics committee at the University of Tübingen (date of approval: November 25, 2019, file number: A2.5.4-106_aa). Regional educational authorities approved the study and the data collection, and we obtained written informed consent from both the participating students and their parents or legal guardians prior to students' participation in the study.

Participants. We recruited N = 381 students in Grade 6 from local academic-track schools via e-mails and invitation letters. To determine the required sample size, we computed an a priori power analysis considering existing findings from experimental studies [96,97]. Since we expected our manipulation to be less salient and effects on behavioural responses less powerful than in these studies, we assumed small to medium effects (f = .20). Based on this, a necessary sample size of $n = 90$ students in each of the four hand-raising conditions was determined for the respective analyses of variance (for two-tailed tests with a .05 alpha level and a minimum power of .90). Due to technical issues during data collection (i.e., visual or audio issues with the HMDs during the IVR experience), data from 28 participants had to be excluded from the analyses. The cleaned sample consisted of N = 353 students (M_{Age} = 11.52 years, SD_{Age} = 0.55; 46.7% girls).

Content and course of the IVR lesson. The IVR lesson's content was adapted from tested and evaluated materials from a course designed to teach kids basic computational thinking skills[98,99]. More specifically, the students learned about the meaning of coding and sequences and loops as basic computational concepts. They also worked on two exercises testing their use of the concepts "sequence" and "loop" (adapted from the Computational Thinking test[100]). The students' self-reports indicated that they found the lesson easy to follow (perceived difficulty assessed with 10 items on a 4-point rating scale, with higher values indicating higher difficulty, yielded a mean value of M = 1.38, SD = .42; Cronbach's alpha .86). We chose this topic because even though computational thinking is regarded as a key 21st-century competence, it is not commonly included in the curricula of primary or lower secondary schools. We therefore expected that (a) the participating students would have little to no prior knowledge of this subject matter and (b) we could investigate social comparisons in an unbiased context largely independently of students' previous experiences in this subject. The entire IVR lesson took place in a simulated classroom showing a typical teaching situation with explanations by the teacher, dialogue between the teacher and the virtual students, and working on exercises independently. We used audio recordings and motion captures stemming from a real classroom to ensure that the pace and content of the virtual students' answers as well as their movements were calibrated to be typical of sixth graders. The IVR experience was designed and rendered using the Unreal Game Engine v4.23.1. Fig. 1 shows the design of the virtual classroom.

Configuration and design of the IVR classroom. We systematically varied the performance level of the IVR class and therefore manipulated the virtual classmates' hand-raising behaviour (i.e., the number of students raising their hands in response to the teacher's questions or indicating that they knew the correct solution to a task). The virtual classmates' hand-raising behaviour was manipulated on four levels, with 20% vs. 35% vs. 65% vs. 80% of students raising their hands and showing high-performing participation.

In addition to the hand-raising conditions, our experimental design included different IVR classroom configurations with regard to the participant's sitting position in the IVR classroom and the virtual avatars of peer learners and the teacher. We varied participating students' sitting position on two levels and allocated them to either the front (i.e., second of four rows) or the back (i.e., fourth of four rows) of the IVR classroom. The virtual classmates' and teacher's avatars were varied on two levels as well and were designed in a more cartoon-like or more realistic manner. We added these configuration conditions because research does not yet provide clear answers as to how an IVR classroom should be programmed to provide ideal conditions as an experimental tool and we wanted to make sure to account for potential factors influencing how the social information we provided in the IVR classroom was perceived. Hence, we implemented a total of sixteen ($4 \times 2 \times 2$) configuration conditions to which participants were randomly assigned (see Fig. 4 for an illustration). Participants' allocation to the different sitting positions and avatar representations was counterbalanced with respect to the main manipulation conditions of different hand-raising behaviours.

We checked for participants' perception of an authentic IVR classroom experience via self-reports. Therefore, we assessed participants' level of experienced (spatial and social) presence in the IVR classroom with nine items (e.g., "I felt like I was sitting in the virtual classroom" or "I felt like the teacher in the virtual classroom really addressed me") based on common conceptualizations of presence[78,79]. Moreover, we asked participants to rate the degree of realism of the IVR lesson with six items that were developed specifically to assess students' perceptions of the IVR classroom situation (e.g., "What I experienced in the virtual classroom could also happen in a real classroom" or "The students in the virtual classroom behaved similarly to real classmates"). The self-reports for experienced presence and perceived realism indicated that the IVR environment was experienced as authentic overall. Both variables were rated on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*) and had acceptable Cronbach's alpha values of 0.77 and 0.78, respectively. The reported mean levels of experienced presence and perceived realism ranged from 2.82 to 2.97 ($0.52 < SD$s < 0.62) in all configuration conditions. None of the configuration conditions had a statistically significant effect on participants' experienced level of presence or perceived realism (all p-values > 0.05).

Experiment procedure. The experiment took place in a quiet room at the participants' school and students participated in groups of up to 10. Before the beginning of each test session, head-mounted displays (HMDs) were set up for each participant. We used the HTC Vive Pro Eye HMD in our experiments. The researchers randomly assigned one of the experimental conditions to each set-up HMD by means of random number generation. Students were then allowed to enter the testing room and were free to choose any seat without knowing the experimental conditions (they were debriefed in detail after they had completed the experiment). All testing sessions followed the same procedure, and the experimental conditions differed only with regard to specific manipulations in the IVR classroom scenario that participants experienced.

First, participants filled out the first part of a paper-based questionnaire that included demographics and basic personality characteristics as well as learning background (i.e., prior experience with the lesson topic and IVR). Second, participants put on the HMDs and were helped to calibrate the included eye trackers. Upon successful calibration of the eye trackers, participants experienced the IVR lesson (which lasted about 15 minutes). Participants all began the IVR lesson at the same time and were instructed to behave as they would in a normal classroom situation (e.g., look around, raise their hands) while remaining seated and quiet. Third, as soon as the participants finished the IVR lesson, they completed the second part of the questionnaire (including measures of self-concepts, experienced presence in the IVR and perceived realism of the IVR classroom), followed by a debriefing. In total, each test session took approximately 45 minutes, including all instructions and preparation, and was supervised by research assistants throughout.

Eye-tracking measures and data pre-processing. To collect eye movement data, we used the Tobii eye tracker integrated into the HTC Vive Pro Eye head-mounted display (HMD). The HMD has a refresh rate of 90 Hz and field of view of 110° (screen resolution 1440 × 1600), and the integrated Tobii eye tracker runs at a 120 Hz sampling rate. Before the start of the IVR lesson, we calibrated the eye tracker based on a 5-point calibration for each participant. During the experiments (duration approximately 850 seconds), continuous measures of HMD orientation, gaze, and eye-related data were collected and assigned to participants via an anonymous identifier. Below, we introduce our central markers based on this data.

With regards to eye-related data, pupil diameter was recorded in millimetres on a millisecond basis. In the course of data pre-processing, we smoothed and normalized the pupil diameter measures using the Savitzky-Golay filter[101] and divisive baseline correction with a baseline duration of approximately 1 second from an interval at the beginning of the experiment[102]. For the purpose of the present study, we averaged the measure across the whole IVR experience and used the mean pupil diameter as an indication of participants' arousal and mental effort[103]. The normalized mean pupil diameter ranged between 0.71 and 1.97 (M = 0.92, SD = 0.14). Since the mean pupil diameter was non-normally distributed (skewness of 3.43, SE = 0.15; kurtosis of 19.95, SE = 0.29), we log transformed the variable for all analyses.

With regards to participants' visual attention, we defined virtual peer learners as the objects of interest (OOIs). Since raw data reported by the eye tracker can be affected negatively by blinks or noisy sensor readings, we first applied a linear polynomial interpolation of degree one to clean the gaze data and account for missing values. Using head pose and gaze data, we applied ray-casting[35,104] to map the gaze into the 3D virtual environment. Calculating the intersections between predefined colliders of the OOIs with the gaze vectors allowed us to identify when participants looked at the OOIs (further details on the ray-casting procedure to identify objects of gaze in the IVR are described in the corresponding subsection below). Considering that objects of gaze may not directly represent visual attention, as participants can unconsciously gaze at an OOI for a very short time when looking around, we set an attention threshold of at least 500 milliseconds to count OOIs. We obtained similar trends across different thresholds tested and chose the selected threshold as a conservative estimate that is larger than classical fixation thresholds applied for both conventional[39] or VR eye-tracking[105] setups. We used the resulting information about the object of gaze to calculate (a) how many different peers were looked at, (b) how often the gaze shifted to peer learners throughout the lesson and (c) how long participants looked at peer learners in total.

To calculate (a) how many different peer learners were looked at, we used the attention threshold of 500 milliseconds to identify different objects of gaze and used this to sum up the *number of peer learners looked at.* We therefore counted each peer learner that was the object of participants' gaze for at least the attention threshold of 500 milliseconds, regardless of how often participants' gaze rested on the respective OOI. There were a total of 24 peer learners in the IVR classroom; the actual number of peers looked at ranged between 1 and 13 (M = 5.58, SD = 2.76).

With regards to (b) how often participants looked at our OOIs, we summed up *the frequency of gazing at peer learners* across the whole VR experiment and used the sum score reflecting total number of gaze shifts towards virtual classmates. Virtual peer learners were counted as 'gazed at' if they were the object of gaze for at least the attention threshold of 500 milliseconds. We counted a new ‘gazed at peer learner’ as soon as the gaze shifted to a virtual peer learner from any other object (i.e., the frequency of gazing at peer learners reflects how often the gaze shifted to a peer learner throughout the IVR lesson). The frequency of gazing at peer learners ranged between 1 and 230 (M = 47.63, SD = 46.45). Frequency of gazing at peers was non-normally distributed (skewness of 1.41, SE = 0.15; kurtosis of 1.66, SE = 0.29); thus, we used a log transformation of the variable for all analyses.

To reflect (c) how long participants looked at our OOIs, we summed up the *gaze time on peer learners* across the whole VR experiment—again, counting only intervals longer than the attention threshold of 500 milliseconds as attention towards a peer—and calculated the sum score reflecting the total gaze time on virtual peer learners. The total gaze time on virtual peer learners ranged between 0.60 and 382.49 seconds (M = 56.10, SD = 65.11). Gaze time on virtual peer learners was non-normally distributed (skewness of 2.03, SE = 0.15; kurtosis of 4.94, SE = 0.29); thus, we log transformed the variable for all analyses.

Ray-casting to identify objects of gaze in the IVR. To identify the object of gaze, we implemented an algorithm to apply ray-casting[35,104], a technique typically used to calculate gaze points from eye-tracking devices[82]. The idea of ray-casting is to forward a persons’ gaze vector, given in 3-D coordinates, and to calculate which object the gaze hits in a 3-D space[106]. Ray casting was performed for every measured time point during the virtual experiment (i.e., on average every 24 milliseconds depending on hardware and software performance). Using predefined functions from the Unreal Engine Blueprint (SetActorLocation, SetActorRotation and GetRotationXVector)[107], we simulated each participants’ position and orientation for each timepoint and calculated the vector orthogonal to the screen surface of the HMD stating the participants’ head direction, adjusted to their position and orientation as measured by the HMD device. Since this orthogonal vector reflects the head but not gaze direction, we needed to rotate the vector to reflect participants’ actual gaze direction in relation to the fixed coordinate system of all objects in the IVR environment. Notably, in an IVR set-up like the present study, the eye-tracking device is not stationary like in traditional eye-tracking experiments but part of the HMD, which can be moved 360 degrees in the virtual space. Therefore, head movements need to be taken into consideration to adequately process eye-tracking data. We used pitch (i.e., the angle at which one is looking up or down) and yaw (i.e., the rotation of one’s head left or right from a vertical axis) as two markers describing orientation in the 3-D space. We then calculated the angles between the normalized gaze vector and the x vector—both given in local coordinates based on participants’ orientation—to rotate the x vector.

Based on the general calculation of an angle (in degrees) of two vectors (v_1, v_2) as

$$\alpha = \arccos\left(\frac{v_1 \cdot v_2}{|v_1| \cdot |v_2|}\right) \cdot \frac{180}{\pi}$$

we calculated the yaw rotation as the angle between the x vector $x = (1\ 0\ 0)^T$ and the flat gaze vector $g_{flat} = (g_1\ g_2\ 0)^T$ and the pitch rotation as the angle between the gaze vector $g = (g_1\ g_2\ g_3)^T$ and g_{flat}. In order to perform the ray-casting, we additionally needed to extract information about the surface of an object to ultimately calculate when it would be hit. We therefore added colliders to the objects in the IVR (i.e., an invisible mesh grid that approximates the shape of an object and describes its surface) which can be used to detect gaze hit. To obtain the object of gaze across the full experiment session, we applied ray-casting frame by frame for the entire IVR lesson for each participant. We counted a gaze hit on an object if

$$\{v_{location} + (v_{direction} \cdot k) | k \in \mathbb{R}\} \cap s_{object} \neq \emptyset$$

where $v_{location} \in \mathbb{R}^3$ describes the coordinates of a person's eye location, $v_{direction} \in \mathbb{R}^3$ describes the normalized combined gaze direction (i.e., the equidistant line between the gaze direction of the left and right eyes) and s_{object} reflects the set of coordinates describing the surface of the object. We used the LineTraceByChannel function from the Unreal Engine Blueprint[107], which outputted the name of the object hit by the ray-cast gaze vector for each frame. Based on these values, we then calculated the object of interest information (i.e., frequency of gazing at virtual classmates, total gaze time on virtual classmates). Notably, in our case, only gaze and head information were collected during the experiment, and we applied the ray-casting algorithm afterwards using a C++ script to map the collected eye-tracking data onto the Unreal Engine while re-running the entire IVR lesson for each participant according to the tracked time stamps.

Self-report measures of students' self-concept. We assessed participants' situational self-concept after the IVR lesson with respect to the specific experience with virtual classmates in the IVR classroom. The self-concept scale consisted of four items (e.g., "Compared to the others, I was really good at giving the robots the right commands"), of which two were reverse-scored and recoded accordingly. The items were based on commonly used wording for self-reports of academic self-evaluations[108] and adapted to be situation- and domain-specific. A 4-point rating scale was applied ranging from 1 (*not true at all*) to 4 (*absolutely true*); the scale had an acceptable Cronbach's α of 0.71.

Covariates. To account for potentially relevant covariates in the social comparison context, we asked teachers for participants' latest grades in maths and German (as a proxy of academic achievement). Moreover, we assessed participants' prior interest in the topic, their intelligence self-concept and their social orientation in the paper-based pretest questionnaire. Prior interest in the topic of computational thinking was assessed with five items (e.g., "I would like to know more about how computer programs or robots work"). Intelligence self-concept was measured with four items (e.g., "I often think I'm not as smart as the others")[108]. Social orientation was assessed with seven items (e.g., "I pay close attention to how I do things compared with my classmates")[89]. All three scales were rated on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*) and had acceptable Cronbach's α values of 0.91, 0.72 and 0.74, respectively.

Regression Analyses. Using the processed and accumulated eye-tracking data, we calculated multiple regression analyses to examine to what extent the experimental variation of virtual peer learners' hand-raising behaviour affected students' pupillary response to and visual attention towards the social comparison information (Research Question 1) as well as how these behavioural responses to the provided social comparison information were related to students' situational self-concept (Research Question 2). We calculated separate models for each of the outcome variables. Prior to the analyses, all independent and dependent continuous variables were z-standardized, and categorical variables were dummy coded. To account for the fact that each testing group consisted of students within the same school, we controlled for cluster effects by using a school variable in all analyses (number of clusters $N = 14$). We added a number of background variables to the regression models to take potential confounding variables into account (see ref. [109] for suggestions for randomized research designs). All models were calculated in Mplus 8.2, using full information maximum likelihood estimation for missing values[110]. As we report standardized regression coefficients, these can be interpreted as effect sizes. We additionally calculated Cohen's *d* for standardized mean differences of dummy-coded categorical variables, whereby values < 0.20 indicate small, values < 0.50 medium-sized, and values > 0.80 large effects[111]. Hypotheses were tested with two-tailed tests with a critical *p*-value and confidence intervals set at an alpha level of 0.05.

We provide access to all data and data analysis scripts including the data pre-processing steps on the Open Science Framework (OSF) under the following link: https://osf.io/xbqg7/?view_only=aee93d91ea634c84a12e40a43e23f6e7.

References

1 Festinger, L. A theory of social comparison processes. *Hum. Relat.* **7**, 117–140, doi:10.1177/001872675400700202 (1954).

2 Wheeler, L. Motivation as a determinant of upward comparison. *J. Exp. Soc. Psychol.* **1**, 27–31, doi:10.1016/0022-1031(66)90062-X (1966).

3 Marsh, H. W. & Shavelson, R. Self-concept: Its multifaceted, hierarchical structure. *Educ. Psychol.* **20**, 107–123, doi:10.1207/s15326985ep2003_1 (1985).

4 Marsh, H. W. & O'Mara, A. Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Pers. Soc. Psychol. Bull.* **34**, 542–552, doi:10.1177/0146167207312313 (2008).

5 Göllner, R., Damian, R. I., Nagengast, B., Roberts, B. W. & Trautwein, U. It's not only who you are but who you are with: High school composition and individuals' attainment over the life course. *Psychol. Sci.* **29**, 1785–1796, doi:10.1177/0956797618794454 (2018).

6 Nagengast, B. & Marsh, H. W. Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *J. Educ. Psychol.* **104**, 1033–1053, doi:10.1037/a0027697 (2012).

7 Trautwein, U. & Möller, J. in *Psychosocial skills and school systems in the twenty-fi rst century: Theory, research, and applications* (eds Anastasiya A. Lipnevich, Franzis Preckel, & R. Roberts) 187–214 (Springer, 2016).

8 Fang, J. *et al.* The big-fish-little-pond effect on academic self-concept: A meta-analysis. *Front. Psychol.* **9**, 1569, doi:10.3389/fpsyg.2018.01569 (2018).

9 Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P. & van der Zee, Y. G. Social comparison in the classroom: A review. *Rev. Educ. Res.* **78**, 828–879, doi:10.3102/0034654308321210 (2008).

10 Marsh, H. W. The big-fish-little-pond effect on academic self-concept. *J. Educ. Psychol.* **79**, 280–295, doi:10.1037/0022-0663.79.3.280 (1987).

11 Marsh, H. W. & Hau, K. T. Big-fish-little-pond effect on academic self-concept. A cross-cultural (26-country) test of the negative effects of academically selective schools. *Am. Psychol.* **58**, 364–376, doi:10.1037/0003-066x.58.5.364 (2003).

12 Seaton, M., Marsh, H. W. & Craven, R. G. Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *J. Educ. Psychol.* **101**, 403–419, doi:10.1037/a0013838 (2009).

13 Nagengast, B. & Marsh, H. W. The negative effect of school-average ability on science self-concept in the UK, the UK countries and the world: the Big-Fish-Little-Pond-Effect for PISA 2006. *Educ. Psychol. (Lond)* **31**, 629–656, doi:10.1080/01443410.2011.586416 (2011).

14 Marsh, H. W., Trautwein, U., Lüdtke, O. & Köller, O. Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *J. Educ. Psychol.* **100**, 510–524, doi:10.1037/0022-0663.100.3.510 (2008).

15 Lüdtke, O., Köller, O., Marsh, H. W. & Trautwein, U. Teacher frame of reference and the big-fish-little-pond effect. *Contemp. Educ. Psychol.* **30**, 263–285, doi:10.1016/j.cedpsych.2004.10.002 (2005).

16 Trautwein, U. & Lüdtke, O. The big-fish-little-pond effect: Future research questions and educational implications. *Z. Pädagog. Psychol.* **19**, 137–140, doi:10.1024/1010-0652.19.3.137 (2005).

17 Levine, J. M. in *Teacher and student perceptions: Implications for learning* (eds J. M. Levine & M. C. Wang) 29–55 (Lawrence Erlbaum, 1983).

18 Marsh, H. W. *et al.* The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educ. Psychol. Rev.* **20**, 319–350, doi:10.1007/s10648-008-9075-6 (2008).

19 Dai, D. Y. & Rinn, A. N. The big-fish-little-pond effect: What do we know and where do we go from here? *Educ. Psychol. Rev.* **20**, 283–317, doi:10.1007/s10648-008-9071-x (2008).

20 Doyle, W. in *Handbook of classroom management: Research, practice and contemporary issues* (eds C. Evertson & C. Weinstein) 97–125 (Erlbaum, 2006).

21 Blanton, H., Gibbons, F. X., Buunk, A. P. & Kuyper, H. When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *J. Pers. Soc. Psychol.* **76**, 420–430, doi:10.1037/0022-3514.76.3.420 (1999).

22 Huguet, P., Dumas, F., Monteil, J. M. & Genestoux, N. Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its benficial impact on performance. *Eur. J. Soc. Psychol.* **31**, 557–578, doi:10.1002/ejsp.81 (2001).

23 Mussweiler, T. & Strack, F. The "relative self": Informational and judgmental consequences of comparative self-evaluation. *J. Pers. Soc. Psychol.* **79**, 23–38, doi:10.1037/0022-3514.79.1.23 (2000).

24 Pyszczynski, T., Greenberg, J. & LaPrelle, J. Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. *J. Exp. Soc. Psychol.* **29**, 195–211, doi:10.1016/0022-1031(85)90015-0 (1985).

25 Seaton, M. *et al.* In search of the big fish: Investigating the coexistence of the big-fish-little-pond effect with the positive effects of upward comparisons. *Br. J. Soc. Psychol.* **47**, 73–103, doi:10.1348/014466607X202309 (2008).

26 Suls, J. & Wheeler, L. in *Handbook of Social Comparison* (eds J. M. Suls & L. Wheeler) 3–19 (Springer, 2000).

27 Dumas, F., Huguet, P., Monteil, J. M., Rastoul, C. & Nezlek, J. Social comparison in the classroom: Is there a tendency to compare upward in elementary school? *Curr. Res. Soc. Psychol.* **10**, 166–187 (2005).

28 Mussweiler, T., Rüter, K. & Epstude, K. The ups and downs of social comparison: Mechanisms of assimilation and contrast. *J. Pers. Soc. Psychol.* **87**, 832–844, doi:10.1037/0022-3514.87.6.832 (2004).

29 Baumeister, R. F., Vohs, K. D. & Funder, D. C. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspect. Psychol. Sci.* **2**, 396–403, doi:10.1111/j.1745-6916.2007.00051.x (2007).

30 Holmqvist, K. *et al. Eye tracking: A comprehensive guide to methods and measures*. (Oxford University Press, 2011).

31 Jarodzka, H., Homqvist, K. & Gruber, H. Eye tracking in educational science: Theoretical frameworks and research agenda. *J. Eye Mov. Res.* **10**, 1–18, doi:10.16910/jemr.10.1.3 (2017).

32 Hutmacher, F. Why is there so much more research on vision than on any other sensory modality? *Front. Psychol.* **10**, 2246, doi:10.3389/fpsyg.2019.02246 (2019).

33 Lodge, J. & Harrison, W. The role of attention in learning in the digital age. *Yale J. Biol. Med.* **92**, 21–28 (2019).

34 Jacobs, R. J. Visual resolution and contour interaction in the fovea and periphery. *Vision Res.* **19**, 1187–1195, doi:10.1016/0042-6989(79)90183-4 (1979).

35 Pietroszek, K. in *Encyclopedia of Computer Graphics and Games* Ch. 180-1, 1–3 (2018).

36 Carrasco, M. Visual attention: The past 25 years. *Vision Res.* **51**, 1484–1525, doi:10.1016/j.visres.2011.04.012 (2011).

37 Kübler, T. C., Rothe, C., Schiefer, U., Rosenstiel, W. & Kasneci, E. SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behav. Res. Methods* **49**, 1048–1064, doi:10.3758/s13428-016-0765-6 (2017).

38 Mudrick, N. V., Azevedo, R. & Taub, M. Integrating metacognitive judgments and eye movements using sequential pattern mining to understand processes underlying multimedia learning. *Comput. Human Behav.* **96**, 223–234, doi:10.1016/j.chb.2018.06.028 (2019).

39 Salvucci, D. D. & Goldberg, J. H. Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 71–78, doi:10.1145/355017.355028 (2000).

40 Mason, L., Pluchino, P., Tornatora, M. C. & Ariasi, N. An eye-tracking study of learning from science text with concrete and abstract illustrations. *J. Exp. Educ.* **81**, 356–384, doi:10.1080/00220973.2012.727885 (2013).

41 Schüler, A. Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration. *Learn. Instr.* **49**, 218–231, doi:10.1016/j.learninstruc.2017.03.001 (2017).

42 Antonietti, A., Colombo, B. & Di Nuzzo, C. Metacognition in self-regulated multimedia learning: Integrating behavioural, psychophysiological and introspective measures. *Learn. Media Technol.* **40**, 187–209, doi:10.1080/17439884.2014.933112 (2014).

43 Binda, P. & Gamlin, P. D. Renewed attention on the pupil light reflex. *Trends Neurosci.* **40**, 455–457, doi:10.1016/j.tins.2017.06.007 (2017).

44 Bernhardt, P. C., Dabbs, J. M. & Riad, J. K. Pupillometry system for use in social psychology. *Behav. Res. Methods Instrum. Comput.* **28**, 61–66, doi:10.3758/BF03203637 (1996).

45 Appel, T., Scharinger, C., Gerjets, P. & Kasneci, E. Cross-subject workload classification using pupil-related measures. *ETRA '18: 2018 Symposium on Eye Tracking Research and Applications*, 1–8, doi:10.1145/3204493.3204531 (2018).

46 Appel, T. *et al.* Predicting cognitive load in an emergency simulation based on behavioral and physiological measures. *2019 International Conference on Multimodal Interaction (ICMI '19)*, 154–163, doi:10.1145/3340555.3353735 (2019).

47 Granholm, R. F., Asarnow, A. J., Sarkin, A. J. & Dykes, K. L. Pupillary responses index cognitive resource limitations. *Psychophysiology* **33**, 457–461, doi:10.1111/j.1469-8986.1996.tb01071.x (1996).

48 Coyne, J. T., Foroughi, C. & Sibley, C. Pupil diameter and performance in a supervisory control task: A measure of effort or individual differences? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **61**, 865–869, doi:10.1177/1541931213601689 (2017).

49 Jainta, S. & Baccino, T. Analyzing the pupil response due to increased cognitive demand: an independent component analysis study. *Int. J. Psychophysiol.* **77**, 1–7, doi:10.1016/j.ijpsycho.2010.03.008 (2010).

50 Goldinger, S. D. & Papesh, M. H. Pupil dilation reflects the creation and retrieval of memories. *Curr. Dir. Psychol. Sci.* **21**, 90–95, doi:10.1177/0963721412436811 (2012).

51 Otero, S. C., Weekes, B. S. & Hutton, S. B. Pupil size changes during recognition memory. *Psychophysiology* **48**, 1346–1353, doi:10.1111/j.1469-8986.2011.01217.x (2011).

52 Privitera, C. M., Renninger, L. W., Carney, T., Klein, S. & Aguilar, M. Pupil dilation during visual target detection. *Journal of Vision* **10**, 1–14, doi:10.1167/10.10.3 (2010).

53 Bradley, M. M., Miccoli, L., Escrig, M. A. & Lang, P. J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **45**, 602–607, doi:10.1111/j.1469-8986.2008.00654.x (2008).

54 Partala, T. & Surakka, V. Pupil size variation as an indication of affective processing. *Int. J. Hum. Comput. Stud.* **59**, 185–198, doi:10.1016/s1071-5819(03)00017-x (2003).

55 Maier, S. U. & Grueschow, M. Pupil dilation predicts individual self-regulation success across domains. *Sci. Rep.* **11**, 14342, doi:10.1038/s41598-021-93121-y (2021).

56 Nassar, M. R. *et al.* Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040–1046, doi:10.1038/nn.3130 (2012).

57 Urai, A. E., Braun, A. & Donner, T. H. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat Commun* **8**, 14637, doi:10.1038/ncomms14637 (2017).

58 Lanier, M. K. *et al.* Virtual reality check: Statistical power, reported results, and the validity of research on the psychology of virtual reality and immersive environments. *Comput. Human Behav.* **100**, 70–78, doi:10.1016/j.chb.2019.06.015 (2019).

59 Parsons, T. D. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Front. Hum. Neurosci.* **9**, 660, doi:10.3389/fnhum.2015.00660 (2015).

60 Slater, M. & Sanchez-Vives, M. V. Enhancing our lives with immersive virtual reality. *Front. Robot. AI* **3**, 1–47, doi:10.3389/frobt.2016.00074 (2016).

61 Bozkir, E., Geisler, D. & Kasneci, E. Person independent, privacy preserving, and real time assessment of cognitive load using eye tracking in a virtual reality setup. *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1834–1837, doi:10.1109/VR.2019.8797758 (2019).

62 Blascovich, J. *et al.* Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* **13**, 103–124, doi:10.1207/S15327965PLI1302_01 (2002).

63 Fox, J., Arena, D. & Bailenson, J. N. Virtual reality: A survival guide for the social scientist. *J. Media Psychol.* **21**, 95–113, doi:10.1027/1864-1105.21.3.95 (2009).

64 Adams, R., Finn, P., Moes, E., Flannery, K. & Rizzo, A. S. Distractibility in attention/deficit/hyperactivity disorder (ADHD): The virtual reality classroom. *Child Neuropsychol.* **15**, 120–135, doi:10.1080/09297040802169077 (2009).

65 Nolin, P. *et al.* ClinicaVR: Classroom-CPT: A virtual reality tool for assessing attention and inhibition in children and adolescents. *Comput. Human Behav.* **59**, 327–333, doi:10.1016/j.chb.2016.02.023 (2016).

66 Rizzo, A. A. *et al.* A virtual reality scenario for all seasons: The virtual classroom. *CNS Spectr.* **11**, 35–44, doi:10.1017/S1092852900024196 (2006).

67 Rizzo, A. A. *et al.* The virtual classroom: A virtual reality environment for the assessment and rehabilitation of attention deficits. *Cyberpsychol. Behav.* **3**, 483–499, doi:10.1089/10949310050078940 (2000).

68 Blume, F. *et al.* Do students learn better when seated close to the teacher? A virtual classroom study considering individual levels of inattention and hyperactivity-impulsivity. *Learn. Instr.* **61**, 138–147, doi:10.1016/j.learninstruc.2018.10.004 (2019).

69 Bailenson, J. N. *et al.* The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *J. Learn. Sci.* **17**, 102–141, doi:10.1080/10508400701793141 (2008).

70 Bailenson, J., Blascovich, J., Beall, A. C. & Loomis, J. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence* **10**, 583–598, doi:10.1162/105474601753272844 (2001).

71 Hoyt, C. L., Blascovich, J. & Swinth, K. R. Social inhibition in immersive virtual environments. *Presence* **12**, 183–195, doi:10.1162/105474603321640932 (2003).

72 Bailenson, J. & Yee, N. Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychol. Sci.* **16**, 814–819, doi:10.1111/j.1467-9280.2005.01619.x (2005).

73 Slater, M. *et al.* A virtual reprise of the Stanley Milgram obedience experiments. *PLoS One* **1**, e39, doi:10.1371/journal.pone.0000039 (2006).

74 Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y. & Muhlberger, A. The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Front. Psychol.* **6**, 26, doi:10.3389/fpsyg.2015.00026 (2015).

75 Huang, S. A. & Bailenson, J. in *Mind, Brain and Technology Educational Communications and Technology: Issues and Innovations* (eds Thomas D. Parsons, Lin Lin, & Deborah Cockerham) Ch. 4, 49–65 (Springer, 2019).

76 Bailey, J. O. & Bailenson, J. in *Cognitive Development in Digital Contexts* (eds Fran C. Blumberg & Patricia J. Brooks) 181–200 (Academic Press, 2017).

77 Slater, M. & Wilbur, S. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence* **6**, 603–616, doi:10.1162/pres.1997.6.6.603 (1997).

78 Schubert, T., Friedmann, F. & Regenbrecht, H. The experience of presence: Factor analytic insights. *Presence* **10**, 266–281, doi:10.1162/105474601300343603 (2001).

79 Lombard, M., Ditton, T. B. & Weinstein, L. Measuring presence: The temple presence inventory. *Proceedings of the 12th Annual International Workshop on Presence*, 1–15 (2009).

80 Biocca, F., Harms, C. & Burgoon, J. K. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence* **12**, 456–480 (2003).

81 Bulu, S. T. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Comput. Educ.* **58**, 154–161, doi:10.1016/j.compedu.2011.08.024 (2012).

82 Tobii Tech. *What is eye tracking?*, https://tech.tobii.com/technology/what-is-eye-tracking/ (2020).

83 Bailenson, J. *et al.* Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. *Proceedings of the 7th Annual International Workshop on Presence*, 1864–1105 (2004).

84 Bailenson, J., Beall, A. C. & Blascovich, J. Gaze and task performance in shared virtual environments. *J. Visual. Comput. Animat.* **13**, 313–320, doi:10.1002/vis.297 (2002).

85 Herr, P. M., Sherman, S. J. & Fazio, R. H. On the consequences of priming: Assimilation and contrast effects. *J. Exp. Soc. Psychol.* **19**, 323–340 (1983).

86 Mussweiler, T. Comparison processes in social judgment: Mechanisms and consequences. *Psychol. Rev.* **110**, 472–489, doi:10.1037/0033-295X.110.3.472 (2003).

87 van Steenbergen, H., Band, G. P. & Hommel, B. Threat but not arousal narrows attention: evidence from pupil dilation and saccade control. *Front. Psychol.* **2**, 281, doi:10.3389/fpsyg.2011.00281 (2011).

88 Babiker, A., Faye, I., Prehn, K. & Malik, A. Machine learning to differentiate between positive and negative emotions using pupil diameter. *Front. Psychol.* **6**, 1921, doi:10.3389/fpsyg.2015.01921 (2015).

89 Gibbons, F. X. & Buunk, A. P. Individual differences in social comparison: Development of a scale of social comparison orientation. *J. Pers. Soc. Psychol.* **76**, 129–142, doi:10.1037/0022-3514.76.1.129 (1999).

90 Huguet, P. *et al.* Clarifying the role of social comparison in the big-fish-little-pond effect (BFLPE): An integrative study. *J. Pers. Soc. Psychol.* **97**, 156–170, doi:10.1037/a0015558 (2009).

91 Jonkmann, K., Becker, M., Marsh, H. W., Lüdtke, O. & Trautwein, U. Personality traits moderate the big-fish-little-pond effect of academic self-concept. *Learn. Individ. Differ.* **22**, 736–746, doi:10.1016/j.lindif.2012.07.020 (2012).

92 Marsh, H. W. *et al.* Moderation of the Big-Fish-Little-Pond effect: Juxtaposition of evolutionary (Darwinian-economic) and achievement motivation theory predictions based on a Delphi approach. *Educ. Psychol. Rev.*, doi:10.1007/s10648-020-09583-5 (2021).

93 Plieninger, H. & Dickhäuser, O. The female fish is more responsive: Gender moderates the BFLPE in the domain of science. *Educ. Psychol. (Lond)* **35**, 213–227, doi:10.1080/01443410.2013.814197 (2013).

94 Seaton, M., Marsh, H. W. & Craven, R. G. Big-fish-little-pond effect: Generalizability and moderation - two sides of the same coin. *Am. Educ. Res. J.* **47**, 390–433, doi:10.3102/0002831209350493 (2010).

95 Jarodzka, H., Skuballa, I. & Gruber, H. Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educ. Psychol. Rev.* **33**, 1–10, doi:10.1007/s10648-020-09565-7 (2021).

96 Wolff, F., Helm, F., Zimmermann, F., Nagy, G. & Möller, J. On the effects of social, temporal, and dimensional comparisons on academic self-concept. *J. Educ. Psychol.* **110**, 1005–1025, doi:10.1037/edu0000248 (2018).

97 Möller, J. & Köller, O. Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *J. Educ. Psychol.* **93**, 826–835, doi:10.1037/0022-0663.93.4.826 (2001).

98 Weintrop, D. *et al.* Defining computational thinking for mathematics and science classrooms. *J. Sci. Educ. Technol.* **25**, 127–147, doi:10.1007/s10956-015-9581-5 (2016).

99 Grover, S. & Pea, R. Computational thinking in K-12: A review of the state of the Field. *Educ. Res.* **42**, 38–43, doi:10.3102/0013189X12463051 (2013).

100 Román-González, M., Pérez-González, J. C. & Jiménez-Fernández, C. Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Comput. Human Behav.* **72**, 678–691, doi:10.1016/j.chb.2016.08.047 (2017).

101 Savitzky, A. & Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1639., doi:10.1021/ac60214a047 (1964).

102 Mathôt, S., Fabius, J., Van Heusden, E. & Van der Stigchel, S. Safe and sensible preprocessing and baseline correction of pupil-size data. *Behav. Res. Methods* **50**, 94–106, doi:10.3758/s13428-017-1007-2 (2018).

103 Attard-Johnson, J., C, O. C. & Bindemann, M. Comparing methods for the analysis of pupillary response. *Behav. Res. Methods* **51**, 83–95, doi:10.3758/s13428-018-1108-6 (2019).

104 Roth, S. D. Ray casting for modeling solids. *Comput. Gr. Image Process.* **18**, 109–144, doi:10.1016/0146-664X(82)90169-1 (1982).

105 Agtzidis, I., Startsev, M. & Dorr, M. 360-degree video gaze behaviour. *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*, 1007–1015, doi:10.1145/3343031.3350947 (2019).

106 Soret, R., Charras, P., Khazar, I., Hurter, C. & Peysakhovich, V. Eye-tracking and virtual reality in 360-degrees: Exploring two ways to assess attentional orienting in rear space. *Symposium on Eye Tracking Research and Applications*, 1–7, doi:10.1145/3379157.3391418 (2020).

107 Epic Games. *Blueprints visual scripting*, https://docs.unrealengine.com/en-US/Engine/Blueprints/index.html (2020).

108 Schwanzer, A. D., Trautwein, U., Lüdtke, O. & Sydow, H. Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica* **51**, 183–194, doi:10.1026/0012-1924.51.4.183 (2005).

109 Mayer, A., Thoemmes, F., Rose, N., Steyer, R. & West, S. G. Theory and analysis of total, direct, and indirect causal effects. *Multivar. Behav. Res.* **49**, 425–442, doi:10.1080/00273171.2014.931797 (2014).

110 Muthén, L. K. & Muthén, B. O. *MPlus user's guide: 8th edition*. (Muthén & Muthén, 1998-2017).

111 Cohen, J. *Statistical power analysis for the behavioral sciences: 2nd edition*. (Lawrence Erlbaum, 1988).

Supplementary Information for:

Do Students Actively Seek Comparisons with Others? Using Eye Movement Data from a Virtual Reality Classroom to Uncover Social Information Processing

Supplementary Appendix 1. Effects on the number of peer learners looked at

The table below provides detailed statistics and standardized regression coefficients for the effects of the experimental hand-raising conditions on the *number of peer learners looked at*.

	b	SE	t	P-value	95% CI
(Intercept)	-0.02	0.07	-0.31	0.758	[-0.17, 0.12]
20% vs. 35% hand-raising	0.00	0.07	0.01	0.994	[-0.13, 0.13]
20% vs. 65% hand-raising	-0.08	0.07	-1.04	0.300	[-0.22, 0.07]
20% vs. 80% hand-raising	0.09	0.04	2.06	0.040	[0.00, 0.17]
Front vs. back sitting position	0.29	0.06	5.21	<0.001	[0.18, 0.40]
Cartoon vs. realistic avatars	-0.24	0.07	-3.51	<0.001	[-0.38, -0.11]
Gender (female vs. male)	-0.11	0.07	-1.50	0.133	[-0.25, 0.03]
Math grade	-0.04	0.06	-0.68	0.494	[-0.16, 0.08]
German grade	0.16	0.07	2.45	0.014	[0.03, 0.29]
Prior CT interest	-0.09	0.06	-1.51	0.132	[-0.21, 0.03]
Self-concept intelligence	0.09	0.04	1.96	0.050	[0.00, 0.17]
Social orientation	0.05	0.07	0.69	0.493	[-0.09, 0.19]

SE, standard error; CI, Confidence interval. P values refer to the standardized regression coefficients b. For categorical variables, the first category mentioned serves as the reference category. Grades were on a scale from 1–6 with higher numbers indicating lower achievement. Prior CT (Computational Thinking) interest, intelligence self-concept and social orientation were assessed on a 4-point rating scale with higher values indicating higher levels of these constructs.

With regards to the covariates, the results indicated that students' German grade and intelligence self-concept had a significant effect on the number of students looked at: Students with worse German grades and higher intelligence self-concept looked at more of their virtual peer learners, indicating that students' level of (domain-specific) confidence in their own abilities might have affected how often they turned their attention towards their peers. Moreover, with regard to the IVR configuration, the results revealed that the number of peer learners looked at *was* statistically significantly higher for participants who were seated in the back of the IVR classroom (d = 0.75), suggesting that the number of peer learners looked at increases when there are more peer learners in the field of view. Moreover, the number of peer learners looked at was significantly higher when the virtual classmates were presented as cartoon-style avatars (d = 0.57), indicating that students generally were more engaged with virtual peer learners when they were presented in a cartoon-style manner.

Supplementary Appendix 2. Effects on the frequency of gazing at peer learners

The table below provides detailed statistics and standardized regression coefficients for the effects of the experimental hand-raising conditions on *frequency of gazing at peer learners.*

	b	SE	t	P-value	95% CI
(Intercept)	-0.01	0.06	-0.24	0.808	[-0.17, 0.10]
20% vs. 35% hand-raising	-0.03	0.03	-0.78	0.434	[-0.09, 0.04]
20% vs. 65% hand-raising	-0.09	0.08	-1.08	0.279	[-0.24, 0.07]
20% vs. 80% hand-raising	0.05	0.03	1.91	0.056	[0.00, 0.11]
Front vs. back sitting position	0.47	0.05	10.00	<0.001	[0.37, 0.56]
Cartoon vs. realistic avatars	-0.29	0.03	-11.78	<0.001	[-0.34, -0.24]
Gender (female vs. male)	-0.02	0.06	-0.34	0.737	[-0.14, 0.10]
Math grade	-0.01	0.06	-0.09	0.932	[-0.12, 0.11]
German grade	0.10	0.08	1.30	0.193	[-0.05, 0.25]
Prior CT interest	-0.10	0.07	-1.43	0.153	[-0.24, 0.04]
Self-concept intelligence	0.13	0.04	3.50	<0.001	[0.06, 0.20]
Social orientation	0.09	0.05	1.94	0.052	[0.00, 0.18]

SE, standard error; CI, Confidence interval. P values refer to the standardized regression coefficients b. For categorical variables, the first category mentioned serves as the reference category. Grades were on a scale from 1–6 with higher numbers indicating lower achievement. Prior CT (Computational Thinking) interest, intelligence self-concept and social orientation were assessed on a 4-point rating scale with higher values indicating higher levels of these constructs.

With regards to the covariates, we found a positive effect of students' general intelligence self-concept on the total gaze time on peer learners, suggesting that whereas peer learners' achievement-related behaviour (i.e., the hand-raising conditions) did not affect how long students looked at their peers, students who felt more comfortable about their intellectual abilities paid more attention to their virtual classmates. Similarly to the number of peer learners looked at (see Supplementary Appendix 1), the results regarding the IVR configuration revealed that the frequency of gazing at peer learners was statistically significantly higher for participants who were seated in the back of the IVR classroom (d = 1.24), suggesting that visual attention to peer learners increases when more peer learners are in the field of view. The frequency of gazing at peer learners was also statistically significantly higher when the virtual classmates were presented as cartoon-style avatars (d = 0.80), indicating that students spent (or needed to spend) more time processing the social information provided by the peer learners' avatars when they were less realistic.

Supplementary Appendix 3. Effects on total gaze time on peer learners

The table below provides detailed statistics and standardized regression coefficients for the effects of the experimental hand-raising conditions on *total gaze time on peer learners.*

	b	SE	t	P-value	95% CI
(Intercept)	-0.01	0.05	-0.26	0.797	[-0.12, 0.09]
20% vs. 35% hand-raising	-0.02	0.04	-0.40	0.690	[-0.09, 0.06]
20% vs. 65% hand-raising	-0.09	0.08	-1.02	0.308	[-0.25, 0.08]
20% vs. 80% hand-raising	0.05	0.03	1.70	0.089	[-0.01, 0.10]
Front vs. back sitting position	0.48	0.05	10.56	<0.001	[0.39, 0.57]
Cartoon vs. realistic avatars	-0.28	0.03	-10.50	<0.001	[-0.33, -0.23]
Gender (female vs. male)	-0.04	0.05	-0.72	0.469	[-0.14, 0.07]
Math grade	-0.01	0.06	-0.10	0.923	[-0.13, 0.12]
German grade	-0.10	0.08	-1.19	0.234	[-0.07, 0.27]
Prior CT interest	-0.09	0.06	-1.47	0.142	[-0.21, 0.03]
Self-concept intelligence	0.13	0.03	3.85	<0.001	[0.07, 0.20]
Social orientation	0.08	0.04	1.72	0.086	[-0.01, 0.16]

SE, standard error; CI, Confidence interval. P values refer to the standardized regression coefficients b. For categorical variables, the first category mentioned serves as the reference category. Grades were on a scale from 1–6 with higher numbers indicating lower achievement. Prior CT (Computational Thinking) interest, intelligence self-concept and social orientation were assessed on a 4-point rating scale with higher values indicating higher levels of these constructs.

Similarly to the model predicting the frequency of gazing at peers (see Supplementary Appendix 2), we found a positive effect of students' general intelligence self-concept on the total gaze time on peer learners, suggesting that whereas peer learners' achievement-related behaviour (i.e., the hand-raising conditions) did not affect how long students looked at their peers, those students who felt more comfortable about their intellectual abilities paid more attention to their virtual classmates. With regards to the IVR configuration—similarly to the number of peers looked at and the frequency of gazing at peers (see Supplementary Appendices 1 and 2)—the results revealed that the total gaze time on peer learners was statistically significantly higher for participants who were seated in the back of the IVR classroom (d = 1.30) and when the virtual classmates were presented as cartoon-style avatars (d = 0.77). This indicates that with more peer learners in the field of view also the visual attention to them increased and students spent (or needed) more time processing the social information provided by the peer learners' avatars when they were less realistic.

Supplementary Appendix 4. Effects on the mean pupil diameter

The table below provides detailed statistics and standardized regression coefficients for the effects of the experimental hand-raising conditions on the *mean pupil diameter.*

	b	SE	t	P-value	95% CI
(Intercept)	0.00	0.04	-0.09	0.930	[-0.08, 0.07]
20% vs. 35% hand-raising	-0.04	0.05	-0.78	0.433	[-0.14, 0.06]
20% vs. 65% hand-raising	0.05	0.06	0.98	0.328	[-0.06, 0.16]
20% vs. 80% hand-raising	0.20	0.08	2.45	0.014	[0.04, 0.35]
Front vs. back sitting position	-0.04	0.06	-0.69	0.493	[-0.15, 0.07]
Cartoon vs. realistic avatars	0.09	0.06	1.45	0.146	[-0.03, 0.22]
Gender (female vs. male)	0.00	0.04	-0.07	0.946	[-0.07, 0.07]
Math grade	-0.07	0.06	-1.24	0.215	[-0.18, 0.04]
German grade	0.07	0.07	0.89	0.376	[-0.08, 0.21]
Prior CT interest	0.05	0.04	1.35	0.177	[-0.02, 0.12]
Self-concept intelligence	-0.06	0.06	-0.95	0.344	[-0.18, 0.06]
Social orientation	-0.14	0.07	-1.97	0.048	[-0.29, 0.00]

SE, standard error; CI, Confidence interval. P values refer to the standardized regression coefficients b. For categorical variables, the first category mentioned serves as the reference category. Grades were on a scale from 1–6 with higher numbers indicating lower achievement. Prior CT (Computational Thinking) interest, intelligence self-concept and social orientation were assessed on a 4-point rating scale with higher values indicating higher levels of these constructs.

Only one of the covariates had an effect on students' pupil diameter. The results showed a small negative effect of social orientation, indicating that students with higher levels of social orientation had a lower mean pupil diameter on average during the IVR lesson. A possible explanation for this finding would be that students with a high social orientation are used to seeking and processing higher levels of social information and thus were less affected by the social information provided in the IVR classroom situation.

Supplementary Appendix 5. Effects on students' situational self-concept

As described in the Results section, students' self-concept was significantly related to (a) the experimental manipulation of peer learners' hand-raising behaviour (particularly comparing the 'extreme' conditions of 20% and 80% hand-raising), and (b) interindividual differences in visual attention towards social comparison information (indicated by the number of peers looked at, the frequency of gazing at peers, and the total time spent gazing at peer learners).

Beyond that, the results showed that students' grades in mathematics and German statistically significantly predicted their situational self-concept in all three regression models using the different eye movements as predictors. More specifically, the results revealed that better grades in mathematics led to a higher situational self-concept regarding the IVR lesson on computational thinking, whereas better grades in German were associated with a lower situational self-concept. These results indicate that whereas students' active social comparisons in the IVR situation—as indicated by the time they spent looking at their peer learners and their mean pupil diameter—had an effect on how they evaluated themselves, students seemed to also base their self-evaluations on their achievement in other (dis)similar subjects.

The table below provides detailed statistics and standardized regression coefficients for the regression models predicting students' situational self-concept using the number of peers looked at (Model 1), the frequency of gazing at peers (Model 2), the total gaze time on peers (Model 3), and the mean pupil diameter (Model 4).

	b	SE	t	P-value	95% CI
Model 1: Number of peers looked at					
(Intercept)	0.00	0.09	-0.02	0.984	[-0.18, 0.18]
Number of peers looked at	-0.13	0.05	-2.56	0.010	[-0.23, -0.03]
20% vs. 35% hand-raising	-0.01	0.06	-0.12	0.902	[-0.12, 0.11]
20% vs. 65% hand-raising	-0.07	0.06	-1.11	0.268	[-0.18, 0.05]
20% vs. 80% hand-raising	-0.10	0.04	-2.42	0.015	[-0.19, -0.02]
Front vs. back sitting position	-0.07	0.06	-1.17	0.242	[-0.18, 0.05]
Cartoon vs. realistic avatars	-0.01	0.07	0.08	0.933	[-0.15, 0.14]
Gender (female vs. male)	0.06	0.10	0.60	0.547	[-0.13, 0.25]
Math grade	-0.12	0.04	-2.70	0.007	[-0.20, -0.03]
German grade	0.20	0.04	5.53	<0.001	[0.13, 0.27]
Prior CT interest	0.04	0.08	0.55	0.580	[-0.11, 0.20]
Self-concept intelligence	0.13	0.07	1.77	0.077	[-0.01, 0.27]
Social orientation	0.01	0.04	0.33	0.744	[-0.07, 0.10]

(continued)

	b	SE	t	P-value	95% CI
Model 2: Frequency of gazing at peers					
(Intercept)	0.00	0.09	0.01	0.992	[-0.18, 0.18]
Frequency of gazing at peers	-0.11	0.05	-2.06	0.040	[-0.21, -0.01]
20% vs. 35% hand-raising	-0.01	0.06	-0.20	0.843	[-0.12, 0.10]
20% vs. 65% hand-raising	-0.07	0.06	-1.13	0.258	[-0.18, 0.05]
20% vs. 80% hand-raising	-0.11	0.05	-2.46	0.014	[-0.20, -0.02]
Front vs. back sitting position	-0.06	0.06	-0.91	0.364	[-0.18, 0.07]
Cartoon vs. realistic avatars	-0.01	0.08	-0.09	0.927	[-0.16, 0.14]
Gender (female vs. male)	0.07	0.10	0.69	0.493	[-0.13, 0.27]
Math grade	-0.11	0.04	-2.68	0.007	[-0.19, -0.03]
German grade	0.19	0.03	5.48	<0.001	[0.12, 0.25]
Prior CT interest	0.05	0.08	0.57	0.571	[-0.11, 0.20]
Self-concept intelligence	0.13	0.07	1.76	0.078	[-0.01, 0.27]
Social orientation	0.02	0.05	0.39	0.696	[-0.07, 0.11]
Model 3: Total gaze time on peers					
(Intercept)	0.00	0.09	-0.01	0.992	[-0.18, 0.18]
Total gaze time on students	-0.10	0.05	-2.27	0.023	[-0.19, -0.01]
20% vs. 35% hand-raising	-0.01	0.06	-0.18	0.860	[-0.12, 0.10]
20% vs. 65% hand-raising	-0.06	0.06	-1.13	0.260	[-0.18, 0.05]
20% vs. 80% hand-raising	-0.11	0.04	-2.50	0.012	[-0.20, -0.02]
Front vs. back sitting position	-0.06	0.06	-0.95	0.341	[-0.17, 0.06]
Cartoon vs. realistic avatars	0.00	0.07	-0.06	0.953	[-0.15, 0.14]
Gender (female vs. male)	0.07	0.10	0.67	0.504	[-0.13, 0.27]
Math grade	-0.11	0.04	-2.68	0.007	[-0.19, -0.03]
German grade	0.19	0.03	5.45	<0.001	[0.12, 0.25]
Prior CT interest	0.05	0.08	0.59	0.556	[-0.11, 0.20]
Self-concept intelligence	0.13	0.07	1.75	0.080	[-0.02, 0.27]
Social orientation	0.02	0.05	0.34	0.734	[-0.07, 0.11]
Model 4: Mean pupil diameter					
(Intercept)	0.00	0.09	0.00	0.997	[-0.18, 0.18]
Mean pupil diameter	-0.07	0.04	-1.65	0.100	[-0.16, 0.01]
20% vs. 35% hand-raising	-0.01	0.05	-0.22	0.829	[-0.11, 0.09]
20% vs. 65% hand-raising	-0.05	0.06	-0.91	0.361	[-0.16, 0.06]
20% vs. 80% hand-raising	-0.10	0.05	-2.17	0.030	[-0.19, -0.01]
Front vs. back sitting position	-0.11	0.06	-1.72	0.086	[-0.23, 0.02]
Cartoon vs. realistic avatars	-0.03	0.07	0.44	0.661	[-0.11, 0.17]
Gender (female vs. male)	0.07	0.10	0.70	0.484	[-0.13, 0.27]
Math grade	-0.12	0.04	-2.64	0.008	[-0.20, -0.03]
German grade	0.18	0.04	5.31	<0.001	[0.12, 0.25]
Prior CT interest	0.06	0.08	0.75	0.451	[-0.09, 0.21]
Self-concept intelligence	0.11	0.07	1.52	0.128	[-0.03, 0.26]
Social orientation	0.00	0.05	-0.01	0.995	[-0.10, 0.09]

SE, standard error; CI, Confidence interval. P values refer to the standardized regression coefficients b. For categorical variables, the first category mentioned serves as the reference category. Grades were on a scale from 1–6 with higher numbers indicating lower achievement. Prior CT (Computational Thinking) interest, intelligence self-concept and social orientation were assessed on a 4-point rating scale with higher values indicating higher levels of these constructs.

5

STUDY 3

Hasenbein, L., Stark, P., Trautwein, U., Queiroz, A. C. M., Bailenson, J., Hahn, J.-U., & Göllner, R. (2021). *Configuring an immersive virtual reality classroom for educational research and practice: Implications from students' gaze-based attention networks.* Manuscript submitted for publication.

The following manuscript has not yet been accepted or published. The version displayed here might not exactly replicate the final version published in the journal. It is not the copy of record.

Abstract

Immersive virtual reality (IVR) provides great potential for experimentally investigating the relevance of different classroom features for student learning and strategically deploying these features to design effective learning environments. The present study examined how three salient classroom features (i.e., students' position in the classroom, visualization style of virtual avatars, and virtual classmates' performance-related behavior) affect students' processing of information presented in the IVR classroom using a large-scale eye-tracking dataset of $N = 274$ sixth graders. Results based on students' gaze-based attention networks during instruction showed that the IVR configurations were systematically associated with differences in gaze centrality on classmates or the instructional content, the connectedness of gazes, and overall uniformity of gaze distribution. In order to gain a deeper understanding of these gaze-based features, their relations to students' IVR learning experience were examined. Gaze-based attention on classmates was negatively related to students' interest in the IVR lesson; specifically, the share of boys observed was negatively related to students' situational self-concept. In turn, gaze-based attention on the instructional content was positively related to students' performance after the IVR lesson. Implications for the design and use of IVR classrooms in educational research and practice are discussed.

1. Introduction

Searching Web of Science for peer-reviewed articles with “virtual reality” and “education” in the abstract yields about 3,600 results—two-thirds of which were published within the last five years (according to a database search as of October 2021). From immersive virtual reality (IVR) applications for engineering education (Alhalabi, 2016), the military (Webster, 2015) and medical training (Bric et al., 2016; Moro et al., 2017) to environmental education (Fauville et al., 2020), virtual field trips and science simulations in elementary and secondary school (Cheng & Tsai, 2019; Makransky et al., 2020; Makransky, Terkildsen, et al., 2019; Queiroz et al., 2018): IVR and its associated affordances are becoming more and more popular in training and education. Most educational IVR applications focus on experiential learning, particularly simulations of experiences that are difficult or impossible for learners to have in real life (Howard et al., 2021; Johnston et al., 2017). However, in addition to IVR simulations that take learners out of the classroom, the transformation of ‘typical’ classrooms into IVR learning environments is a promising methodology for educational research and practice (a) to examine the relevance of different classroom features for student learning and (b) strategically deploy these features to design effective learning environments (Bailenson et al., 2008; Karutz & Bailenson, 2015; Rizzo et al., 2006).

Notably, classroom situations in the real world are complex and dynamic, and students’ classroom learning is substantially shaped by numerous contextual and peer-related factors (Brophy & Good, 1974; Harker & Tymms, 2004; Marsh et al., 2012). The (perceived) learning environment—which is strongly characterized by peer learners—has been found to be related to students’ achievement and academic trajectories (Göllner et al., 2018; Gottfried, 2012; Hattie, 2002; Hochweber et al., 2014; Lavy et al., 2011) as well as their emotions and motivation during learning (Frenzel et al., 2007; Hardré & Sullivan, 2008; Pekrun et al., 2019). Hence, a central goal when designing IVR classrooms for educational research and practice should be to authentically simulate classroom scenarios in order to (a) use them as an experimental tool to gain insights into the processes underlying students’ learning in the classroom (i.e., in a standardized yet authentic setting; Blascovich et al., 2002; Fox et al., 2009) and subsequently to (b) strategically deploy certain configurations for more effective learning (e.g., in remote learning scenarios or using virtual peer learners as pedagogical agents; Bailenson et al., 2008; Hudson & Hurter, 2016; Makransky, Wismer, et al., 2019). Whereas the use of IVR classrooms in educational research and practice has been increasing (see examples by Adams et al., 2009; Bailenson et al., 2008; Blume et al., 2019; Nolin et al., 2016; Rizzo et al., 2000), systematic insights into how different configurations, specifically in the IVR classroom, affect users' perception of the IVR environment are scarce. Importantly, the majority of existing studies about individual IVR experiences are based on samples of (young) adults; hence, a clear understanding of how children perceive IVR environments and social interactions in the virtual space is lacking (Bailey & Bailenson, 2017). Assuming that contextual factors and peer learners substantially shape students' learning experience not just in real-world classrooms (Dijkstra et al., 2008; Trautwein et al., 2015) but also in IVR classroom settings, it is crucial to understand how different IVR classroom configurations affect how and to what extent students attend to the provided (social) information during an IVR lesson.

Of course, there are countless ways to configure an IVR classroom and therefore many features that could potentially influence students' processing of (social) information in the IVR environment. However, some configuration features are more salient than others, such as the perspective from which students experience the lesson or the visualization style and behavior of their social counterparts. Do students focus more on their virtual classmates versus the instructional content when they sit in the back of the IVR classroom? What role does a more or less stylized visualization of virtual classmates play? Finally, does virtual peer learners' performance-related behavior (e.g., more or less hand-raising) affect students' learning experience in the IVR classroom?

In the present study, we aim to provide answers to these questions by examining students' learning experience in an IVR classroom with different configurations. More specifically, we examined three salient features of IVR classroom configurations that are decisive for how students perceive and process what is happening during a virtual classroom lesson (see Section 1.1). To gain insights into students' learning experience in the IVR classroom, we used students' gaze data and analyzed their gaze-based attention networks as an indicator of (social) information processing in the different IVR configurations (see Section 1.2). In order to provide insights into the meaning of the gaze-based features used, we additionally examined how they are related to central learning outcomes, namely students' interest in the lesson, their situational self-concept and post-lesson achievement.

1.1. Configuration of immersive virtual reality classrooms for educational research and practice

Given the myriad of decisions involved in the configuration of IVR classrooms, findings from educational psychology research and already existing studies in IVR (classroom) contexts point to central features that seem to affect students' learning experience in the classroom and therefore need to be carefully considered when configuring IVR classrooms.

First, one of the most salient features for students' classroom learning is the seating arrangement and position of students within the classroom (MacAulay, 1990; Wannarka & Ruhl, 2008). Research on the effect of students' position in the classroom has provided mixed findings regarding different outcome variables, indicating a positive effect of a front sitting position close to the teacher on students' performance, but also null effects for performance outcomes or only effects on students' motivation, not performance (Fernandes et al., 2011; LaCroix & LaCroix, 2017; Levine et al., 1980; Meeks et al., 2013; Montello, 1988; Perkins & Wieman, 2005; Schwebel & Cherlin, 1972; Will et al., 2020). This is not surprising considering that due to natural limitations in the classroom, existing studies are situated in very different classroom environments and have not always randomly allocated students to sitting positions (i.e., results that indicate better learning outcomes for students sitting in the front might be confounded by the seating choices of higher-performing students and/or changes in teachers' instructional practices in response to certain classroom compositions). In order to address this issue, some IVR studies have systematically varied students' position in the classroom in order to provide experimental evidence. For instance, Bailenson et al. (2008) manipulated participants' position in the classroom, including their distance to the teacher, while keeping all other factors constant and found effects on students' subsequent learning outcomes (see Experiments 2 and 3). Similarly, Blume et al. (2019) found that students who were assigned to a position closer to the virtual teacher performed better in a posttest compared to students that were placed in the back of the IVR classroom. Most importantly for the scope of the present study, existing research has not moved beyond learning outcomes as a measure of distinct effects of sitting positions in the classroom on students' learning experience. Thus, how students' position in the classroom actually affects how they attend to and process (social) information in the classroom (e.g., the instructional content and social information provided by their classmates) remains an open question. Whereas sitting in the front is most likely associated with increased attention to the teacher and instructional content, paying some attention to peers might also be desirable, particularly when considering potentially beneficial effects of peers (e.g., as pedagogical agents or as a motivating reference group).

Second, one of the most salient features when configuring IVR classrooms is the visualization style of social counterparts such as virtual classmates and the virtual teacher (Cheng et al., 2002). As the Uncanny Valley effect (Mori et al., 2012) and related works (e.g., Ho & MacDorman, 2010; MacDorman, 2006) indicate, avatars' more human-like appearance is not the only decisive factor, and more importantly, not always desirable if the goal is for users to have a favorable perception of virtual avatars (Mathur & Reichling, 2016; Strait et al., 2015). Notably, IVR studies examining this effect mostly compare the two ends of the spectrum, i.e., full-body human-like avatars versus non-human-like visualizations such as avatars with only a head and hands (Heidicker et al., 2017), a drone that functions as a pedagogical agent (Makransky, Wismer, et al., 2019) or avatars with a more animal-like appearance (Zanbaka et al., 2006). If the aim is to configure an IVR classroom with a teacher and classmates that are clearly recognizable as such and able to simulate a real-world classroom scenario, how realistically these human-like avatars need to be visualized remains an open question. Does a cartoonish visualization of virtual classmates and the virtual teacher lead to the same perceptions as more stylized representations? Particularly given that animation and design costs increase with increasing realism, it seems worth examining what degree of realism is necessary when designing virtual avatars. Moreover, in addition to the question of what is perceived as authentic and realistic—which has been the focus of most avatar-related research to date, another open question concerns how different avatar visualization styles affect students' processing of the social information provided in the classroom (e.g., virtual classmates' behavior in contrast to the instructional content). For instance, previous studies have found longer fixation durations in an IVR classroom with cartoon-style avatars (Gao et al., 2021) and longer dwell time on peer learners visualized as cartoonish characters (Bozkir et al., 2021). When interpreting the results, the authors argued that the unusual appearance of cartoon-style peer learners and the increased difficulty of decoding social information from less realistic avatars might lead to these results.

Third, another central feature to consider when configuring IVR classroom scenarios is peer learners' behavior (e.g., performance level and active participation). Educational psychology research has repeatedly demonstrated that classmates substantially shape student learning, highlighting the role of what can be called "classroom composition" effects (see, e.g., Trautwein et al., 2015). On the one hand, there is evidence for so-called positive spillover effects of higher-achieving peers on students' achievement and self-evaluations, in the sense that learners benefit from high-achieving peers and perform better when they are surrounded by high-performing classmates (Brewer & Weber, 1994; De Fraine et al., 2003; Fruehwirth, 2013; Pelham & Wachsmuth, 1995). On the other hand, there is a large body of evidence for negative contrast effects in the face of high-achieving classmates, suggesting that higher-achieving peers have a negative impact on students' evaluations of their own competence, controlling for individual achievement (the so-called Big-Fish-Little-Pond Effect, see latest reviews by Fang et al., 2018; Marsh et al., 2017; Marsh & Seaton, 2015).

Notably, whereas the aforementioned classroom composition effects have been studied exhaustively in educational psychology research—typically by examining students' test performance and self-reports of their own competencies in relation to their peers' average test performance, the effect of peer learners' actual (performance-related) behavior has received little attention. In other words, the actual processes underlying compositional effects in a classroom situation (e.g., the effect of peer learners' behavior on students' learning and attention distribution) remain largely unexplored. IVR classrooms provide the opportunity to examine such effects in an authentic yet controlled setting, as demonstrated for instance by Bailenson et al. (2008). The authors manipulated virtual classmates' attention-related behavior (i.e., peer learners being attentive or distracting during instruction; see Experiment 4) and found positive effects of more attentive virtual classmates on students' performance after the IVR lesson. Similarly, a study in the field of economics manipulated virtual co-workers' productivity and found a positive relation with the performance of participants doing the same task as the virtual co-workers in an IVR environment (Bönsch et al., 2017; Gürerk et al., 2019). In sum, based on existing IVR studies, it can be assumed that classmates' performance affects students' learning in IVR settings; however, it is still unclear how exactly peers' (performance-related) behavior in an IVR classroom needs to be configured in order to be recognized by K-12 students in an IVR classroom scenario. Beyond the opportunity to use IVR classrooms as a tool to examine the effects of peers' performance-related behavior, such studies have important implications for the design and use of virtual peer learners as pedagogical agents in IVR classroom-based learning applications (see e.g., Bailenson et al., 2008; Hudson & Hurter, 2016; Makransky, Wismer, et al., 2019).

Taken together, the three outlined features of IVR classrooms (i.e., students' position, visualization style of virtual avatars, and virtual peer learners' behavior) play an important role in student learning. However, with previous studies based either on real-world classroom research or self-reported experiences in IVR (classroom) settings, it remains unclear how exactly these features affect how students process different types of information in an IVR classroom environment. Students' gaze data provides an opportunity to obtain such insights.

1.2. Students' gaze-based attention networks as indicators of (social) information processing

Students' gaze behavior allows for insights into how students process information presented to them in an IVR classroom environment (Holmqvist et al., 2011; Hutmacher, 2019; Jarodzka et al., 2017). Moreover, compared to real-life classrooms, gaze data from an IVR classroom provides the opportunity to combine the high methodological rigor of a standardized environment with an authentic representation of a classroom situation with all its accompanying dynamics. Thanks to recent technological advances, state-of-the-art IVR equipment comes with integrated eye trackers, making it possible to unobtrusively examine students' gaze behavior to gain an unbiased and in-depth understanding of their learning experience in the IVR.

Importantly, the interpretation of eye-tracking data is known to be context-specific, and the appropriate analysis technique to understand how attention is distributed must be chosen carefully (Kaakinen, 2021; Lai et al., 2013). Most learning-related studies analyzing eye movement data have used temporal and count measures such as number of fixations, number of saccades, and fixation durations (Lai et al., 2013), which are easy to collect with available software. However, these commonly used eye-tracking features are often analyzed in isolation, and it is difficult to establish an interpretable link between eye movements and underlying cognitive processes (e.g., Strohmaier et al., 2020). As Lai et al. (2013) point out, more sophisticated measures are necessary to investigate meta-cognitive skills in-depth. Based on the assumption that students guide their attention in the classroom and focus on certain objects (e.g., their classmates) while ignoring others (e.g., the teacher and instructional content on the screen), what students look at can serve as an indication of what they pay attention to. Such so-called overt spatial or visual attention (Bundesen, 1990; Carrasco, 2011; Kübler et al., 2017; Lodge & Harrison, 2019) can be analyzed using eye-tracking data. Considering that an IVR classroom is a relatively static environment where spatial relations between objects do not change substantially over time, it can be assumed that students are able to willfully direct their attention to certain objects at least to a substantial degree (Katsuki & Constantinidis, 2014; Theeuwes et al., 2000). Corresponding processes of active information gathering are reflected in longer gaze movement periods, such as consecutive gaze shifts from object to object (instead of eye movement features like fixations and saccades operating on a level of milliseconds; e.g., Kaakinen, 2021).

Taking these aspects into consideration, in the present study, we opt for a rather novel approach and apply the methodology of network analysis to the analysis of gaze data to gain insights into students' gaze-based attention distribution in an IVR classroom. Network analysis (based on the mathematical theory of graphs; Diestel, 2017) is a prominent method in various scientific fields, including biology, geography and the social sciences (e.g., Charitou et al., 2016; Chiesi, 2001; Curtin, 2018). However, this approach has so far received little attention in eye-tracking research and there are only few studies performing network analysis with gaze data (Guillon et al., 2015; Sadria et al., 2019; Schneider et al., 2013; Yazdan-Shahmorad et al., 2020). We argue that particularly when it comes to students' processing of (social) information in a classroom situation, the analysis of gaze-based attention networks provides novel and most importantly explainable and interpretable insights into students' gaze behavior during the IVR experience. It makes it possible to identify the degree to which certain objects of interest (e.g., peer learners, the teacher or the instructional content) are in the center or focus of gaze transitions (via so-called gaze centrality markers, see, e.g., Yazdan-Shahmorad et al., 2020). Moreover, the analysis of gaze-based attention networks allows for extracting information about the overall gaze activity and connectedness of gazes between certain objects of interest (e.g., how intensely different peer learners are attended to) or the overall distribution of gaze between objects of interest in the environment (e.g., how often students' gaze goes back and forth between the teacher and peer learners). A detailed description of the method and corresponding visual attention measures can be found in Methods Section 3.6.1 and in Appendix A.

2. The Present Study: Aims and Research Questions

The present study aims to gain insights into how different configurations of an IVR classroom with a full class of more than 20 virtual peer learners impact students' attention distribution towards (social) information in the IVR environment and their learning experience. To extend existing research, we focused particularly on the IVR experiences of children, who have been the subject of considerably less IVR experience-related research to date (Bailey & Bailenson, 2017). To this end, the present study examined how different IVR classroom configuration features affect students' gaze-based attention networks—as an indicator of (social) information processing—during instruction in an IVR classroom.

We focused on three configuration features that we consider of particular interest when aiming to answer the question of how to design an IVR classroom for ideal learning outcomes as well as research purposes, namely (a) participants' positioning in the IVR classroom, (b) the visualization style of virtual avatars of peer learners and the teacher, and (c) virtual peer learners' performance-related behavior. We (a) placed participating students either in a front or the back row of the IVR classroom and (b) visualized virtual avatars either in a cartoonish or more stylized (i.e., more realistic) manner. Moreover, we (c) used peer learners' hand-raising behavior as an indicator of students' behavioral engagement and performance and varied the proportion of virtual classmates who raised their hands to respond to the virtual teachers' question during the IVR lesson (i.e., 20%, 35%, 65% or 80%). Drawing on graph theory, we mapped students' visual attention patterns during the IVR lesson in terms of the gaze allocation to different objects of interest (OOIs; i.e., virtual peer learners, the virtual teacher, and the screen with instructional content) in the form of a graph. We then extracted different features that allowed us to describe students' gaze-based attention networks with regard to the focus of gaze transitions on OOIs, the connectedness of gazes between OOIs and the uniformity of gaze distribution across OOIs in the IVR classroom (see details in Methods Section 3.6.1). We used these features to examine differences in students' gaze-based attention networks with regard to the different IVR configuration conditions, asking:

RQ 1. How do different IVR configuration**s** affect students' gaze-based attention networks in the IVR classroom? More specifically, how do participants' position in the IVR classroom (front vs. back), the visualization style of virtual avatars (cartoonish vs. stylized) and the performance-related behavior of virtual peers (proportion of classmates who raise their hands) affect (a) the degree to which an OOI is in the center/focus of gaze transition, (b) the connectedness of gazes to peers, and (c) the uniformity of gaze distribution across OOIs and in the IVR classroom in general? We used different structural features to assess (a) to (c) respectively and tested the following hypotheses:

H1a. Given that students positioned in the back row of the classroom had the whole class of virtual peer learners in front of them, while students who were positioned in the front had only one row of students between themselves and the teacher and screen, we expected being positioned in the back of the virtual classroom leads to more gaze centrality on virtual peer learners (and less centered gaze networks on the virtual teacher and screen), more connectedness of gazes among peers, and due to the increased field of view, a more uniformly distributed gaze in the IVR classroom.

H1b. Based on existing findings regarding the less usual appearance of cartoonish peer learners and the increased difficulty of decoding social information from them due to their less fine-grained visualization (Bozkir et al., 2021; Gao et al., 2021), we hypothesized that a cartoonish visualization of avatars leads to more gaze centrality on virtual peer learners and less gaze centrality on the virtual teacher (gaze centrality on the screen not affected), more connectedness of gazes among peers, and due to the increased focus on peer learners, a less uniformly distributed gaze overall in the IVR classroom.

H1c. Assuming that increased activity of virtual peer learners attracts more attention from students, we expected that more hand-raising behavior of virtual classmates leads to more gaze centrality on peer learners (and gaze networks less centered on the teacher and screen), more connected gazes among peers, and based on the desire to obtain a comprehensive picture of peer learners' behavior, a more uniformly distributed gaze in the IVR classroom.

H1d. We expected the effects of avatar visualization style (H1b) and the variation in peers' performance-related behavior (H1c) on students' gaze-based attention networks to be particularly pronounced when seated in the back, where more peer learners were in the field of view. We therefore explored interaction effects between the configuration conditions.

Given that the methodological approach of network analysis has never before been used to analyze eye-tracking data from a classroom situation and with regard to learning-related outcomes, we sought to obtain a more substantiated understanding of how these indicators of students' gaze behavior were related to their learning experience in the IVR classroom. Therefore, in a second step, we asked:

RQ 2. How do structural features of students' gaze-based attention networks (i.e., the degree to which an OOI is in the center/focus of gaze transition, the connectedness of the gaze networks among peers, and the uniformity of gaze distribution across OOIs and in the IVR classroom in general) relate to their learning experience in the IVR classroom? Students' learning experience in the IVR classroom was examined in terms of (a) their interest in the IVR lesson, (b) their evaluation of their own competence in the IVR lesson (i.e., situational self-concept), and (c) their performance on a posttest assessing the IVR lesson content. We examined the following exploratory hypotheses:

H2a. We expected that more gaze centrality on peers and a higher connectedness of gazes among peers (i.e., more processing of social information) are related to lower interest in the IVR lesson, lower situational self-concept and lower test performance after the IVR lesson.

H2b. Based on the assumption that increased focus of visual attention on the instructional content is beneficial for students' learning outcomes, we hypothesized that more gaze centrality on the teacher and screen (i.e., more focus on the instructional content and less processing of social information) are related to higher interest in the IVR lesson, higher situational self-concept and better test performance after the IVR lesson.

Moreover, we explored how uniformity of gaze distribution across OOIs in the IVR classroom (as an indicator of rather balanced processing of social information) is related to students' interest in the IVR lesson, situational self-concept and test performance after the IVR lesson.

3. Method

The present study was approved by the regional educational authorities and the ethics committee of the University of Tübingen who confirmed that the procedures were in line with ethical standards for research on human subjects (date of approval: 11/25/2019, file number: A2.5.4-106_aa).

3.1. Participants

We collected data from a total of N = 381 sixth-grade students. In this study, we used data from N = 274 students with a sufficiently high eye-tracking ratio (> 90%). The lack of suitable eye-tracking data from the excluded students was mostly caused by hardware-related problems during data collection (e.g., incorrect eye-tracker calibration, unexpected crashing and restart of the IVR experience) and synchronization issues during data pre-processing. Importantly, the availability of suitable eye-tracking data was unsystematic with regard to the different testing groups and central sample characteristics (see respective statistics in the supplemental material). Similarly to the full sample, the students in our study stem from a total of 25 sixth-grade classes at 14 academic-track schools (M_{Age} = 11.50, SD_{Age} = 0.55, 50.4% girls). None of the children in our sample had participated in any previous IVR studies, but 57.8% indicated that they had experienced an IVR environment as a consumer at least once before.

3.2. Research design

This study follows a 2 × 2 × 4 between-subjects design in which we examined three different IVR configuration features, namely (a) participants' positioning in the IVR classroom, (b) the visualization style of virtual avatars, and (c) the performance-related behavior of virtual peer learners. Participants' position in the IVR classroom and virtual avatar visualization were varied on two levels (front vs. back and cartoon vs. stylized, respectively). Virtual peer learners' performance-related behavior was manipulated on four levels via varying proportions of students raising their hands in response to questions from the virtual teacher. A more detailed description of the IVR configurations is provided in Materials Section 3.3.1. Participating students were randomly assigned to one of the 16 (2 × 2 × 4) IVR configuration conditions via random number generation at the individual level. Table 1 shows the descriptive sample statistics after randomization.

Assuming small- to medium-sized effects ($f = .20$), we computed an a priori power analysis for respective analyses of variance with two-tailed tests at a .05 alpha level and a minimum power of .90. Based on this, a necessary sample size of $N = 22$ students per group was determined.

Table 1

Descriptive Sample Statistics after Randomization to One of the IVR Configuration Conditions

Variable	Total	Front (*N* = 122)		Back (*N* = 152)	
		Cartoonish (*N* = 56)	Stylized (*N* = 66)	Cartoonish (*N* = 94)	Stylized (*N* = 58)
Age	11.50 (0.55)	11.57 (0.57)	11.49 (0.53)	11.47 (0.58)	11.52 (0.50)
Gender					
Female	138	23	36	43	36
Male	136	33	30	51	22
Grades [a]					
Math	2.61 (0.89)	2.72 (0.87)	2.57 (0.77)	2.51 (0.96)	2.74 (0.91)
German	2.48 (0.72)	2.50 (0.68)	2.51 (0.83)	2.40 (0.72)	2.57 (0.62)
Prior IVR experience [b]					
No	114	21	25	38	30
Yes	156	35	41	54	26
n/a	4	-	-	2	2
General self-concept intelligence [c]	3.07 (0.61)	3.20 (0.61)	2.96 (0.56)	3.12 (0.64)	3.00 (0.60)
Initial CT interest [c]	3.14 (0.76)	3.27 (0.63)	3.18 (0.75)	3.10 (0.82)	3.04 (0.77)

Note. Mean values and standard deviations M (SD) are shown for continuous variables, categorical variables are shown in absolute numbers. Values are averaged across hand-raising conditions. CT = Computational Thinking; IVR = Immersive Virtual Reality.

[a] Grades were on a scale from 1–6 with lower numbers indicating better achievement; [b] Prior IVR experience was assessed via one item asking whether participants had previously used IVR glasses; [c] Measured on a 4-point rating scale with higher values indicating higher levels of the respective variable.

Figure 1

IVR Configuration Conditions

Note. The images stem from the exact same moment in the IVR lesson in different configuration conditions: The top image shows the avatar visualization in cartoon style; the bottom image depicts the more stylized (i.e., more realistic) avatar visualization. In the top image, 20% of peer learners raise their hand, compared to 80% in the bottom image. The numbers in black circles indicate the sitting positions (1) in the front (i.e., the second row) and (2) in the back (i.e., the last of four rows).

3.3. Materials

3.3.1. IVR Classroom Configuration Conditions

We implemented different IVR configurations with regard to the three features we consider of particular importance when designing an IVR classroom for ideal learning and research outcomes. We varied the positioning of participating students in the IVR classroom, placing them either in a front or a back row. This made participants experience the IVR lesson either (a) from a position close to the instructional center, with only one row of students between themselves and the teacher and screen on which the lesson content was presented, or (b) from a position in the back row of the classroom with the whole class of peer learners between themselves and the teacher and screen (see Figure 1).

Moreover, we varied the visualization style of the virtual avatars (i.e., teacher and peer learners). Participants were either surrounded by cartoonish or stylized (i.e., more realistically visualized) virtual avatars (see Figures 1 and 2 for an impression). The visualization style only concerned the look of the virtual avatars (i.e., tinier arms and legs, larger heads and eyes, and less fine-grained facial expressions for the cartoonish avatars), whereas audio and motions were the same in all conditions. Lastly, we varied the performance-related behavior of virtual peer learners via their hand-raising behavior; whenever the virtual teacher asked a question during the IVR lesson, either 20%, 35%, 65% or 80% of the virtual peer learners raised their hands to indicate that they knew the correct answer (see Figure 1 for an image of the two extreme conditions).

All configuration conditions aimed at a high degree of behavioral realism (Bailenson et al., 2004; Guadagno et al., 2007) and considered the Uncanny Valley effect in terms of appropriate avatar visualization (MacDorman, 2006; Mori et al., 2012). To ensure that the IVR lesson was consistent with a typical classroom experience for sixth graders and perceived as authentic, we used audio recordings and motion captures from a real sixth-grade classroom. We recorded and motion-captured six different students for the whole 15-minute IVR lesson duration and used individual sequences of recorded movements to distinctly animate the virtual peer learners in the classroom. Notably, participants reported similarly high levels of perceived realism and experienced presence in the IVR classroom across all configuration conditions.[20]

3.3.2. IVR Lesson Content

The IVR experience was a 15-minute simulation of a lesson on computational thinking. We chose computational thinking as the IVR lesson content because—despite being considered a central 21st century skill—this topic is not yet widely taught in primary and early secondary education, except for some extracurricular activities (Grover & Pea, 2013; Weintrop et al., 2016). Aiming to provide participants with novel content they had little (or no) prior knowledge or learning experiences with, the 15-minute IVR lesson introduced sequences and loops as basic computational concepts. The IVR lesson proceeded in the fashion of a typical classroom

[20] We assessed participants' perceived realism and experienced presence in the IVR classroom with six and nine items each in the posttest questionnaire. The measures were based on conceptualizations of presence by Schubert et al. (2001) and Lombard et al. (2009) and adapted to assess students' perception of and experience with the specific IVR environment in the present study. Both perceived realism (e.g., "What I experienced in the virtual classroom could also happen in a real classroom") and experienced presence (e.g., "I felt like I was sitting in the virtual classroom") were rated on 4-point rating scales ranging from 1 to 4, with higher values indicating higher levels of perceived realism and experienced presence (Cronbach's alpha values of .76 and .77, respectively). The IVR configuration had no statistically significant effect on participants' perceived realism and experienced presence; mean values ranged between 2.73 and 3.08 ($0.26 < SDs < 0.75$) for both variables across all IVR conditions.

situation; the virtual teacher introduced the topic, asked questions to include the students, presented two exercises that students had some time to think about individually, and lastly, the virtual teacher discussed the solutions to these tasks. Virtual peer learners were programmed to raise their hands in response to the teacher's questions and to respond when the virtual teacher called on them. To ensure that the virtual peer learners' hand-raising behavior was unambiguously attributed to their performance level (i.e., more hand-raising peers leading to the perception of a higher-performing class)[21], the hand-raising virtual classmates' answers were always correct, which was communicated accordingly by the virtual teacher. Participants experienced the IVR lesson from the perspective of a student in the IVR classroom surrounded by 24 virtual peer learners. Participants could raise their hands; however, they were not called on by the teacher since the whole IVR lesson was fully preprogrammed.

3.4. Apparatus

We used HTC Vive Pro Eye head-mounted displays with a refresh rate of 90 Hz and a 110° field of view for our experiment (1440 x 1600 screen resolution for each eye). To collect participants' eye-tracking data during the IVR lesson, we used the integrated Tobii eye tracker in the HTC Vive Pro Eye with a 120 Hz sampling rate and a default calibration accuracy of 0.5°-1.1° (based on a standard 5-point calibration). The IVR classroom scenario was designed and rendered using the Unreal Game Engine v4.23.1.

3.5. Study procedure

Participants took part in the experiment in a quiet room at their school in groups of up to ten. Each of the test sessions followed the same procedure, which consisted of three parts (see Figure 2 for an overview).

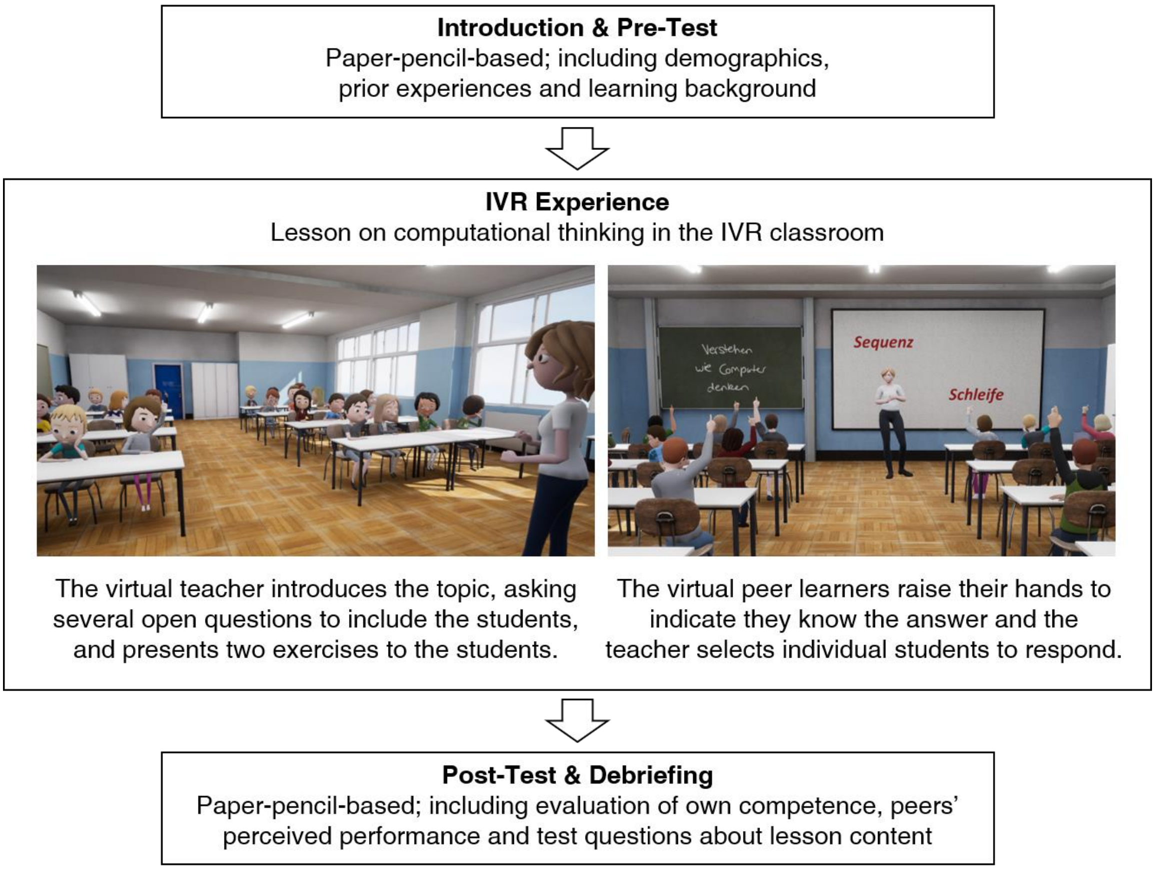
After a general introduction to the study procedure, participants completed a paper-based pretest questionnaire including demographics and relevant background variables (e.g., intelligence self-concept, initial interest in the lesson topic and previous IVR experience). Following the pretest, participants put on the head-mounted displays and experienced the 15-minute IVR lesson once the integrated eye tracker was calibrated. The IVR lesson was introduced as a learning experience that participants were free to explore as they liked. Participants were seated in desks in the real world, congruent to their virtual IVR classroom

[21] A manipulation check indicated that virtual classmates' hand-raising was significantly positively related to the perceived performance level of the class (assessed via self-reports from participants after the IVR lesson; Spearman's rho $\rho = .41$, $p < .001$). Mean differences in the perceived performance level continuously increased from 20% hand-raising ($M = 2.89$, $SD = 0.53$) to 80% hand-raising ($M = 3.45$, $SD = 0.43$).

experience; they were instructed to remain seated but otherwise behave like they would in a normal classroom situation. Participating students were unaware of the different IVR configuration conditions during the experiment. Upon completion of the IVR lesson, participants filled out the paper-based posttest questionnaire, including measures of their self-concept with reference to the IVR situation and their overall experience of the IVR experience. The testing session ended with a debriefing about the study aims and design (including information about the random assignment to different IVR configurations) after approximately 45 min in total.

Figure 2

Study Procedure and IVR Lesson Content



Note. The images depict a situation during the IVR lesson when the virtual teacher asked a question and virtual students raised their hands to indicate that they know the answer. The image on the left shows virtual avatars represented in a cartoon-style manner (and with less hand-raising), the image on the right depicts stylized virtual avatars (with more hand-raising). Participants experienced the classroom situation either from the second row from the front or the back row.

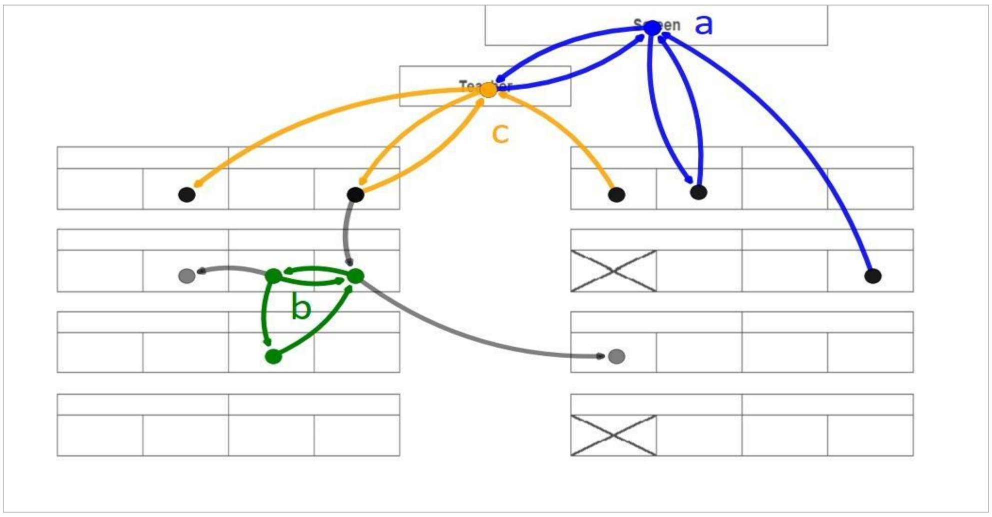
3.6. Measures

3.6.1. Structural variables describing gaze-based attention networks

In order to generate gaze-based attention networks from the gaze and head movement data, various data preprocessing steps were necessary. First, we used a technique known as gaze ray-casting (Alghamdi & Alhalabi, 2019; Pietroszek, 2018) to extract the information about what participants looked at during the IVR lesson. Because not every virtual object in the environment was of interest for our study, we afterward defined and only included specific objects of interest (OOIs) with regard to our research questions, namely the virtual peer learners, the virtual teacher, and the screen on which the instructional content was presented. We then counted participants' gaze transitions between these OOIs as gaze shifts and aggregated the number of gaze shifts across all OOIs during the whole experiment. The collected gaze information was then transformed into a graph that treated the observed OOIs as nodes and the gaze shifts between these nodes as weighted edge connections (i.e., one edge consists of a bidirectional connection defined by the frequencies of gaze shifts; see Figure 3 for a visualization). A detailed description of the pre-processing of eye-tracking data and the creation of graphs is provided in Appendix A.

Figure 3

Example Visualization of Structural Graph Variables



Note. Three different structural variables are visualized for illustrative purposes, one from each respective category. **a** = gaze centrality of the screen (visualized in blue): all incoming and outgoing edge weights are summed up into one centrality marker. **b** = clique among peers (visualized in green): all green nodes are connected to each other by edges and therefore form a clique; a connection is considered to exist if there is at least one edge between two nodes. **c** = cut size between teacher and peers (visualized in yellow): summing up all edge weights between the teacher and all peer nodes.

Based on these graphs for each participant for the full experiment duration, we calculated different structural variables that describe the graph structure and associated gaze-based attention network for each participant. The calculated structural variables can be assigned to three categories, namely gaze centrality, connectedness of gazes and uniformity of gaze distribution.

Gaze centrality was assessed via three variables regarding central OOIs in the IVR classroom: the degree centrality of (a) the peer learners, (b) the virtual teacher, and (c) the screen with the instructional content. Degree centrality (Sadria et al., 2019; Yazdan-Shahmorad et al., 2020), as a measure of the gaze centrality of the OOIs, indicates to what extent these OOIs are in the center of gaze transitions and describes the focus of attention towards these OOIs. For each node (or bundle of nodes) in the graph, degree centrality is defined as the sum of weights of all incoming and outgoing edges (or the sum of all edge weights for more than one node). To calculate degree centrality in our gaze-based attention networks, we summed up the frequency of gaze shifts from and towards the selected OOIs.

Connectedness of gazes was measured with three variables regarding so-called cliques in the gaze network. Cliques are highly connected clusters (i.e., substructures) in a graph and therefore provide information about the connectedness of gazes—i.e., the extent of gaze transitions between the OOIs—in the gaze-based attention network. Because cliques can only be calculated in undirected graphs, we transformed each directed graph into an undirected one by calculating the weight of each undirected edge as the sum of both directed edge weights. Furthermore, we calculated all maximal cliques among virtual peer learners (i.e., the subset of nodes that contains the maximal number of nodes that share an edge with every other node in the subset). Because two connected nodes build a trivial clique, we only considered cliques that contain more than two nodes. After calculating all cliques in a network, we are able to state (a) the number of cliques, and (b) their average size. Given that this part of the analysis focused specifically on visual attention towards virtual peer learners, we took this opportunity to conduct more fine-grained analyses, such as the gender composition of cliques. Therefore, we calculated (c) the proportion of boys in the observed cliques.

Uniformity of gaze distribution was measured with three different variables that describe how gaze shifts were distributed across the OOIs in the classroom. Firstly, we calculated (a) a weighted degree centrality measure (as proposed by Candeloro et al., 2016) that includes uniformity of edge weights. We consequently used the weighted degree centrality (WDC) of the screen as an indicator of how uniformly gaze was distributed from the screen to different peer learners. Secondly, we calculated (b) the cut size between the teacher/screen and peer learners as an indicator of how much students' visual attention shifted back and forth between the two versus staying on one group (e.g., students mostly focused on the teacher/screen). Cut size was calculated by summing up the edge weights of edges that pass between the two subsets (one subset being the teacher and screen and the other being all peer learners). Thirdly, we looked at the overall distribution of all edge weights in the network and tested for (c) overall uniformity. Therefore, for each participant, we stated the chi-square test statistic value calculated for a sample containing all edge weights of this person's gaze shifts.

3.6.2. Students' learning outcomes

Interest. Participants' interest in the IVR lesson was measured at posttest with six items (e.g., "I liked the topic of the IVR lesson" or "I would like to learn more about the topic of the IVR lesson") on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*), yielding a Cronbach's alpha of .91 (M = 3.18, SD = 0.69).

Situational self-concept. Participants' situational self-concept after the IVR lesson was assessed with a four-item scale that was based on the commonly used wording by Schwanzer et al. (2005) and adapted for the specific situation with virtual peer learners (e.g., "I could solve the robot tasks faster than the others" and "It was harder for me to understand the robot tasks than for the other students"). Participants indicated their responses on a 4-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*). Two items were reverse-scored and recoded accordingly, yielding an acceptable Cronbach's alpha of .69 in our sample overall (M = 3.41, SD = 0.54).

Learning. The posttest questionnaire included a short test of how much participants learned during the IVR lesson about computational thinking. The test consisted of 12 questions covering the IVR lesson content on basic computational principles (i.e., sequences and loops). Participants had to indicate whether 12 given statements were correct or incorrect (e.g., "The order of commands does not matter in a loop" [false] or "Following a recipe when cooking is an example of a sequence" [correct]). Participants were given one point for each correct answer; thus, posttest scores ranged from 0 to a maximum of 12 points. Obtained scores ranged from 4 to 12 (M = 10.45, SD = 1.59). The 12 items had a low but acceptable Cronbach's alpha of .53.

3.6.3. Covariates

Participants' gender, general intelligence self-concept (based on Schwanzer et al., 2005) and initial interest in the lesson topic of computational thinking were included as covariates in the models, as they could potentially influence participants' learning outcomes. Intelligence self-concept was measured with four items (e.g., "I often think I'm not as smart as the others"; Schwanzer et al., 2005), and initial interest in the lesson topic was measured with five items (e.g., "I would like to know more about how computer applications and robots work" or "I am interested in topics related to technology"). A four-point rating scale ranging from 1 (*not true at all*) to 4 (*absolutely true*) were used for both scales, yielding Cronbach's alpha values of .74 for intelligence self-concept (M = 3.07, SD = 0.61) and .91 for initial interest in the lesson topic (M = 3.14, SD = 0.76).

3.7. Statistical analyses

We applied three-way full factorial ANOVAs to examine differences in structural variables of students' gaze-based attention networks in the different IVR configuration conditions (RQ1, H1a-d). To answer RQ2, we used partial correlations to examine the relation between structural variables of students' gaze-based attention networks and (a) their interest in the IVR lesson, (b) their situational self-concept, and (c) their posttest learning score after the IVR lesson. We added students' gender, general intelligence self-concept and initial interest in the lesson topic as covariates, to account for the fact that these variables could potentially influence students' learning outcomes (a respective correlation matrix of the covariates and outcome variables is provided in the supplemental material). Moreover, as we sought to obtain insights into the general meaning of the gaze features used, we controlled for the IVR configuration conditions (i.e., participants' position in the classroom, the virtual avatars' visualization styles, and virtual peers' performance-related behavior) to examine the relations between students' visual attention and their learning experience after removing the influence of our experimental manipulation.

Prior to all analyses, we checked for a normal distribution of our data with the Shapiro-Wilk Test. If the Shapiro-Wilk test was significant, and graphical representations and variable skewness and kurtosis also indicated a lack of normality, we calculated Spearman's rho for non-parametric correlations. For non-parametric ANOVA procedures, we applied full-factorial aligned rank transformation using the ARTool package in R (Wobbrock et al., 2011). We used Tukey's HSD test for post-hoc comparisons and calculated partial eta squared (η_p^2) to describe effect sizes of the ANOVA, with cut-off values of ≥ 0.06 for medium and ≥ 0.14 for large effects (Cohen, 1988). All analyses were done in R (R Core Team, 2020) and we set the critical p-value and confidence intervals at an alpha level of .05 for all hypothesis tests. We report and interpret results based on both statistical significance and effect sizes.

We posted all data and data analysis scripts on the Open Science Framework under the following link: https://osf.io/pek4q/?view_only=ef151fd06ac8413a827020d4264b3c8d.

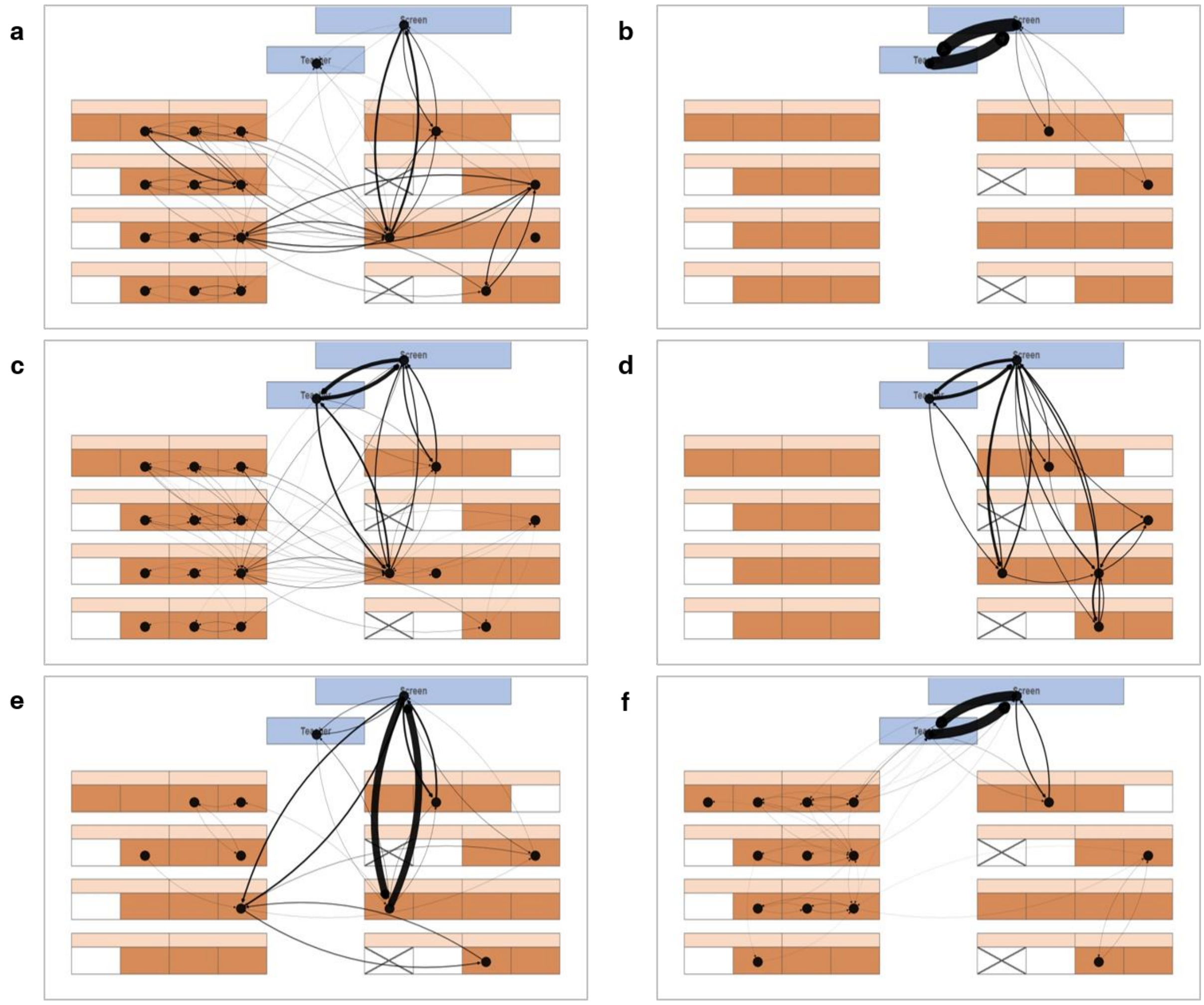
4. Results

We used different structural features to analyze students' gaze-based attention networks describing gaze centrality, the connectedness of gazes among peers and the overall uniformity of gaze distribution. Figure 4 depicts examples of gaze-based attention networks from selected students. Table 2 provides descriptive statistics and a correlation matrix for the structural features describing students' gaze-based attention networks. We found significant moderate to high correlations between almost all of the features.[22] As also depicted in Figure 4, the correlational pattern between the structural variables indicated that students' gaze-based attention networks largely reflect two types: Students tended to focus their gazes either on their peers *or* on the instructional content (i.e., the teacher and screen; see highly negative correlation of degree centrality peers with degree centrality teacher, $\rho = -.93$, $p < .001$, and degree centrality screen, $\rho = -.91$, $p < .001$). The more students' gaze centered on the instructional content (i.e., the teacher or screen), the less uniformly their gazes were distributed across the classroom, as can be seen, for instance, in the comparably high negative correlations of the uniformity of gaze distribution between the screen and peers (i.e., WDC screen) with the degree centrality of the teacher ($\rho = -.87$, $p < .001$) and screen ($\rho = -.64$, $p < .001$), in contrast to highly positive correlations with the degree centrality of peers ($\rho = .80$, $p < .001$).

[22] On the one hand, high correlations are to be expected for markers that are based on similar calculations, such as degree centrality of peers, teacher and screen, or the number and average size of cliques—particularly in light of the highly standardized environment. On the other hand, each of the selected features provides distinct information about the processing of (social) information in the classroom, as visualized in Figure 3.

Figure 4

Example Gaze-Based Attention Networks for Different Participants



Note. a–f represent the gaze-based attention networks over the course of the 15-minute IVR lesson for six selected participants. The crossed-out seats indicate the participants' position in the front (b, c, and f) versus the back (a, d and e) of the classroom. Colored seats were occupied by a virtual peer learner, white seats were empty. Black bullets represent nodes (i.e., OOIs gazed at); the width of the black lines indicates the frequency of gaze transitions between the OOIs. **a** = high gaze centrality of peers, low gaze centrality of teacher and screen. **b** = low gaze centrality of peers (no cliques), high gaze centrality of teacher and screen, low weighted degree centrality screen (uniformity in gaze distribution between screen and peers) and low cut size (transitions between teacher/screen and peers). **c** = high number and average size of cliques among peers. **d** = high weighted degree centrality screen and high uniformity of gaze distribution across all OOIs, as indicated by similar width of all connecting lines. **e** = High cut size and medium uniformity of gaze distribution across OOIs. **f** = Low uniformity and cut size, medium number and size of cliques.

Table 2

Descriptive Statistics and Correlation Matrix for Structural Variables Describing Students' Gaze-Based Attention Networks

Variable	*M* (*SD*)	*Mdn* (*MAD*)	*Min* / *Max*	1	2	3	4	5	6	7	8
1. DC peers	0.72 (0.35)	0.73 (0.39)	0.04 / 1.74	—							
2. DC teacher	0.58 (0.20)	0.58 (0.21)	0.02 / 0.97	-.93*** [-0.95, -0.91]	—						
3. DC screen	0.71 (0.18)	0.73 (0.18)	0.12 / 1.00	-.91*** [-0.93, -0.88]	.73*** [0.66, 0.79]	—					
4. N cliques peers	2.78 (2.75)	2.00 (2.97)	0.00 / 12.00	.59*** [0.50, 0.67]	-.48*** [-0.57, -0.38]	-.67*** [-0.74, -0.59]	—				
5. Avg clique size peers	2.31 (0.71)	2.38 (0.56)	0.00 / 3.58	.56*** [0.47, 0.64]	-.48*** [-0.57, -0.38]	-.62*** [-0.69, -0.53]	.92*** [0.90, 0.94]	—			
6. Proportion boys in cliques	0.53 (0.19)	0.56 (0.13)	0.00 / 1.99	.57*** [0.48, 0.65]	-.53*** [-0.62, -0.43]	-.54*** [-0.62, -0.44]	.35*** [0.24, 0.45]	.34*** [0.23, 0.44]	—		
7. WDC screen	1.93 (0.54)	1.89 (0.68)	1.05 / 4.00	.80*** [0.75, 0.84]	-.87*** [-0.90, -0.83]	-.64*** [-0.71, -0.56]	.40*** [0.29, 0.50]	.40*** [0.29, 0.50]	.53*** [0.43, 0.62]	—	
8. CS teacher/ screen – peers	1.78 (0.57)	1.75 (0.63)	0.57 / 3.45	.51*** [0.41, 0.60]	-.60*** [-0.68, -0.51]	-.28*** [-0.39, -0.16]	-.07 [-0.19, 0.05]	-.06 [-0.18, 0.06]	.23*** [0.11, 0.34]	.60*** [0.51, 0.68]	—
9. Uniformity overall GD	-4.53 (2.33)	-4.21 (2.43)	-11.27 / -0.35	.36*** [0.25, 0.47]	-.36*** [-0.46, -0.25]	-.25*** [-0.35, -0.13]	-.25*** [-0.36, -0.13]	-.24*** [-0.35, -0.12]	.11 [-0.22, 0.01]	.30*** [0.19, 0.41]	.67*** [0.59, 0.74]

Note. Mean values and standard deviations (*M* and *SD*) as well as medians and median absolute deviations (*Mdn* and *MAD*) are reported. Variables 1-9 are non-normally distributed; thus, Spearman's rho is reported. 95% confidence intervals are given in brackets. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.
*** $p < .001$.

4.1. How do different IVR configurations affect the structure of students' gaze-based attention networks in the IVR classroom?

The first research question (RQ1) asked how the IVR classroom configurations (i.e., participants' position, visualization style of virtual avatars, and virtual peers' performance-related behavior) affect students' gaze-based attention networks during the IVR lesson. In this section, we describe the results of three-way full-factorial aligned rank transformation ANOVAs examining the effects of the IVR configurations on the structural features of students' gaze-based attention networks (i.e., gaze centrality, connectedness of gazes among peers, uniformity of gaze distribution). We only report detailed statistics for significant results in the main text; full statistics for all analyses are provided in the supplemental material. Table 3 provides an overview of the observed main effects.

Table 3

Summary of Main Effects of IVR Configuration Conditions on Structural Network Features

	Sitting position	Avatar visualization style	Proportion of hand-raising peer learners
DC peers	**front < back**	**cartoon > stylized**	20% > 65% < 80%
DC teacher	**front > back**	**cartoon < stylized**	20% < 65% > 80%
DC screen	**front > back**	**no difference**	20% < 65% > 80%
N cliques peers	**front < back**	**cartoon > stylized**	no difference
Avg clique size peers	**front < back**	**cartoon > stylized**	no difference
Proportion boys in cliques	**front < back**	no difference	no difference
WDC screen	**front < back**	**cartoon > stylized**	20% > 35% > 65% < 80%
CS teacher/screen – peers	**front < back**	**cartoon > stylized**	no difference
Uniformity overall	**front < back**	no difference	no difference

Note. Only statistically significant differences are shown. < and > indicate the direction of the effect. Results in bold represent findings in line with hypotheses. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.

4.1.1. Effects of students' sitting position on the structure of their gaze-based attention networks

Students' position in the IVR classroom had a significant effect on all of the structural features describing students' gaze-based attention networks (descriptive statistics in Table 4).

First, students sitting in the front position showed a significantly different focus of gaze transitions on their virtual peer learners, the virtual teacher and the screen with the lesson content (i.e., measured by degree centrality) compared to students in the back sitting position: Degree centrality of the virtual peers was significantly higher when students were located in the back row of the IVR classroom, $F(1,258) = 138.55$, $p < .001$, $\eta_p^2 = .35$. In turn, degree centrality of the virtual teacher was significantly higher for students who were located in the front of the IVR classroom, $F(1,258) = 204.07$, $p < .001$, $\eta_p^2 = .44$, and similarly the screen was more the focus of students' gaze transitions when they were sitting in the front, $F(1,258) = 60.56$, $p < .001$, $\eta_p^2 = .19$.

Table 4

Descriptive Statistics for Structural Network Features in Different Sitting Positions

	Front (N = 122)		Back (N = 152)	
	M (SD)	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*
DC peers	0.5 (0.27)	0.45 (0.31)	0.89 (0.30)	0.89 (0.25)
DC teacher	0.71 (0.14)	0.72 (0.16)	0.47 (0.17)	0.46 (0.17)
DC screen	0.79 (0.15)	0.83 (0.14)	0.64 (0.18)	0.65 (0.15)
N cliques peers	2.21 (2.39)	2.00 (2.97)	3.24 (2.93)	3.00 (4.45)
Avg clique size peers	2.20 (0.70)	2.32 (0.48)	2.40 (0.71)	2.44 (0.58)
Proportion boys in cliques	0.45 (0.17)	0.50 (0.11)	0.60 (0.18)	0.63 (0.09)
WDC screen	1.50 (0.28)	1.43 (0.22)	2.27 (0.44)	2.29 (0.39)
CS teacher/screen – peers	1.45 (0.47)	1.38 (0.43)	2.05 (0.51)	2.06 (0.50)
Uniformity overall	-5.34 (2.52)	-5.28 (2.72)	-3.89 (1.94)	-3.86 (1.96)

Note. Mean values and standard deviations (*M* and *SD*, respectively) as well as medians and median absolute deviations (*Mdn* and *MAD*, respectively) are reported. Values for each of the configuration conditions are averaged across the other conditions. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.

Second, with regard to the gaze activity among virtual peer learners, we found that the front vs. back position in the IVR classroom led to significant differences in the number and average size of cliques and the proportion of boys in the observed cliques: The number of cliques among peers and the average size of these cliques were significantly higher when students were positioned in the back of the IVR classroom, whereby also the proportion of boys in the observed cliques was significantly higher in the back position; $F(1,258) = 4.50$, $p = .035$, $\eta_p^2 = .02$ and $F(1,258) = 8.06$, $p = .005$, $\eta_p^2 = .03$ and $F(1,258) = 101.35$, $p < .001$, $\eta_p^2 = .28$ for the number and average size of cliques and the proportion of boys in the observed cliques, respectively.

Third, results showed higher levels of uniformity in gaze distribution for students in the back sitting position for all three indicators of uniformity: Students distributed their gazes more evenly from the screen to different peers (i.e., weighted degree centrality screen) in the back position, $F(1,258) = 333.16$, $p < .001$, $\eta_p^2 = .56$, and had more gaze transitions between the instructional content (teacher/screen) and peers (i.e., cut size) when sitting in the back, $F(1,258) = 109.23$, $p < .001$, $\eta_p^2 = .30$. Moreover, students' gazes were distributed more uniformly across all observed OOIs in the IVR classroom when they were sitting in the back; $F(1,258) = 33.11$, $p < .001$, $\eta_p^2 = .11$.

In sum, these results fully support our Hypothesis ***H1a***: The position in the back of the virtual classroom led to more gaze centrality on virtual peer learners (and less on the virtual teacher and screen), more connectedness of gazes among peers and more uniformly distributed gazes between OOIs in the IVR classroom (see summary in Table 3).

4.1.2. Effects of virtual avatar visualization style on the structure of students' gaze-based attention networks

The virtual avatar visualization style had a significant effect on a number of the structural features describing students' gaze-based attention networks (descriptive statistics in Table 5).

Table 5

Descriptive Statistics for Structural Network Features in Different Avatar Visualization Styles

	Cartoonish (N = 150)		Stylized (N = 124)	
	M (SD)	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*
DC peers	0.79 (0.35)	0.80 (0.36)	0.63 (0.32)	0.60 (0.38)
DC teacher	0.52 (0.21)	0.49 (0.22)	0.65 (0.16)	0.63 (0.17)
DC screen	0.69 (0.18)	0.71 (0.17)	0.73 (0.19)	0.76 (0.19)
N cliques peers	3.24 (2.80)	3.00 (2.97)	2.23 (2.59)	1.00 (1.48)
Avg clique size peers	2.44 (0.62)	2.48 (0.50)	2.15 (0.79)	2.26 (0.39)
Proportion boys in cliques	0.54 (0.16)	0.57 (0.12)	0.52 (0.22)	0.55 (0.13)
WDC screen	2.04 (0.54)	2.08 (0.69)	1.80 (0.51)	1.70 (0.52)
CS teacher/screen – peers	1.88 (0.60)	1.87 (0.65)	1.66 (0.50)	1.66 (0.50)
Uniformity overall GD	-4.56 (2.43)	-4.17 (2.33)	-4.50 (2.20)	-4.32 (2.54)

Note. Mean values and standard deviations (*M* and *SD*, respectively) as well as medians and median absolute deviations (*Mdn* and *MAD*, respectively) are reported. Values are averaged across the other configuration conditions. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.

With regards to the degree to which OOIs were at the focus of students' gaze transitions (i.e., measured by degree centrality), we found greater degree centrality of peer learners when they were presented in cartoon style compared to a more stylized visualization, $F(1,258) = 20.46$, $p < .001$, $\eta_p^2 = .07$. In contrast, degree centrality of the teacher was higher in the stylized visualization compared to the cartoonish one, $F(1,258) = 52.03$, $p < .001$, $\eta_p^2 = .17$. We found no statistically significant differences for the degree centrality of the screen based on different avatar visualizations.

Turning to the gaze activity among virtual peer learners, the number and the average size of cliques among peers differed significantly between the visualization styles of the virtual avatars in the IVR. Both the number of cliques among peers and the average clique size were statistically significantly higher when virtual peer learners were visualized in cartoon style; $F(1,258) = 9.81$, $p = .002$, $\eta_p^2 = .04$, and $F(1,258) = 12.47$, $p < .001$, $\eta_p^2 = .05$ for the number and average size of cliques, respectively. The proportion of boys in the observed cliques was not affected by the visualization style of virtual avatars.

Moreover, we found more evenly distributed gazes between the screen and peers (i.e., higher weighted degree centrality screen) when avatars were visualized in cartoon style, $F(1,258) = 34.82$, $p < .001$, $\eta_p^2 = .12$. Similarly, the results showed more gaze transitions between the teacher/screen and peers (i.e., higher cut size) for the cartoonish visualization, $F(1,258) = 14.30$, $p < .001$, $\eta_p^2 = .05$. The visualization style of virtual avatars had no effect on the overall uniformity of gaze distribution across OOIs in the IVR classroom.

In sum, these results partially support our Hypothesis ***H1b*** (see summary in Table 3).

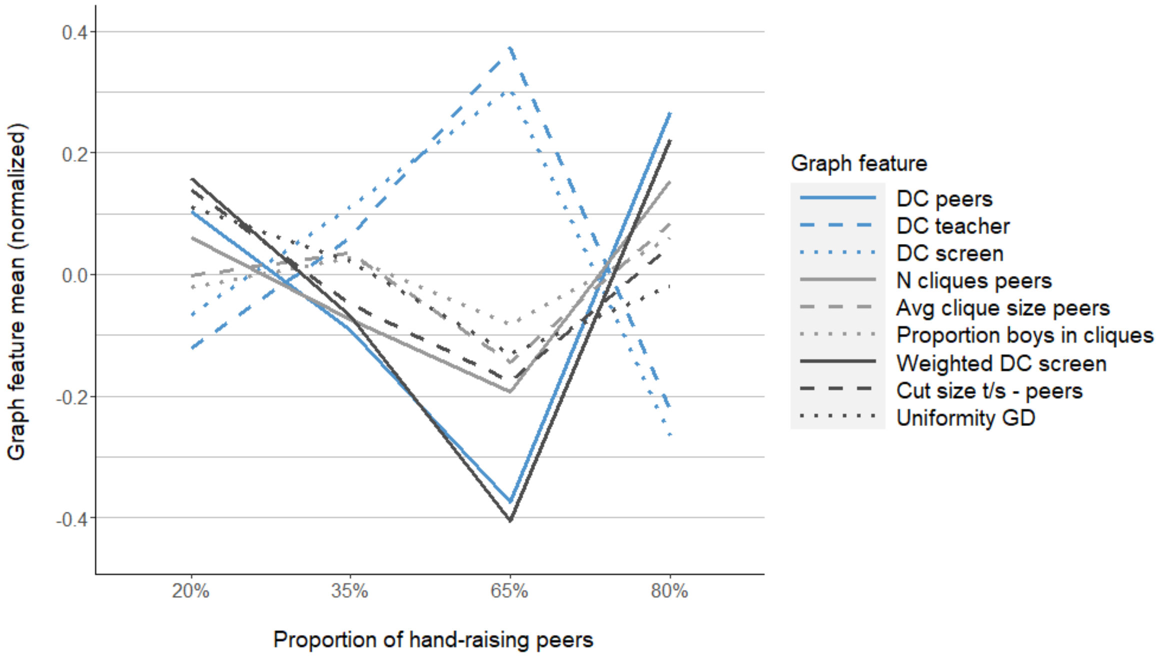
4.1.3. Effects of virtual peers' hand-raising behavior on the structure of students' gaze-based attention networks

The manipulation of virtual peers' performance-related behavior, specifically their hand-raising behavior, had a statistically significant effect on students' gaze-based attention networks. In particular, the degree to which virtual peer learners, the virtual teacher and the screen with the lesson content were the focus of students' gaze transitions (measured by degree centrality) was affected by the hand-raising conditions; $F(3,258) = 7.76$, $p < .001$, $\eta_p^2 = .08$ and $F(3,258) = 8.09$, $p < .001$, $\eta_p^2 = .09$ and $F(3,258) = 4.94$, $p = .002$, $\eta_p^2 = .05$ for the degree centrality of virtual peer learners, the virtual teacher and the screen, respectively. As can be seen in Figure 5, descriptively speaking, the degree centrality of virtual peers was highest in the 'extreme' hand-raising conditions of 20% and 80% (see solid blue line), whereas the degree centrality of the teacher and the screen showed the opposite pattern (see dotted and dashed blue lines).

Tukey's HSD test for multiple comparisons showed that the degree centrality of peers was significantly higher in the 20% hand-raising condition compared to the 65% condition ($p = .004$) and significantly lower for the 65% compared to the 80% hand-raising condition ($p < .001$). In turn, the degree centrality of the teacher and screen were significantly lower in the 20% hand-raising condition compared to the 65% condition ($p = .002$ and $p = .039$ for the teacher and screen, respectively) and significantly higher for 65% compared to 80% hand-raising ($p < .001$ and $p = .002$ for the teacher and screen, respectively).

Figure 5

Normalized Mean Values of Gaze Features by Hand-Raising Conditions



Note. Values are averaged across sitting position and avatar visualization. DC = Degree Centrality; N = Number; Avg = Average; t/s = teacher/screen; GD = Gaze Distribution.

With regard to gaze activity among virtual peer learners, the hand-raising conditions had no statistically significant effect on any of the respective features (i.e., number and average size of cliques as well as proportion of boys in the observed cliques; see light grey lines in Figure 5).

Lastly, whereas the hand-raising behavior of virtual peers also had no effect on the amount of transitions between teacher/screen and peers (i.e., cut size) and the overall uniformity of gaze distribution across OOIs in the IVR classroom, the results indicated significantly different levels of weighted degree centrality of the screen in the different hand-raising conditions, $F(3,258) = 12.92$, $p < .001$, $\eta_p^2 = .13$. A similar pattern as for the gaze centrality markers was observed: Weighted degree centrality of the screen (i.e., uniformity of gaze distribution between screen and peers) was highest in the 20% and 80% hand-raising conditions (see solid dark grey line in Figure 5). Tukey's HSD test for multiple comparisons indicated statistically significant differences between 20% and 65% ($p < .001$), 35% and 65% ($p = .028$), 35% and 80% ($p = .017$), and 65% and 80% ($p < .001$).

Table 6

Descriptive Statistics for Structural Network Features in Different Hand-Raising Conditions

	20% (*N* = 72)		35% (*N* = 64)		65% (*N* = 60)		80% (*N* = 78)	
	M (SD)	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*	*M (SD)*	*Mdn (MAD)*
DC peers	0.75 (0.35)	0.80 (0.35)	0.68 (0.32)	0.75 (0.35)	0.59 (0.32)	0.56 (0.33)	0.81 (0.36)	0.79 (0.39)
DC teacher	0.55 (0.19)	0.51 (0.20)	0.59 (0.20)	0.59 (0.22)	0.65 (0.18)	0.67 (0.20)	0.53 (0.20)	0.52 (0.21)
DC screen	0.70 (0.18)	0.72 (0.18)	0.73 (0.16)	0.73 (0.17)	0.76 (0.17)	0.82 (0.13)	0.66 (0.19)	0.68 (0.19)
N cliques peers	2.94 (3.06)	2.00 (2.97)	2.58 (2.72)	2.00 (2.97)	2.25 (2.45)	2.00 (2.97)	3.21 (2.66)	3.00 (2.97)
Avg clique size peers	2.31 (0.77)	2.32 (0.47)	2.33 (0.66)	2.38 (0.56)	2.21 (0.68)	2.33 (0.49)	2.37 (0.72)	2.44 (0.45)
Proportion boys in cliques	0.53 (0.20)	0.57 (0.11)	0.53 (0.17)	0.55 (0.11)	0.51 (0.20)	0.54 (0.18)	0.54 (0.19)	0.57 (0.14)
WDC screen	2.01 (0.54)	2.04 (0.69)	1.89 (0.57)	1.83 (0.67)	1.71 (0.45)	1.58 (0.40)	2.05 (0.52)	2.16 (0.59)
CS teacher/screen – peers	1.86 (0.57)	1.82 (0.64)	1.75 (0.63)	1.72 (0.59)	1.68 (0.55)	1.74 (0.57)	1.81 (0.52)	1.77 (0.65)
Uniformity overall GD	-4.28 (2.15)	-3.92 (2.00)	-4.48 (2.23)	-4.08 (2.04)	-4.84 (2.17)	-4.93 (2.22)	-4.58 (2.67)	-4.13 (2.81)

Note. Mean values and standard deviations (*M* and *SD*, respectively) as well as medians and median absolute deviations (*Mdn* and *MAD*, respectively) are reported. Values are averaged across the other configuration conditions. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution

In sum, the pattern of results for the effects of peer learners' hand-raising on students' gaze-based attention networks in the IVR classroom was different than we hypothesized (***H1c***). Descriptive statistics for the structural features of students' gaze-based attention networks in the different hand-raising conditions are given in Table 6. Detailed statistics for the post-hoc comparisons can be found in the supplemental material.

4.1.4. Interaction effects of IVR configurations on the structure of students' gaze-based attention networks

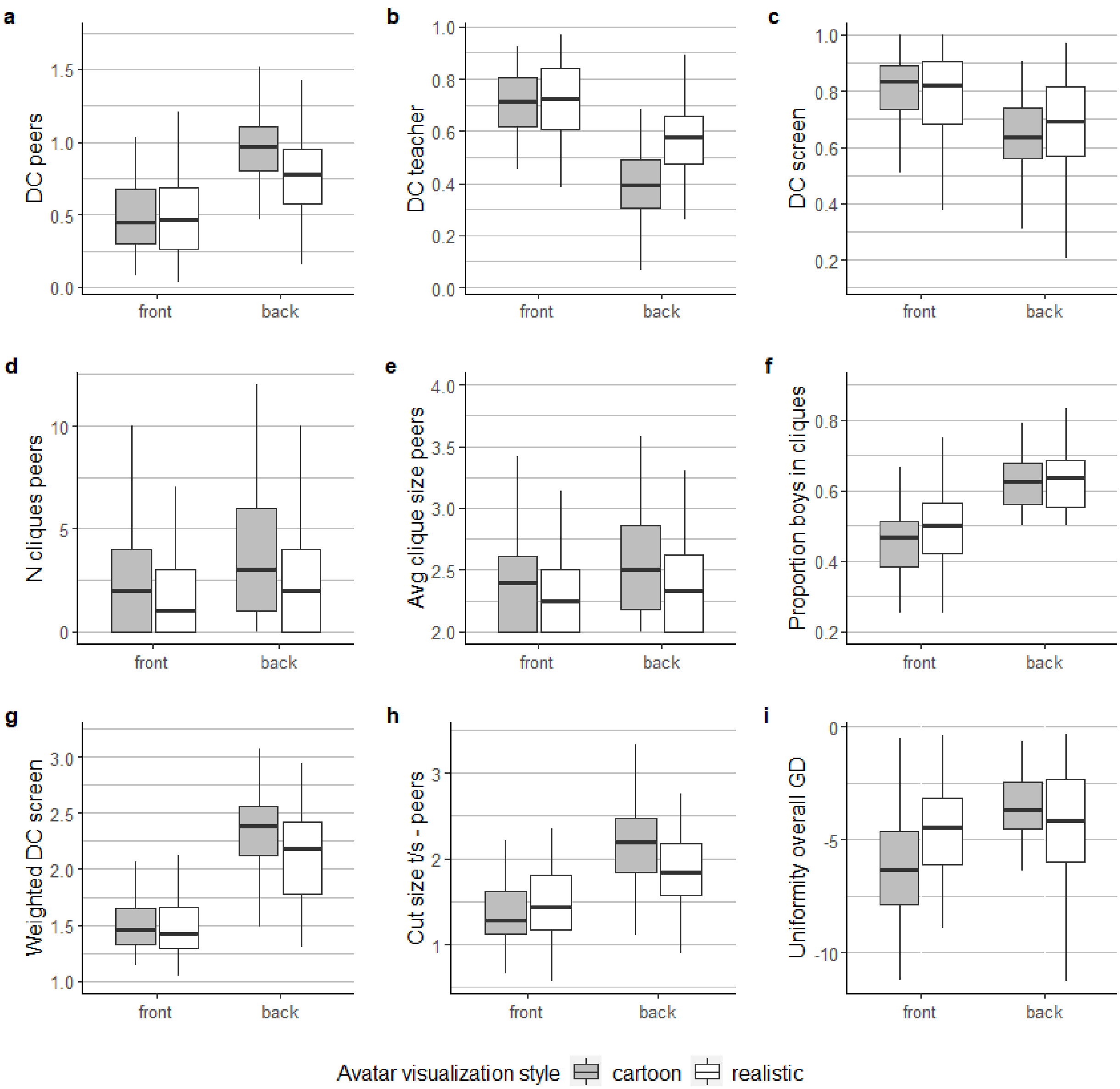
In addition to the main effects described above, we explored interaction effects of the different IVR configuration conditions on students' gaze-based attention networks. Whereas the effects of virtual peers' hand-raising behavior on students' gaze-based attention networks (see main effects in Section 4.1.1) were not affected by students' sitting position, the results showed that the effects of the virtual avatar visualization style on students' gaze-based attention networks (see main effects in Section 4.1.2) were more pronounced when participants were sitting in the back of the IVR classroom. The results thus provided partial support for Hypothesis ***H1d***.

As Figure 6a shows, the focus of students' gaze transitions on virtual peer learners (i.e., degree centrality peers) visualized in a cartoon-style way was significantly greater in the back position; $F(1,258) = 11.87$, $p < .001$, $\eta_p^2 = .04$. Similarly, as can be seen in Figure 6b, the degree centrality of the stylized teacher was significantly lower when sitting in the back; $F(1,258) = 19.77$, $p < .001$, $\eta_p^2 = .07$. There were no significant interaction effects of the IVR configuration conditions for connectedness of gazes (see Figure 3d-e).

Moreover, as Figures 6g-i show, the higher levels of the uniformity markers in the cartoon-style visualization of virtual avatars were particularly pronounced in the back position. The interaction effects were small compared to the main effects, but indicated statistically significant differences for the weighted degree centrality of the screen, $F(1,258) = 6.08$, $p = .014$, $\eta_p^2 = .02$, cut size, $F(1,258) = 12.50$, $p < .001$, $\eta_p^2 = .05$, and overall uniformity of gaze distribution, $F(1,258) = 21.44$, $p < .001$, $\eta_p^2 = .08$.

Figure 6

Boxplots of Structural Network Features by Sitting Position and Avatar Visualization



Note. Values are averaged across hand-raising conditions. DC = Degree Centrality; N = Number; Avg = Average; t/s = teacher/screen; GD = Gaze Distribution.

4.2. How does the structure of students' gaze-based attention networks relate to their learning experience in the IVR classroom?

The second research question asked how students' gaze-based attention networks in the IVR classroom relate to (a) their interest in the IVR lesson on computational thinking, (b) their evaluation of their own competence (i.e., situational self-concept) and (c) their performance on a test about the IVR lesson content afterwards. In this section, we report results of partial correlations between these outcomes and markers of students' gaze-based attention networks (i.e., structural variables describing gaze centrality, gaze connectedness among peers and overall uniformity of gaze distribution), controlling for students' gender, general intelligence self-concept and initial interest in the lesson topic as well as the IVR configuration conditions. Table 7 provides an overview of the results.

Table 7

Partial Correlations of Gaze-Based Features with Interest in the Lesson, Situational Self-Concept, and Posttest Score

	Interest in IVR lesson	Situational self-concept	Posttest score
DC peers	-.14* [-0.26, -0.02]	-.10 [-0.21, 0.02]	-.12 [-0.24, 0.00]
DC teacher	.11 [-0.01, 0.23]	.07 [-0.05, 0.19]	.06 [-0.07, 0.17]
DC screen	.15* [0.04, 0.27]	.10 [-0.02, 0.21]	.13* [0.02, 0.25]
N cliques peers	-.16** [-0.27, -0.04]	-.06 [-0.18, 0.06]	-.04 [-0.16, 0.08]
Avg clique size peers	-.17** [-0.28, -0.05]	-.04 [-0.16, 0.08]	.00 [-0.12, 0.12]
Proportion boys in cliques	-.18** [-0.30, -0.06]	-.12* [-0.24, 0.00]	-.11 [-0.23, 0.01]
WDC screen	-.10 [-0.22, 0.02]	.00 [-0.12, 0.12]	-.03 [-0.15, 0.09]
CS teacher/screen – peers	.07 [-0.05, 0.19]	-.04 [-0.16, 0.08]	.01 [-0.11, 0.13]
Uniformity overall GD	.09 [-0.03, 0.21]	-.01 [-0.13, 0.11]	-.07 [-0.19, 0.05]

Note. Partial correlations controlling for gender, intelligence self-concept and initial interest in the lesson topic computational thinking as well as the IVR classroom configuration conditions. Variables are non-normally distributed; thus, Spearman's rho is reported. The Bonferroni correction was used to adjust for multiple significance tests. 95% confidence intervals are given in brackets. DC = Degree Centrality; N = Number; Avg = Average; WDC = Weighted Degree Centrality; CS = Cut Size; GD = Gaze Distribution.
** $p < .01$. * $p < .05$.

The results indicated that markers of students' gaze-based attention were most consistently related to their interest in the IVR lesson in the present study: Higher gaze centrality on the screen where the lesson content was presented was associated with higher interest in the IVR lesson topic ($\rho = .15$, $p = .011$); conversely, the greater the gaze centrality on virtual peers, the lower the reported interest in the IVR lesson topic (Spearman's rho $\rho = -.14$, $p = .021$). Similarly, students' interest in the IVR lesson topic was negatively related to the number of observed cliques among peers ($\rho = -.16$, $p = .009$), the average clique size ($\rho = -.17$, $p = .007$) and the proportion of boys in the observed cliques ($\rho = -.18$, $p = .003$).

In addition, only one gaze-based feature each exhibited a small statistically significant relation with students' evaluation of their own competence in the IVR lesson (i.e., situational self-concept) and their performance on the posttest. The proportion of boys in the observed cliques was significantly related to students' situational self-concept: The more boys were in the observed cliques, the lower a student's situational self-concept (ρ = -.12, p = .043). In turn, degree centrality of the screen was positively related to students' performance on the posttest: The more students' gaze-based networks centered on the screen, the better their performance on the posttest (ρ = .13, p = .028).

In sum, the results partly supported Hypotheses ***H2a*** and ***H2b***.

5. Discussion

The present study aimed to answer central questions about the configuration of IVR classrooms for educational research and practice and therefore examined how different IVR classroom configuration features affect how students process different types of (social) information provided in the IVR classroom (RQ 1). We focused on three IVR classroom configuration features that we consider particularly important with regard to students' gaze-based attention distribution in the IVR classroom, namely (a) students' positioning in the front vs. back of the IVR classroom, (b) the visualization style of virtual avatars as cartoonish vs. stylized, and (c) virtual peers' performance-related behavior in terms of different proportions of hand-raising students. Students' visual attention behavior was assessed via students' eye-tracking data, more specifically via features reflecting the structure of students' gaze-based attention networks (i.e., the gaze centrality on OOIs, connectedness of gazes among OOIs and uniformity of gaze distribution across OOIs). The results showed statistically significant differences between the IVR classroom configuration conditions for all structural features of students' gaze-based attention networks in the classroom. To gain a more in-depth understanding of the structural features, in a second step, we examined how the structure of students' gaze-based attention networks relates to how students experienced the IVR classroom scenario (RQ 2). The results showed statistically significant relations to students' interest in the IVR lesson as well as their evaluation of their own competence (i.e., situational self-concept) and performance after the IVR lesson. In the following sections, we discuss our findings in more detail.

5.1. Implications of students' gaze-based attention networks for the configuration of IVR classrooms

We examined students' gaze-based attention networks in an IVR classroom with different configuration conditions to obtain in-depth and objectively measurable insights into how different IVR classroom features affect how students attend to (social) information in the IVR classroom scenario. Taken together, our findings indicate that students' sitting position as well as the visualization style and performance-related behavior of virtual avatars in an IVR classroom need to be carefully considered when using IVR for learning purposes or experimental classroom research. The present study's findings have important implications for educators and scholars aiming to select the best IVR classroom configuration.

First, regarding participants' position in the IVR classroom, our findings indicated that positioning students in the front of an IVR classroom led to gaze-based attention networks that were more centered on the instructional content, whereas a position in the back was associated with a more comprehensive perception of all (social) information provided in the IVR classroom. Extending the findings of existing IVR studies suggesting better learning outcomes when sitting in the front of an IVR classroom (Bailenson et al., 2008; Blume et al., 2019), our results provide evidence that, indeed, sitting in the front of an IVR classroom centers students' gaze more on the teacher and instructional content. Although this finding might be intuitive, considering that students sitting in the back have the whole class of peers in front of them, our results also indicated that students sitting in the back did not just focus more on peer learners, but distributed their attention more evenly across the classroom in general. More balanced gaze transitions from the instructional content on the screen to different peers and more gaze transitions back and forth between the teacher or screen and peer learners rather than solely focusing on the instructional content might be an indication of more integrated (social) information processing in the classroom (Hutmacher, 2019; Jarodzka et al., 2017). Notably, in the present IVR lesson, the learning content was mainly provided by the virtual teacher and on the screen in the front of the classroom; therefore, students' attentional focus on the teacher and screen was desirable with regard to learning outcomes. However, given that peer learners can serve as an important source of information during instruction (e.g., Dijkstra et al., 2008), a sitting position in the back of the classroom might be more beneficial for learning in cases where virtual classmates are designed to be actively involved in the process of knowledge acquisition (e.g., as role models, to clarify misconceptions, etc.).

Second, regarding the visualization style of the virtual avatars, our findings indicate that for our sample of sixth graders, visualizing peer learners in a cartoon style was not just more cost- and time-efficient, but yielded no considerable disadvantages compared to a more realistic (i.e., stylized) visualization of peers: In fact, the students showed higher visual attention focus and gaze activity on cartoonish virtual peer learners, with particularly pronounced effects in the back sitting position. Notably, alongside existing explanations for these findings (e.g., cartoonish peer learners are unusual and therefore attract more attention and cartoonish peers have larger head sizes which leads to increased visual attention; Bozkir et al., 2021; Gao et al., 2021), the results of the present study point to an additional important aspect: When virtual avatars were visualized in cartoon style, we found (a) more equally distributed gazes between and screen and different peers and (b) more gaze transitions between instructional content (i.e., teacher and screen) and virtual peer learners, indicating that cartoon-style learners do not just attract attention to themselves, but are more engaging for students in an IVR classroom in general. This finding is not just important given that programming costs increase exponentially as virtual avatars become increasingly realistic, it also points to potential affordances of cartoonish characters when aiming to design IVR classroom environments that invite high engagement with virtual avatars (e.g., in collaborative learning scenarios or with virtual classmates as emotional support).

Third, with regard to peer learners' performance-related behavior, our findings indicated that virtual peer learners' hand-raising had the greatest effect on students' visual attention distribution in the IVR classroom when it was most salient and unambiguous (i.e., a clear minority or majority of peers raising their hands). Against our expectation that more hand-raising would lead to more gaze centrality on peers (and respectively less on the teacher and screen), we found the highest gaze centrality on peers in the 'extreme' conditions of 20% and 80% hand-raising (and respectively the highest gaze centrality on the teacher and screen in the more moderate conditions of 35% and 65% hand-raising). Importantly, these effects were not affected by sitting position, suggesting that the most salient hand-raising conditions of 20% and 80% were recognized most by students regardless of whether they were positioned in the front or the back of the IVR classroom. Indicating that 'social manipulations' in an IVR classroom are particularly effective when they are very clearly interpretable (i.e., almost none or pretty much all peers are raising their hands), this finding has important implications for the design of peer learners' behavior in IVR classrooms in both educational research and practice. More specifically, this finding suggests that peers' (performance-related) behavior needs to be configured to be as unambiguous as possible (a) to investigate respective effects of peer behavior on student learning, and (b) to strategically deploy respective behaviors in the design of virtual peer learners as pedagogical agents in IVR classroom-based learning applications (see e.g., Bailenson et al., 2008; Hudson & Hurter, 2016; Makransky, Wismer, et al., 2019).

5.2. Gaze-based attention networks as indicators of students' (social) information processing

Given that the use of graph-based analysis is a relatively new approach for analyzing gaze data and visual attention, especially in an IVR classroom setting and in relation to students' learning experiences, we were interested in how the structure of students' gaze-based attention networks relates to central outcome variables in the context of classroom learning (i.e., students' interest, situational self-concept and performance).

In line with our expectations, we found significant relations between students' learning experiences in the IVR classroom and their gaze centrality on peers, the teacher and the screen as well as with the connectedness of gazes among peers. Notably, the examined structural features of students' gaze-based attention networks allowed us to capture specific aspects of social information processing, such as gaze centrality on certain objects of interest or visual attention focus on different subgroups (e.g., the proportion of boys in the observed cliques). In the end, we found relations between educational outcomes and the structure of students' gaze-based attention networks exclusively for features describing visual attention tied to different objects of interest (e.g., degree centrality on the screen, proportion of boys in observed cliques), whereas more general descriptions of students' gaze behavior (i.e., markers of uniformity) were not related to any of the examined educational outcomes.

As expected, the more interested students were in the IVR lesson content, the more they focused on the instructional content and the less they processed social information provided by their peers. Accordingly, students' test performance after the IVR lesson was positively related to their visual attention focus on the screen. Given that everything necessary to obtain a good test score was presented on the screen, this finding is in line with our expectations. At the same time, considering that the most important content of the IVR lesson was also presented orally by the teacher and the audio was the same in all IVR configurations, it is not surprising that the effect of visual attention on the lesson content was comparably small. Notably, students' performance after the IVR lesson was not related to their visual attention on peers. On the one hand, this finding might be considered reassuring given the potential detrimental effect of peers as distractions from the instructional content; on the other hand, considering the effect of only the 'extreme' hand-raising behaviors on students' gaze-based attention networks, this finding additionally highlights that the manipulation of hand-raising behavior in the present study was not ‘powerful’ enough to make use of potential beneficial effects of peers—for instance, as pedagogical agents. Hence, it might be worthwhile to implement fewer but very salient peer avatars (in line with suggestions by Liao et al., 2019).

Regarding students’ situational self-concept, only the proportion of boys in the observed cliques exhibited a negative relation to how students evaluated their own competence during the IVR lesson. In line with common assumptions in research on reference group effects, we expected similar findings for the degree centrality of peer learners as well as for the number and average size of cliques among virtual peers. Hence, the present study’s finding highlights the role of very specific social information for students’ self-evaluations. Although the observed effect is small and needs further investigation in future studies, we argue that this result is particularly interesting given that our IVR lesson concerned the topic of computational thinking, which might be associated with gender stereotypes that affect who students compare themselves to and how they consequently evaluate themselves (Fryer & Levitt, 2010; Plieninger & Dickhäuser, 2013; Preckel et al., 2008; Tiedemann, 2000).

5.3. Limitations and future directions

In the present study, we applied graph theory and network analysis to students’ eye-tracking data from an IVR lesson to examine differences in students’ gaze-based attention networks in an IVR classroom with different configurations. We manipulated three central configuration features of the IVR classroom (i.e., students’ position, the visualization style of virtual avatars, and performance-related behavior of virtual peer learners) and examined central structural variables describing students’ gaze-based attention networks in three categories (i.e., gaze centrality, connectedness of gazes among peer learners, and uniformity of gaze distribution across OOIs). Notably, although our approach yielded a number of important findings, we would also like to point out some limitations that provide great potential for future research regarding the configurations of IVR classroom environments and students’ individual responses to them.

In terms of IVR classroom design, we would like to highlight four critical aspects. First, we applied a neutral design of the classroom environment and focused on specificities of virtual avatar design. However, it should be noted that the overall IVR classroom design (e.g., wall color, posters, lighting, etc.) provides many additional opportunities to further guide students’ attentional focus and affect their perception of the IVR classroom scenario (see, e.g., Cheryan et al., 2011). Second, whereas we found significant positive effects of the cartoonish avatar visualization on students' overall engagement with virtual peer learners, it needs to be considered that our sample consisted of sixth graders who might have been more engaged with cartoon learners in comparison to adults or older students. Future research should extend these findings and examine the effects in different age groups. Third, in order to render 24 virtual peer agents, we needed to keep the visual realism under a certain level; hence, even our 'more realistic' (i.e., stylized) avatars do not represent the highest degree of realism that is currently possible. In addition, we only varied the visualization style of the virtual avatars; both visualization styles were based on the same motion captures and therefore, the avatars' movements and gestures were identical across the different visualizations. In light of previous work demonstrating the importance of a good match between behavioral and photographic realism (Bailenson et al., 2005; Garau et al., 2003), it seems worthwhile for future research to further explore different visualization styles of peer learners in combination with different simulations of performance-related behaviors. Fourth, we only varied the appearance of the virtual avatars of the peer learners and the teacher, whereas the participating students were not represented by an avatar in our IVR classroom. In light of the substantive body of research examining the effects of self-representation via avatars on users' behavior and experience in IVR (see meta-analysis by Ratan et al., 2020), implementing representations of participating students in the IVR classroom seems worth investigating further.

With regard to the virtual peer learners, we manipulated their performance-related behavior via hand-raising as an indicator of their performance and overall behavioral engagement (Chang et al., 2019; Lee et al., 2014). Based on the present study's findings suggesting (a) an effect of peers' hand-raising on students' gaze-based attention networks in the classroom and (b) a relation between gaze-based attentional focus on peers and central learning outcomes, we argue that future research should extend this line of research and consider additional variations of peer behavior and classroom composition. In addition, given that the proportion of observed boys also had an effect on students' self-evaluations, future research should make use of the affordances of IVR to examine gender differences with regard to peer effects.

Moreover, our IVR was fully preprogrammed, which allowed for maximum standardization and therefore systematic insights into students' IVR experience. However, we believe that the implementation of some interaction options for participating students might provide additional valuable insights into reference group effects in an interactive yet standardized setting. For instance, Liao et al. (2019) demonstrated the impact of virtual classmates on students' learning by implementing interactive virtual classmates with timeanchored comments and behaviors based on content and valence analyses of participants' prior comments during instruction. Combining such approaches with analyses of students' actual gaze-based attentional networks in the classroom seems like a promising avenue to gain insights into (a) how students make use of (social) information provided in the IVR classroom and (b) how the ideal IVR classroom for student learning should consequently be configured.

Lastly, our IVR lesson lasted only 15 minutes, and we aggregated students' gaze-based attention networks over the entire lesson period in order to gain insights into their processing of (social) information provided in the IVR classroom. Whereas our approach yielded important insights into the effects of different IVR configurations (see Section 5.1) and the use of gaze-based attention networks as indicators of students' processing of (social) information in the IVR classroom (see Section 5.2), we argue that future research should see whether our findings replicate in other and longer IVR classroom scenarios. Moreover, the graph-based analysis of gaze data has great potential for additionally examining dynamics and changes in students' gaze-based attention networks (see, e.g., Wang et al., 2019), such as whether attentional focus on the teacher decreases over time or whether students focus their gaze on certain students at important conversational points. In addition, we encourage future research to explore integrating behavioral information from hand or head movements into corresponding analyses (see for the potential of head movements, e.g., Jun et al., 2020; Miller et al., 2021). Such studies might yield additional valuable insights into students' processing of (social) information in the IVR classroom during different phases of instruction.

6. Conclusion

The present study answers central questions about the configuration of IVR classrooms with a full class of virtual peer learners in educational research and practice. Overall, our results underline the potential of transforming traditional classrooms into immersive virtual reality scenarios for research purposes and effective learning scenarios. With regard to IVR configuration, the present study's findings indicate that the positioning of students in the IVR classroom, the visualization style of virtual avatars, as well as the performance-related behavior of virtual peer learners are decisive features to consider when configuring an IVR classroom. By examining students' gaze-based attention networks during instruction in an IVR classroom, we were able to gain valuable insights into the effects of different IVR classroom configurations on students' perception of the IVR classroom environment and processing of respective (social) information. Both educational researchers and practitioners are encouraged to carefully consider potential (side) effects of different IVR classroom configurations in light of their individual intentions and (research or learning) goals for using an IVR classroom.

References

Adams, R., Finn, P., Moes, E., Flannery, K., & Rizzo, A. S. (2009). Distractibility in attention/deficit/hyperactivity disorder (ADHD): The virtual reality classroom. *Child Neuropsychology, 15*(2), 120–135. https://doi.org/10.1080/09297040802169077

Alghamdi, N., & Alhalabi, W. (2019). Fixation detection with ray-casting in immersive virtual reality. *International Journal of Advanced Computer Science and Applications, 10*(7), 60–65. https://doi.org/10.14569/IJACSA.2019.0100710

Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology, 35*(11), 919–925. https://doi.org/10.1080/0144929x.2016.1212931

Bailenson, J., Aharoni, E., Beall, A. C., Guadagno, R. E., Dimov, A., & Blascovich, J. (2004). Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. In *Proceedings of the 7th Annual International Workshop on Presence* (pp. 1864–1105). International Society for Presence Research.

Bailenson, J., Swinth, K., Hoyt, C., Persky, S., Dimov, A., & Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence, 14*(4), 379–393. https://doi.org/10.1162/105474605774785235

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences, 17*(1), 102–141. https://doi.org/10.1080/10508400701793141

Bailey, J. O., & Bailenson, J. (2017). Immersive virtual reality and the developing child. In F. C. Blumberg & P. J. Brooks (Eds.), *Cognitive development in digital contexts* (pp. 181–200). Academic Press. https://doi.org/10.1016/b978-0-12-809481-5.00009-2

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Blume, F., Göllner, R., Moeller, K., Dresler, T., Ehlis, A.-C., & Gawrilow, C. (2019). Do students learn better when seated close to the teacher? A virtual classroom study considering individual levels of inattention and hyperactivity-impulsivity. *Learning and Instruction, 61*, 138–147. https://doi.org/10.1016/j.learninstruc.2018.10.004

Bönsch, A., Wendt, J., Overath, H., Gurerk, O., Harbring, C., Grund, C., Kittsteiner, T., & Kuhlen, T. W. (2017). Peers at work: Economic real-effort experiments in the presence of virtual co-workers. In *2017 IEEE Virtual Reality (VR)* (pp. 301–302). IEEE. https://doi.org/10.1109/VR.2017.7892296

Bozkir, E., Stark, P., Gao, H., Hasenbein, L., Hahn, J.-U., Kasneci, E., & Göllner, R. (2021). Exploiting object-of-interest information to understand attention in VR classrooms. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 597–605). IEEE. https://doi.org/10.1109/VR50410.2021.00085

Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology, 66*(2), 268–275. https://doi.org/10.1037/0022-3514.66.2.268

Bric, J. D., Lumbard, D. C., Frelich, M. J., & Gould, J. C. (2016). Current state of virtual reality simulation in robotic surgery training: A review. *Surgical Endoscopy, 30*(6), 2169–2178. https://doi.org/10.1007/s00464-015-4517-y

Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. Holt, Rinehart and Winston.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*(4), 523–547. https://doi.org/10.1037/0033-295x.97.4.523

Candeloro, L., Savini, L., & Conte, A. (2016). A new weighted degree centrality measure: The application in an animal disease epidemic. *PloS One*, *11*(11), Article e0165781. https://doi.org/10.1371/journal.pone.0165781

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Chang, F., Luo, M., Walton, G., Aguilar, L., & Bailenson, J. (2019). Stereotype threat in virtual learning environments: Effects of avatar gender and sexist behavior on women's math learning outcomes. *Cyberpsychology, Behavior, and Social Networking*, *22*(10), 634–640. https://doi.org/10.1089/cyber.2019.0106

Charitou, T., Bryan, K., & Lynn, D. J. (2016). Using biological networks to integrate, visualize and analyze genomics data. *Genetics Selection Evolution, 48*(27), 1–12. https://doi.org/10.1186/s12711-016-0205-1

Cheng, K.-H., & Tsai, C.-C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education*, *140*, Article 103600. https://doi.org/10.1016/j.compedu.2019.103600

Cheng, L., Farnham, S., & Stone, L. (2002). Lessons learned: Building and deploying shared virtual environments. In R. Schroeder (Ed.), *The Social Life of Avatars* (pp. 90-111). Springer. https://doi.org/10.1007/978-1-4471-0277-9_6

Cheryan, S., Meltzoff, A. N., & Kim, S. (2011). Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education, 57*(2), 1825–1835. https://doi.org/10.1016/j.compedu.2011.02.004

Chiesi, A. M. (2001). Network analysis. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 10499–10502). Pergamon. https://doi.org/10.1016/B0-08-043076-7/04211-X

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences: 2nd edition*. Lawrence Erlbaum. https://doi.org/10.4324/9780203771587

Curtin, K. M. (2018). Network analysis. In B. Huang (Ed.), *Comprehensive geographic information systems* (pp. 153–161). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09599-3

De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal, 29*(6), 841–859. https://doi.org/10.1080/0141192032000137330

Diestel, R. (2017). *Graph theory* (5th ed.). Springer. https://doi.org/10.1007/978-3-662-53622-3

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research, 78*(4), 828–879. https://doi.org/10.3102/0034654308321210

Erciyes, K. (2021). *Discrete mathematics and graph theory*. Springer. https://doi.org/10.1007/978-3-030-61115-6

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The Big-Fish-Little-Pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology, 9,* Article 1569. https://doi.org/10.3389/fpsyg.2018.01569

Fauville, G., Queiroz, A. C. M., Hambrick, L., Brown, B. A., & Bailenson, J. N. (2020). Participatory research on using virtual reality to teach ocean acidification: A study in the marine education community. *Environmental Education Research, 27*(2), 254–278. https://doi.org/10.1080/13504622.2020.1803797

Fernandes, A. C., Huang, J., & Rinaldo, V. (2011). Does where a student sits really matter? The impact of seating locations on student classroom learning. *International Journal of Applied Educational Studies, 10*(1), 66–77.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology, 21*(3), 95–113. https://doi.org/10.1027/1864-1105.21.3.95

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Perceived learning environment and students' emotional experiences: A multilevel analysis of mathematics classrooms. *Learning and Instruction, 5*, 478–493. https://doi.org/10.1016/j.learninstruc.2007.09.001

Fruehwirth, J. C. (2013). Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics, 4*(1), 85–124. https://doi.org/10.3982/QE93

Fryer, R. G. J., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics, 2*(2), 210–240. https://doi.org/10.1257/app

Gao, H., Bozkir, E., Hasenbein, L., Hahn, J.-U., Göllner, R., & Kasneci, E. (2021). Digital transformations of classrooms in virtual reality. In *CHI '21: Proceedings of the 2021 CHI*

Conference on Human Factors in Computing Systems (pp. 1–10). ACM. https://doi.org/10.1145/3411764.3445596

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 529–536). https://doi.org/10.1145/642611.642703

Göllner, R., Damian, R. I., Nagengast, B., Roberts, B. W., & Trautwein, U. (2018). It's not only who you are but who you are with: High school composition and individuals' attainment over the life course. *Psychological Science, 29*(11), 1785–1796. https://doi.org/10.1177/0956797618794454

Gottfried, M. A. (2012). Peer effects in urban schools: Assessing the impact of classroom composition on student achievement. *Educational Policy, 28*(5), 607–647. https://doi.org/10.1177/0895904812467082

Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the Field. *Educational Researcher, 42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

Guadagno, R. E., Blascovich, J., Bailenson, J., & McCall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology, 10*(1), 1–22. https://doi.org/10.108/15213260701300865

Guillon, Q., Afzali, M. H., Roge, B., Baduel, S., Kruck, J., & Hadjikhani, N. (2015). The importance of networking in autism gaze analysis. *PloS One, 10*(10), Article e0141191. https://doi.org/10.1371/journal.pone.0141191

Gürerk, Ö., Bönsch, A., Kittsteiner, T., & Staffeldt, A. (2019). Virtual humans as co-workers: A novel methodology to study peer effects. *Journal of Behavioral and Experimental Economics, 78*, 17–29. https://doi.org/10.1016/j.socec.2018.11.003

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference (SciPy 2008)* (pp. 11–15). SciPy.

Hardré, P. L., & Sullivan, D. W. (2008). Student differences and environment perceptions: How they contribute to student motivation in rural high schools. *Learning and Individual Differences, 18*(4), 471–485. https://doi.org/10.1016/j.lindif.2007.11.010

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15*(2), 177–199. https://doi.org/10.1076/sesi.15.2.177.30432

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research, 37*(5), 449–481. https://doi.org/10.1016/s0883-0355(03)00015-6

Heidicker, P., Langbehn, E., & Steinicke, F. (2017). Influence of avatar appearance on presence in social VR. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. https://doi.org/10.1109/3DUI.2017.7893357

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior, 26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Hochweber, J., Hosenfeld, I., & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal of Educational Psychology, 106*(1), 289–300. https://doi.org/10.1037/a0033829

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Howard, M. C., Gutworth, M. B., & Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior*, *121*, Article 106808. https://doi.org/https://doi.org/10.1016/j.chb.2021.106808

Hudson, I., & Hurter, J. (2016). Avatar types matter: Review of avatar literature for performance purposes. In S. Lackey & R. Shumaker (Eds.), *VAMR 2016: International conference on virtual, augmented and mixed reality* (pp. 14–21). Springer. https://doi.org/10.1007/978-3-319-39907-2_2

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology, 10*, Article 2246. https://doi.org/10.3389/fpsyg.2019.02246

Jarodzka, H., Homqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agenda. *Journal of Eye Movement Research, 10*(1), 1–18. https://doi.org/10.16910/jemr.10.1.3

Johnston, E., Olivas, G., Steele, P., Smith, C., & Bailey, L. (2017). Exploring pedagogical foundations of existing virtual reality educational applications: A content analysis study. *Journal of Educational Technology Systems, 46*(4), 414–439. https://doi.org/10.1177/0047239517745560

Jun, H., Miller, M. R., Herrera, F., Reeves, B., & Bailenson, J. (2020). Stimulus sampling with 360-videos: Examining head movements, arousal, presence, simulator sickness, and preference on a large sample of participants and videos. In *IEEE Transactions on Affective Computing*. https://doi.org/10.1109/TAFFC.2020.3004617

Kaakinen, J. K. (2021). What can eye movements tell us about visual perception processes in classroom contexts? Commentary on a special issue. *Educational Psychology Review, 33*, 169–179. https://doi.org/10.1007/s10648-020-09573-7

Karutz, C. O., & Bailenson, J. (2015). Immersive virtual environments and the classrooms of tomorrow. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 290–310). John Wiley & Sons. https://doi.org/10.1002/9781118426456.ch13

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. *Neuroscientist, 20*(5), 509–521. https://doi.org/10.1177/1073858413514136

Kübler, T. C., Rothe, C., Schiefer, U., Rosenstiel, W., & Kasneci, E. (2017). SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods, 49*(3), 1048–1064. https://doi.org/10.3758/s13428-016-0765-6

LaCroix, K., & LaCroix, S. (2017). Does seat location matter? A review of the proximity effect in large and small classrooms. *Community College Enterprise, 23*(2), 58–69.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90–115. https://doi.org/10.1016/j.edurev.2013.10.001

Lavy, V., Paserman, M. D., & Schlosser, A. (2011). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal, 122*, 208–237. https://doi.org/10.1111/j.1468-0297.2011.02463.x

Lee, J. E., Nass, C. I., & Bailenson, J. N. (2014). Does the mask govern the mind? Effects of arbitrary gender representation on quantitative task performance in avatar-represented virtual groups. *Cyberpsychology, Behavior, and Social Networking, 17*(4), 248–254. https://doi.org/10.1089/cyber.2013.0358

Levine, D. W., O'Neal, E. C., Garwood, S. G., & McDonald, P. J. (1980). Classroom ecology: The effects of seating position on grades and participation. *Personality and Social Psychology Bulletin, 6*(3), 409–412. https://doi.org/10.1177/014616728063012

Liao, M. Y., Sunq, C. Y., Wang, H. C., & Lin, W. C. (2019). Virtual classmates: Embodying historical learners' messages as learning companions in a VR classroom through comment mapping. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 163–171). IEEE. https://doi.org/10.1109/VR.2019.8797708

Lodge, J., & Harrison, W. (2019). The role of attention in learning in the digital age. *The Yale Journal of Biology and Medicine, 92*(1), 21–28.

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). Measuring telepresence: The temple presence inventory. In *12th Annual International Workshop on Presence* (pp. 1–15). International Society for Presence Research.

MacAulay, D. J. (1990). Classroom environment: A literature review. *Educational Psychology, 10*(3), 239–253. https://doi.org/10.1080/0144341900100305

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science* (pp. 25–29).

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology, 113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Makransky, G., Wismer, P., & Mayer, R. E. (2019). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning, 35*(3), 349–358. https://doi.org/10.1111/jcal.12335

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Marsh, H. W., Martin, A. J., Yeung, A. S., & Craven, R. (2017). Competence self-perceptions. In A. J. Elliot, C. Dweck, & D. Yeager (Eds.), *Handbook of competence and motivation* (2nd ed., pp. 85–115). Guilford Press.

Marsh, H. W., & Seaton, M. (2015). The Big-Fish-Little-Pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. *Advances in Motivation Science, 2*, 127–184. https://doi.org/10.1016/bs.adms.2015.05.002

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition, 146*, 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Meeks, M., Knotts, T., James, K., Williams, F., Vassar, J., & Wren, A. (2013). The impact of seating location and seating type on student performance. *Education Sciences, 3*(4), 375–386. https://doi.org/10.3390/educsci3040375

Miller, M. R., Jun, H., & Bailenson, J. (2021). Motion and meaning: Sample-level nonlinear analyses of virtual reality tracking data. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. https://doi.org/10.1109/54149.2021.00039

Montello, D. R. (1988). Classroom seating location and its effect on course achievement, participation, and attitudes. *Journal of Environmental Psychology, 8*(2), 149–157.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Moro, C., Stromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education, 10*(6), 549–559. https://doi.org/10.1002/ase.1696

Nolin, P., Stipanicic, A., Henry, M., Lachapelle, Y., Lussier-Desrochers, D., Rizzo, A. S., & Allain, P. (2016). ClinicaVR: Classroom-CPT: A virtual reality tool for assessing attention and inhibition in children and adolescents. *Computers in Human Behavior*, *59*, 327–333. https://doi.org/10.1016/j.chb.2016.02.023

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. *Journal of Personality and Social Psychology*, *117*(1), 166–185. https://doi.org/10.1037/pspp0000230

Pelham, B. W., & Wachsmuth, J. O. (1995). The waxing and waning of the social self: Assimilation and contrast in social comparison. *Journal of Personality and Social Psychology*, *69*(5), 825–838. https://doi.org/10.1037/0022-3514.69.5.825

Perkins, K. K., & Wieman, C. E. (2005). The surprising impact of seat location on student performance. *The Physics Teacher*, *43*(1), 30–33. https://doi.org/10.1119/1.1845987

Pietroszek, K. (2018). Raycasting in virtual reality. In *Encyclopedia of Computer Graphics and Games* (pp. 1–3). https://doi.org/10.1007/978-3-319-08234-9_180-1

Plieninger, H., & Dickhäuser, O. (2013). The female fish is more responsive: Gender moderates the BFLPE in the domain of science. *Educational Psychology*, *35*(2), 213–227. https://doi.org/10.1080/01443410.2013.814197

Preckel, F., Zeidner, M., Goetz, T., & Schleyer, E. J. (2008). Female ‘big fish’ swimming against the tide: The ‘big-fish-little-pond effect’ and gender-ratio in special gifted classes. *Contemporary Educational Psychology*, *33*(1), 78–96. https://doi.org/10.1016/j.cedpsych.2006.08.001

Queiroz, A. C. M., Nascimento, A. M., Tori, R., & Da Silva Leme, M. I. (2018). Using HMD-based immersive virtual environments in primary/K-12 education. In D. Beck, C. A. L. Morgado, J. Pirker, A. Peña-Rios, T. Ogl, J. Richter, & C. Gütl (Eds.), *Immersive Learning Research Network: 4th International Conference iLRN 2018* (pp. 160–173). Springer. https://doi.org/10.1007/978-3-319-93596-6

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ratan, R., Beyea, D., Li, B. J., & Graciano, L. (2020). Avatar characteristics induce users’ behavioral conformity with small-to-medium effect sizes: A meta-analysis of the proteus effect. *Media Psychology*, *23*(5), 651–675. https://doi.org/10.1080/15213269.2019.1623698

Rizzo, A. A., Bowerly, T., Buckwalter, J. G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2006). A virtual reality scenario for all seasons: The virtual classroom. *CNS Spectrum*, *11*(1), 35–44. https://doi.org/10.1017/S1092852900024196

Rizzo, A. A., Buckwalter, J. G., Bowerly, T., Van Der Zaag, C., Humphrey, L., Neumann, U., Chua, C., Kyriakakis, C., Van Rooyen, A., & Sisemore, D. (2000). The virtual classroom: A virtual reality environment for the assessment and rehabilitation of attention deficits. *Cyberpsychology & Behavior*, *3*(3), 483–499. https://doi.org/10.1089/10949310050078940

Sadria, M., Karimi, S., & Layton, A. T. (2019). Network centrality analysis of eye-gaze data in autism spectrum disorder. *Computers in Biology and Medicine, 111*, Article 103332. https://doi.org/10.1016/j.compbiomed.2019.103332

Schneider, B., Abu-El-Haija, S., Reesman, J., & Pea, R. (2013). Toward collaboration sensing: applying network analysis techniques to collaborative eye-tracking data. In *LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 107–111). ACM. https://doi.org/10.1145/2460296.2460317

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence, 10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica, 51*(4), 183–194. https://doi.org/10.1026/0012-1924.51.4.183

Schwebel, A. I., & Cherlin, D. L. (1972). Physical and social distancing in teacher-pupil relationships. *Journal of Educational Psychology, 63*(6), 543–550. https://doi.org/10.1037/h0034081

Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015). Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers. In *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3593–3602). ACM. https://doi.org/10.1145/2702123.2702415

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics, 104*(2), 147–200. https://doi.org/10.1007/s10649-020-09948-1

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 104–124). The International Association for the Study of Attention and Performance.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92*(1), 144–151. https://doi.org/10.1037//0022-0663.92.U44

Trautwein, U., Dumont, H., & Dicke, A.-L. (2015). Schooling: Impact on cognitive and motivational development. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol. 21, pp. 119–124). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26056-X

Wang, Y., Yuan, Y., Ma, Y., & Wang, G. (2019). Time-dependent graphs: Definitions, applications, and algorithms. *Data Science and Engineering, 4*(4), 352–366. https://doi.org/10.1007/s41019-019-00105-0

Wannarka, R., & Ruhl, K. (2008). Seating arrangements that promote positive academic and behavioural outcomes: A review of empirical research. *Support for Learning, 23*(2), 89–93. https://doi.org/10.1111/j.1467-9604.2008.00375.x

Webster, R. (2015). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments, 24*(6), 1319–1333. https://doi.org/10.1080/10494820.2014.994533

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Will, P., Bischof, W. F., & Kingstone, A. (2020). The impact of classroom seating location and computer use on student academic performance. *PloS One*, *15*(8), Article e0236131. https://doi.org/10.1371/journal.pone.0236131

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 143–146). ACM. https://doi.org/10.1145/1978942.1978963

Yazdan-Shahmorad, P., Sammaknejad, N., & Bakouie, F. (2020). Graph-based analysis of visual scanning patterns: A developmental study on green and normal images. *Nature Scientific Reports*, *10*(1), Article 7791. https://doi.org/10.1038/s41598-020-63951-3

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1153–1162). ACM. https://doi.org/10.1145/1124772.1124945

Appendix A

Details on the Pre-processing of Eye-Tracking Data and the Creation of Graphs

We analyzed attentional processes by performing network analysis on gaze shift movements to trace the path of visual attention throughout the virtual space. To access information about where participants looked during the IVR lesson, we used a technique known as gaze ray-casting (Alghamdi & Alhalabi, 2019; Pietroszek, 2018). Gaze ray-casting combines information from each frame about the participant's head location, head orientation, and gaze direction to calculate the gaze direction in the virtual environment and pinpoint the exact location the participant is looking at. One could imagine a gaze ray-cast as a laser beam pointing from the participant's (combined) eye location into the virtual space and hitting a specific physical object there. By identifying hits of the gaze ray with a virtual object at every split second, we were able to continuously track which object in the IVR classroom participants observed (Alghamdi & Alhalabi, 2019). We used the Python programming language (Python Software Foundation, https://www.python.org/) to process the ray-casting information calculated during the experimental session. Since not every virtual object in the environment was of interest for our study, we only included specific objects of interest (OOIs), namely the virtual peer learners, the virtual teacher, and the screen on which the instructional content was presented. The gaze shift movement from one object to another was then identified as a transition from a specific object to another if the transition duration was no longer than 10 seconds. The transitions between OOIs across the entire experiment session were then summed up and stored in an adjacency matrix (i.e., a $n \times n$ matrix A, with n being the number of OOIs). Consequently, each cell in the matrix a_{ij} stated the number of gaze shifts transitioning from OOI i to OOI j, resulting in a transition matrix for each participant encompassing the number of gaze shifts for the full experiment session (similar to transitions matrices for scanpath analyses; Kübler et al., 2017).

According to graph theory, each adjacency matrix can be considered as a weighted directed cyclic graph (Erciyes, 2021). Building a graph from the normalized transition matrix involves treating the OOIs as nodes of the graph and the normalized number of transitions as the edge weights, with an edge created between nodes if at least two transitions between them occurred during the full experiment. We built the graphs from the adjacency matrices using the NetworkX package (Hagberg et al., 2008). Based on a graph for each participant, we were able to calculate different variables describing the graph structure and graph properties (see overview in Table A.1). We provide access to all data and data analysis scripts including the data pre-processing steps on the Open Science Framework (OSF) under the following link: https://osf.io/pek4q/?view_only=ef151fd06ac8413a827020d4264b3c8d.

Taken together, we argue that analyzing gaze shift movements as gaze-based attention networks has a number of advantages: First, the existing large body of research on graph theory and network analysis provides a rich set of graph properties and variables to describe graph structures, which can be computed from the given gaze-based attention networks. Second, these structural variables are already presented as aggregated variables, which allows for statistical testing and significance analysis without the need for machine learning or complex non-linear regression models. Third and finally, the calculated structural variables can be interpreted on a theoretical level because they directly reflect characteristics of gaze-based visual attention. Notably, the gaze-based graphs also include aggregated information about head movements; because participants' field of view did not capture all OOIs in the classroom simultaneously, even when participations were positioned in the back of the IVR classroom, the gaze-based graphs include information about the extent to which participants moved their heads to the left and right to change their field of view (e.g., high gaze activity among peers is only possible with head movements).

Table A.1

Overview of Structural Variables Used to Examine Students' Gaze-Based Attention Networks in the IVR Classroom

Category	Feature Example / Application in IVR classroom	Technical description	Meaning
Gaze centrality	Degree centrality (DC) / DC peers DC teacher DC screen	Sum of incoming and outgoing edges, based on directed weighted graph edges	Indicates to what extent an OOI is the center/focus of gaze transition
Connectedness of gazes	Cliques / N cliques among peers Avg size of cliques among peers Proportion of boys in observed cliques	Subset of nodes that contains the maximal number of nodes that share an edge with every other node in the subset, based on undirected edge weights	Indicates how connected gazes between OOIs are, i.e., to what extent gaze transitions between OOIs occur
Uniformity of gaze distribution	Weighted degree centrality (WDC) / WDC screen	Sum of incoming and outgoing edges, including uniformity of edge weights	Indicates how uniformly gaze is distributed between the screen and different peer learners
	Cut size / Cut size teacher/screen – peers	Sum of edge weights that pass between the two predefined subsets	Indicates the degree to which attention goes back and forth between the instructional content (i.e., teacher/screen) and peer learners
	Overall uniformity	Chi-square test statistic for a sample containing all edge weights of a person's gaze shifts	Indicates the overall uniformity of gaze distribution

Note. OOI = Object of interest (i.e., in our case virtual peer learners, the virtual teacher or the screen); N = Number; Avg = Average.

Supplemental Material

The supplemental material includes details on the present study's sample compared to the full sample (Table S1), detailed statistics of the aligned rank transformation ANOVAs (Tables S2-6) as well as details on the correlations between covariates and students' learning experience in the IVR (Table S7).

Table of Contents

Table S1

Central Sample Characteristics and Distribution Across Experimental Groups in the Full Compared to the Present Study Sample

	Full Sample (*N* = 381)	Study Sample (*N* = 274)
Data structure		
Schools / classes	11 / 25	11 / 25
Students per class	13.24 (4.19)	9.64 (3.06)
Age	11.51 (0.56)	11.50 (0.55)
Gender		
Female / Male	47% / 53%	50.4% / 49.6%
Grades [a]		
Math	2.63 (0.88)	2.61 (0.89)
German	2.49 (0.71)	2.48 (0.72)
Prior IVR experience [b]		
No / Yes / n/a	39.9% / 58.8% / 1.3%	41.6% / 56.9% / 1.5%
General self-concept intelligence [c]	3.05 (0.60)	3.07 (0.61)
Initial CT interest [c]	3.14 (0.77)	3.14 (0.76)
Distribution across experimental conditions		
Front / back sitting position	42.8% / 57.2%	44.5% / 55.5%
Cartoonish / stylized avatar visualization	52.2% / 47.5%	54.7% / 45.3%
20% / 35% / 65% / 80% hand-raising	26.0% / 24.7% / 23.9% / 25.5%	26.3% / 23.4% / 21.9% / 28.5%

Note. Mean values and standard deviations M (SD) are shown for continuous variables. Number of schools and classes are shown in absolute numbers.

CT = Computational Thinking; IVR = Immersive Virtual Reality.

[a] Grades were on a scale from 1–6 with lower numbers indicating better achievement; [b] Prior IVR experience was assessed via one item asking whether the participant had previously used IVR glasses; [c] Measured on a 4-point rating scale with higher values indicating higher levels of the respective variable.

Table S2

Results of Aligned Rank Transformation ANOVA for Gaze Centrality Markers

	Degree centrality peers				Degree centrality teacher				Degree centrality screen			
	df	*F*	*p*	partial η^2	*df*	*F*	*p*	partial η^2	*df*	*F*	*p*	partial η^2
Sitting position	1, 258	138.55	< .001	.35	1, 258	204.07	< .001	.44	1, 258	60.56	< .001	.19
Visualization style	1, 258	20.46	< .001	.07	1, 258	52.03	< .001	.17	1, 258	3.36	.068	.01
Hand-raising	3, 258	7.76	< .001	.08	3, 258	8.09	< .001	.09	3, 258	4.94	.002	.05
Position * Visualization style	1, 258	11.87	< .001	.04	1, 258	19.77	< .001	.07	1, 258	3.81	.052	.01
Position * Hand-raising	3, 258	0.37	.774	.00	3, 258	0.64	.592	.01	3, 258	0.31	.816	.00
Visualization style * Hand-raising	3, 258	1.58	.195	.02	3, 258	0.74	.527	.01	3, 258	1.63	.182	.02
Position * Visualization style * Hand-raising	3, 258	0.15	.931	.00	3, 258	0.10	.961	.00	3, 258	0.31	.817	.00

Note. ANOVA = analysis of variance. df denotes the degrees of freedom (df) and df residuals. Sitting position with two levels (front vs. back). Visualization style with two levels (cartoonish vs. stylized). Hand-raising with four levels (20% vs. 35% vs. 65% vs. 80%).

Table S3

Tukey's HSD Post-hoc Test for Differences in Gaze Centrality Markers Between Hand-Raising Conditions

Hand-raising conditions	Degree centrality peers				Degree centrality teacher				Degree centrality screen			
	Estimate	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*	*Estimate*	*SE*	*t*	*p*
20% - 35%	18.90	13.68	1.38	.512	-20.39	13.66	-1.49	.443	-16.23	13.89	-1.17	.647
20% - 65%	47.29	13.83	3.42	.004	-50.75	13.81	-3.67	.002	-37.66	14.04	-2.68	.039
20% - 80%	-16.21	13.41	-1.21	. 622	13.16	13.39	0.98	.759	13.59	13.61	1.00	.751
35% - 65%	28.39	14.07	2.02	.184	-30.36	14.05	-2.16	.137	-21.42	14.28	-1.50	.439
35% - 80%	-35.11	13.65	-2.57	.052	33.54	13.62	2.46	.068	29.82	13.86	2.15	.140
65% - 80%	-63.50	13.80	-4.60	< .001	63.91	13.78	4.64	< .001	51.24	14.01	3.66	.002

Note. Based on aligned rank transformation analysis of variance (ANOVA), the estimates and respective standard errors (*SE*) are on the scale of the ranks and not the data.

Table S4

Results of Aligned Rank Transformation ANOVA for Markers of Gaze Connectedness Among Peers

	Number of cliques among peers				Average clique size among peers				Proportion boys in observed cliques			
	df	*F*	*p*	partial η^2	*df*	*F*	*p*	partial η^2	*df*	*F*	*p*	partial η^2
Sitting position	1, 258	4.50	.035	.02	1, 258	8.06	.005	.03	1, 258	101.35	< .001	.28
Visualization style	1, 258	9.81	.002	.04	1, 258	12.47	< .001	.05	1, 258	0.17	.677	.00
Hand-raising	3, 258	1.84	.140	.02	3, 258	1.59	.192	.02	3, 258	1.14	.334	.01
Position * Visualization style	1, 258	0.05	.827	.00	1, 258	1.45	.230	.01	1, 258	1.50	.221	.01
Position * Hand-raising	3, 258	0.47	.706	.01	3, 258	0.76	.518	.01	3, 258	1.20	.309	.01
Visualization style * Hand-raising	3, 258	1.46	.224	.02	3, 258	1.78	.152	.02	3, 258	0.85	.470	.01
Position * Visualization style * Hand-raising	3, 258	0.06	.982	.00	3, 258	0.47	.707	.01	3, 258	1.00	.394	.01

Note. ANOVA = analysis of variance. df denotes the degrees of freedom (df) and df residuals. Sitting position with two levels (front vs. back). Visualization style with two levels (cartoonish vs. stylized). Hand-raising with four levels (20% vs. 35% vs. 65% vs. 80%).

Table S5

Results of Aligned Rank Transformation ANOVA for Gaze Uniformity Markers

	Weighted DC screen				Cut size teacher/screen – peers				Uniformity (overall)			
	df	*F*	*p*	partial η^2	*df*	*F*	*p*	partial η^2	*df*	*F*	*p*	partial η^2
Sitting position	1, 258	333.16	< .001	.56	1, 258	109.23	< .001	.30	1, 258	33.11	< .001	.11
Visualization style	1, 258	34.82	< .001	.12	1, 258	14.30	< .001	.05	1, 258	0.62	.430	.00
Hand-raising	3, 258	12.92	< .001	.13	3, 258	1.81	.145	.02	3, 258	1.07	.361	.01
Position * Visualization style	1, 258	6.08	.014	.02	1, 258	12.50	< .001	.05	1, 258	21.44	< .001	.08
Position * Hand-raising	3, 258	2.31	.077	.03	3, 258	0.39	.758	.01	3, 258	1.39	.247	.02
Visualization style * Hand-raising	3, 258	0.36	.785	.00	3, 258	0.48	.695	.01	3, 258	1.00	.394	.01
Position * Visualization style * Hand-raising	3, 258	0.68	.567	.01	3, 258	0.02	.996	.00	3, 258	1.70	.167	.02

Note. ANOVA = analysis of variance. df denotes the degrees of freedom (df) and df residuals. Sitting position with two levels (front vs. back). Visualization style with two levels (cartoonish vs. stylized). Hand-raising with four levels (20% vs. 35% vs. 65% vs. 80%).

Table S6

Tukey's HSD Post-hoc Test for Differences in Gaze Uniformity Markers Between Hand-Raising Conditions

Hand-raising conditions	Weighted DC screen			
	Estimate	*SE*	*t*	*p*
20% - 35%	26.56	13.30	2.00	.192
20% - 65%	64.82	13.45	4.82	< .001
20% - 80%	-12.78	13.03	-0.98	.760
35% - 65%	38.26	13.67	2.80	.028
35% - 80%	-39.35	13.27	-2.97	.017
65% - 80%	-77.61	13.42	-5.78	< .001

Note. Based on aligned rank transformation analysis of variance (ANOVA), the estimates and respective standard errors (*SE*) are on the scale of the ranks and not the data. DC = Degree Centrality.

Table S7

Correlation Matrix of Covariates and Interest in the IVR Lesson, Situational Self-Concept, and Posttest Score

Variables	1	2	3	4	5
1. Gender [a]	—				
2. Self-Concept Intelligence [b]	.05 [-0.07, 0.17]	—			
3. Initial CT Interest [b]	.35*** [0.24, 0.46]	.03 [-0.09, 0.15]	—		
4. Interest in IVR lesson [b]	.23*** [0.11, 0.34]	.10 [-0.02, 0.22]	.71*** [0.63, 0.77]	—	
5. Situational Self-Concept [b]	.13* [0.01, 0.25]	.13* [0.01, 0.24]	.21*** [0.09, 0.32]	.22*** [0.10, 0.33]	—
6. Posttest Score [c]	-.17** [-0.29, -0.05]	.25*** [0.13, 0.36]	-.06 [-0.18, 0.06]	.06 [-0.06, 0.18]	.08 [-0.04, 0.20]

Note. CT = Computational thinking (lesson topic).
[a] Gender female = 1, male = 2; [b] Continuous variables were measured on a 4-point rating scale with higher values indicating higher levels of the respective variable; [c] Posttest scores ranged from 0 to 12 points. Variables 2-6 are non-normally distributed; thus, Spearman’s rho is reported. 95% confidence intervals are given in brackets.
*** $p < .001$. ** $p < .01$. * $p < .05$

6

GENERAL DISCUSSION

6 General Discussion

Social comparisons are commonplace in every classroom and widely acknowledged as central determinants of students' academic self-evaluations (see, e.g., Dijkstra et al., 2008; Trautwein & Möller, 2016). BFLPE research has repeatedly shown the importance of the social environment of a classroom in shaping students' academic-self-concept (see, e.g., Marsh et al., 2017; Marsh & Seaton, 2015). However, whereas existing research has provided compelling evidence of the effects of social comparisons on students' self-evaluations, and a considerable number of experimental studies have provided general insights into the mechanisms behind social comparisons, the actual processes involved in social comparisons in the classroom are still quite a black box. The present dissertation was aimed at gaining insights into the underlying mechanisms (i.e., the "inner workings" of this black box) by using immersive virtual reality (IVR) as a research tool.

To this end, I proposed a theoretical model that can be applied to examine social comparisons and the corresponding processing of social information in the classroom (see Figure 5, Chapter 1.2.3). The model is aimed at describing the black box of social comparisons in the classroom through (a) *intra*personal covert cognitive responses to social comparison information and (b) *inter*personal overt behavioral responses to the social environment. Aiming to develop a theoretically advanced understanding of social comparisons in the classroom, Studies 1 and 2 examined these different aspects of social comparison behavior, respectively. Study 3 focused on the overall methodological approach of using an immersive virtual reality classroom to authentically simulate and control the social environment and social comparison information to gain corresponding insights into the black box of social comparisons in the classroom.

In the following chapter, I first summarize the findings of the three empirical studies and discuss them with regard to the present dissertation's aims and research questions as well as to a broader research context (Chapter 6.1). Following this, I outline the dissertation's strengths and limitations (Chapter 6.2) and discuss implications for future research and practice (Chapter 6.3) before closing with a general conclusion (Chapter 6.4).

6.1 Discussion of the Results

In line with the present dissertation's two subordinate objectives, I will use the following chapter first to discuss theoretical contributions with regard to insights into the black box of social comparison processes in the classroom (Chapter 6.1.1) and second to discuss methodological contributions with respect to the use of immersive virtual reality as an experimental tool (Chapter 6.1.2). Afterwards, I will integrate the two perspectives to reflect on the extent to which the present dissertation provides insights into the overarching research aim of revisiting social comparison processes in the classroom using an IVR environment as an experimental tool (Chapter 6.1.3).

6.1.1 Theoretical Contributions: Insights into Social Comparison Processes in the Classroom

In educational psychology research, it is nearly impossible to read about social comparisons in the classroom without coming across the BFLPE. The negative reference group effect, while controlling for individual achievement, is one of the most researched phenomena in the field. And yet, despite the large body of supporting evidence (see, e.g., reviews by Fang et al., 2018; Marsh et al., 2017; Marsh & Seaton, 2015), BFLPE research has faced two major limitations (see Chapter 1.2.1): Such studies have mostly relied on cross-sectional data from large-scale surveys and corresponding correlational analyses; hence, experimental evidence for the BFLPE—especially in authentic classroom settings—is largely missing (see Zell & Alicke, 2009a; Zell & Alicke, 2010, for an exception). Moreover, as most prominently highlighted by Dai and Rinn (2008), BFLPE research has typically assumed but never directly examined or measured social comparisons as the underlying mechanism, leading to the respective pattern of results (see the few exceptions by Huguet et al., 2009; Marsh et al., 2014; Marsh, Trautwein, et al., 2008). Whereas social psychology research has provided a large body of experimental studies on social comparisons that have provided insights into the underlying mechanisms and the processing of social information, none of these have been situated in a real classroom situation with naturally occurring social comparison behavior among students (see Chapter 1.2.2 for an overview).

With regard to theoretical contributions to research on social comparisons in the classroom, and specifically the BFLPE, the studies included in this dissertation addressed the lack of experimental evidence for the BFLPE in an authentic classroom setting: Students were randomly assigned to an IVR classroom with a specific proportion of students who showed high-achieving behavior (i.e., more hand-raising), and—in line with the BFLPE—higher proportions of high-achieving classmates led to lower situational self-concept (Study 1). Moreover, the present studies addressed what Dai and Rinn (2008) pointed out in their critical review of BFLPE research when they criticized that the explanation and interpretation of the respective findings on social comparisons is “based on blanket assumptions rather than direct evidence” (p. 297). What the authors criticized as “blanked assumptions” is what the present dissertation described as the black box of social comparisons in the classroom and aimed to uncover by explicitly examining the processing of social information in the classroom. More specifically, the present studies provided insights into the explanation of the BFLPE through insights into the black box of social comparison on the basis of students’ cognitive (i.e., intrapersonal) and behavioral (i.e., interpersonal) responses to social comparison information.

First, with respect to *intra*personal behavior associated with social comparisons, the results showed that students’ cognitive responses to the social comparison information that was provided was critical for explaining the BFLPE. Study 1 demonstrated that students’ interpretation of their peer learners’ performance-related behavior as an indicator of performance fully explained the BFLPE that was found on the basis of the experimental manipulation of the class’ performance level. Students were asked (a) how many hand-raising classmates they remembered and (b) how they perceived the class’ overall performance level. Importantly, these two variables were only moderately correlated, and only the perceived performance level of the class (i.e., students’ interpretation of what they observed) was found to explain the BFLPE on students’ situational self-concept, but the number of hand-raising classmates they recalled was not. Particularly against the background that social comparison information was only implicitly provided via the classmates’ performance-related behavior (i.e., hand-raising), this finding highlights the role of subjective interpretations of certain situational cues as self-relevant social comparison information. The findings are in line with self-concept conceptualizations (see Chapters 1.1.2 and 1.1.3) that highlight subjective experiences of success or failure in a specific situation as major determinants of students’ self-concepts. In other words, the results suggest that students who perceived the class’ overall level of performance as lower were more likely to experience their own performance as a success relative to their classmates and hence showed higher levels of situational self-concept afterwards. Compared with existing attempts to integrate measures of students’ intrapersonal social comparison behavior into BFLPE research designs (i.e., the achievement of individually selected classmates or self-reports of a student’s perceived relative standing in class; Huguet et al., 2009; Marsh et al., 2014; Marsh, Trautwein, et al., 2008), the present dissertation worked with a standardized setting that at the same time allowed social comparisons to occur naturally.

In this vein, the insights into the role of intrapersonal social comparison behavior in Study 1 also relied on students' self-reports but (a) from an experimentally controlled environment and (b) without forcing or limiting the comparison information that participants might consider by asking about specific comparison targets.

Second, with regard to *inter*personal processes, the results showed that students' behavioral responses to the social comparison information that was provided was critical for explaining the BFLPE. Similar to some physiological response studies that have used visual attention patterns to experimentally examine social comparisons (see Chapter 1.2.2; e.g., Bauer, Schneider, Waldorf, Braks, et al., 2017; Michinov et al., 2015), Studies 2 and 3 used students' eye-tracking data and its affordances to describe overt social comparison behavior (see Chapter 1.3.3 and a respective discussion of the measurements in the IVR classroom environment in the Discussion in Chapter 6.1.2). Results showed that general measures of visual attention to social comparison information (i.e., number of peer learners looked at and the frequency and total duration of gazes at peers; Study 2) and particularly visual attention to more specific social information (i.e., the proportion of boys in observed peer learners; Study 3) were related to individual differences in situational self-concept: The more students paid attention to their classmates (and particularly to male classmates), the lower their situational self-concept. Compared with existing research, using behavioral measures of social comparisons from an authentic classroom setting with naturally occurring interpersonal behavior is completely novel for research on social comparisons and even more so for research on the BFLPE. Not only do the respective results allow researchers to describe interpersonal social comparison behavior objectively with standardized data, but they provide insights into the extent to which students actively engage in social comparisons in the classroom without being asked to do so. To this end, it is important to note that students' overt social comparison behavior indicated that they attended and responded to their classmates' performance-related behavior, particularly when it could be clearly interpreted; visual attention to virtual classmates (Study 2) as well as gaze-based attention networks in the IVR classroom (Study 3) have shown that students attended more to their virtual classmates when a clear minority or majority of them (i.e., 20% or 80%) raised their hands. This finding is particularly important as it highlights that students' behavioral responses to their peer learners go beyond an "action-reaction relationship," by which attention would increase after the activity of the peers increased (as would be expected if visual attention to peers was purely a bottom-up attentional process reflecting a spontaneous reaction to situational cues). In fact, the current finding indicates that visual attention to peers—at least to a substantial degree—reflects some more willfully directed visual attention oriented toward social information (related to top-down attentional processes; Richardson & Gobel, 2015; Theeuwes et al., 2000).

In sum, the studies in this dissertation were able to identify both (a) intrapersonal cognitive responses and (b) interpersonal behavioral responses to social comparison information that explain the BFLPE. I would like to highlight that indicators of both covert (Study 1) and overt (Studies 2 and 3) social comparison behavior showed that some behaviors—as would be intended and expected on the basis of the experimental manipulation—were affected by peers' performance-related behavior in the IVR classroom situation; however, only some of these showed a relationship to students' self-concept, whereas others seemed to primarily represent a reaction to situational cues that was not directly linked to self-evaluations (i.e., the number of hand-raising peers recalled in Study 1 or the mean pupil diameter in Study 2). In turn, other behavioral markers that were not affected by the experimentally manipulated social comparison information (e.g., the frequency with which and the total time that students looked at their peers in Study 2 or the proportion of boys in observed groups of peers in Study 3) showed significant relationships to individual differences in students' situational self-concepts. In line with conceptualizations of self-concept from social psychology (see, e.g., Markus & Wurf, 1987) and the theoretical model that was proposed for examining social comparisons in the classroom (see Figure 5, Chapter 1.2.3), this pattern of results points to the importance of individual characteristics when examining social comparison behavior—particularly the extent to which students actively process social comparison information—and the resulting individual differences in self-concept. Whereas the influence of individual characteristics on social comparisons and the corresponding processing of social comparison information in the classroom was not the focus of the studies from this dissertation, such variables were included as covariates in the studies' analyses. Notably, the covariates that were included did not have an effect on the studies' results but showed a consistent pattern across the different studies,[23] which points to the role of individual differences in social comparison behaviors and the corresponding processing of social comparison information in the classroom that needs to be considered when looking for explanations for the BFLPE (see the implications for future research in Chapter 6.3.1).

[23] On the one hand, grades (as a proxy for domain-specific achievement) were related to measures of situational self-concept, indicating that students incorporated their existing performance-related experiences into their situational self-evaluations (see the dimensional comparison effect in Studies 1 and 2). On the other hand, students' gender and general self-concept of intelligence as well as their interest in the lesson were related to their overt social comparison behavior (i.e., their visual attention to peers; see Studies 2 and 3, respectively).

6.1.2 Methodological Contributions: The Use of an IVR Classroom as an Experimental Tool

The present dissertation aimed to use an IVR classroom as a research tool to get the best of both worlds—experimental control from lab settings and authenticity from real-world research—in order to gain insights into social comparisons in the classroom (see Figure 10 for an illustration).

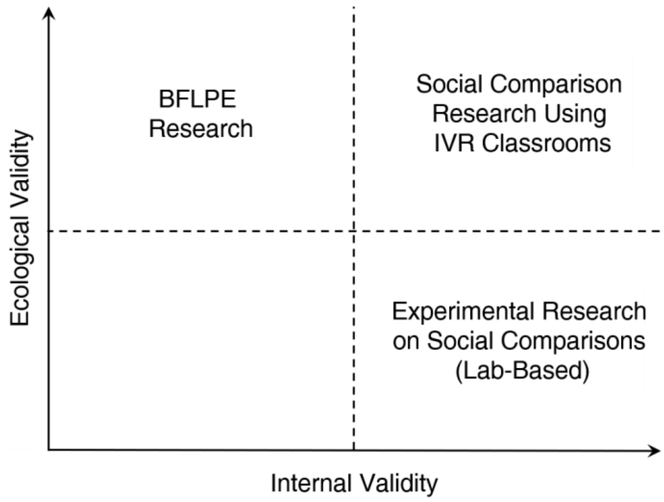


Figure 10. The potential of IVR classrooms as an experimental tool.

Whereas the potential of IVR as an experimental tool for educational and social psychology research is evident (see, e.g., Blascovich et al., 2002; Fox et al., 2009), there are not many studies that have tapped this potential, specifically not to simulate classroom environments. I therefore identified three critical aspects that need to be considered for the successful use of IVR as an experimental tool for classroom research, namely, (a) the authenticity of immersive virtual simulations of classroom realities (see Chapter 1.3.1), (b) decisions about experimental control when manipulating an IVR classroom (see Chapter 1.3.2), and (c) the potential of standardized process data from IVR classroom environments (see Chapter 1.3.3). In the following, I discuss how the present dissertation contributes to understanding how an IVR classroom can be used as an experimental tool with respect to these three aspects.

As outlined in Chapter 1.3.1, state-of-the-art IVR environments are expected to lead to a highly authentic experience in the IVR environment, characterized by full immersion and the experience of a high degree of presence and perceived realism (e.g., Sanchez-Vives & Slater, 2005). Students' self-reports of their IVR experiences in the present studies indicated that this was also the case for the IVR classroom simulation (see Studies 1 to 3). Most importantly, however, as a proof of concept for the use of an IVR classroom as an experimental tool for authentic classroom research is the fact that the empirically well-established BFLPE could be reproduced in the IVR setting (Study 1). More specifically, students recognized their virtual classmates' performance-related (i.e., hand-raising) behavior as an indicator of performance, and their responses to it were in line with the BFLPE (Study 1). This provides support for the authenticity of the IVR classroom simulation, especially against the background that hand-raising in real-world classrooms has also been found to be associated with high-performing characteristics of students, such as behavioral engagement, achievement, and self-concept (Böheim, Knogler, et al., 2020; Böheim, Urdan, et al., 2020).

With regard to the configuration of IVR classrooms and the corresponding experimental manipulations, the studies in this dissertation provided additional important insights. Not only was virtual classmates' hand-raising behavior perceived as an indicator of performance, but students actively attended to the social information provided by their virtual peer learners (Studies 2 and 3). In fact, students attended and responded to their classmates' performance-related behavior the most when it was unambivalent, as indicated by their visual attention and pupil diameters (Study 2) as well as their gaze-based attention networks (Study 3). From a methodological perspective, this highlights the importance of salience when manipulating certain information in the IVR classroom. Compared with traditional experimental studies that have manipulated social comparison information via fictional scenarios (see, e.g., Mussweiler et al., 2004; Mussweiler & Strack, 2000), the experimental design in the present studies presented students with much more freedom to explore. Whereas such freedom is desirable in terms of authenticity and is much closer to a real-world classroom experience with naturally occurring social comparison behavior, the results indicate that the experimental manipulations need to be as unambiguous as possible in order to have an effect (see the effect on situational self-concept; Study 1).[24]

Moreover, the results of the present studies highlighted some features related to the configuration of the IVR classroom environment that need to be considered when using it as an experimental tool: As Studies 2 and 3 demonstrated, students' natural field of view in the IVR classroom (from a front or back row) as well as the style in which the virtual avatars were represented affected how students distributed their attention in the IVR classroom. In order to

[24] It is important to consider that the results of the present studies are based on a 15-min IVR classroom simulation; the length of the IVR classroom simulation (i.e., the time of "exposure" to certain information) might make the salience of the respective manipulations more or less critical (see the corresponding implications for future research in Chapter 6.3.1).

examine social comparisons in the classroom, attention to virtual peer learners was considered a particularly critical outcome, which occurred more when students had the whole class in their field of view (i.e., from a position in the back of the classroom) and when the virtual avatars were represented in more of a cartoon-styled manner (Studies 2 and 3). In turn, students' visual attention was more focused on the contents of the lesson when they sat in the front where they were closer to the teacher and where the lesson materials were being presented (Study 3). On the one hand, these findings demonstrate how students' gaze data from an IVR classroom—as an indicator of their actual processing of information in the IVR classroom—can be used to extend existing research that has typically examined the effects of different IVR classroom configurations on student learning via test scores and self-reports (e.g., Bailenson et al., 2008; Blume et al., 2019; Makransky, Wismer, et al., 2019). On a more general level, with regard to the use of IVR classrooms as an experimental tool, these findings suggest that the respective configuration and design features of the IVR classroom are decisive factors with regard to what students' pay attention to and should be chosen carefully with the respective research objective in mind.

Finally, the present studies provided evidence for the potential of IVR technology to collect standardized process data (see Chapter 1.3.3). Regarding students' self-reports of their interpretations of the respective social comparison information (Study 1) and their situational self-concepts (Studies 1 to 3), the IVR setting in the present studies made it possible to overcome the existing shortcoming of experimental research on social comparisons, which is that the respective measures have relied on "detached classroom and experimental situations" (Demo, 1992, p. 304). Beyond students' self-reports of their IVR experiences and their interpretations of the information they encountered in the IVR classroom, the present studies analyzed students' visual attention as an indicator of their (social) information processing in the IVR classroom (Studies 2 and 3). In this vein, not only did the studies provide a more naturalistic and yet standardized experimental setting for the measures associated with social comparisons (see a call for respective designs, e.g., by Collins, 2000), but they also objectively traced students' overt behavioral responses to social comparison information in the classroom. More specifically, the studies used measures of students' visual attention to their peers (Study 2) and their gaze-based networks (i.e., transitions between objects of interest; Study 3) to examine students' processing of different types of information in the classroom. Whereas visual attention to objects of interest is a rather well-established way of analyzing gaze data (e.g., Kaakinen, 2021), the analysis of visual attention based on a gaze-based network is a comparably new approach (see, e.g., Sadria et al., 2019; Yazdan-Shahmorad et al., 2020). Notably, with regard to analyzing how students distribute their attention in an IVR classroom setting, both ways of analyzing visual attention present relatively new approaches. Against the background that the observed markers of visual attention were related to students' self-concept (Study 2) and additionally their interest in the IVR lesson and performance (Study 3), the studies in this dissertation make a substantial contribution to existing research on the use of these approaches to gain insights into how students process (social) information, particularly in an IVR classroom.

Taken together, the present studies provide a proof of concept for the use of IVR classrooms as an experimental tool in educational and social psychology research, specifically to investigate social comparisons in the classroom (as suggested in Figure 10). Limitations in BFLPE research and the lack of experimental evidence (see Chapter 1.2.1 and the respective discussion in Chapter 7.1.1) have primarily been due to natural restrictions in field research where it is not feasible (or ethically possible) to manipulate the composition of a classroom as in an ideal experimental design to examine the BFLPE. The results of the present studies indicate that it is possible to overcome these limitations by using an IVR classroom as an experimental tool that (a) provides an authentic and yet fully controllable research setting and (b) allows insights into underlying mechanisms via standardized process data such as visual attention to social comparison information in the classroom.

6.1.3 Revisiting Social Comparisons in an IVR Classroom: An Interim Conclusion

As outlined in the previous chapters, the studies in this dissertation have provided insights into intra- and interpersonal processes of social comparisons in the classroom (see Chapter 6.1.1) and how an IVR classroom can be used as an experimental tool to gain respective insights (see Chapter 6.1.2). Whereas I discussed the findings of each study separately regarding their contributions to each of the subordinate research objectives, it is important to note that the theoretical and methodological advancements in the present dissertation are naturally intertwined. On the one hand, the theoretical contributions to research on social comparisons in the classroom are based on the methodological approach of using IVR as an experimental tool; on the other hand, the insights into the methodological affordances of IVR are based on the specific application to research on social comparisons in the classroom.

On the basis of the discussion of the dissertation's two subordinate objectives (theoretical contributions in Chapter 6.1.1 and methodological contributions in Chapter 6.1.2), I will use the following chapter to draw an interim conclusion about the overarching aim of the dissertation to revisit social comparisons in the classroom and gain insights into the underlying mechanisms from the use of IVR as an experimental tool. I will therefore circle back to the proposed theoretical model to examine social comparisons in the classroom (see Figure 5, Chapter 1.2.3) and evaluate how the present studies contribute to the theoretical foundations and methodological implementation of the model from using an IVR environment to examine social comparisons in the classroom.

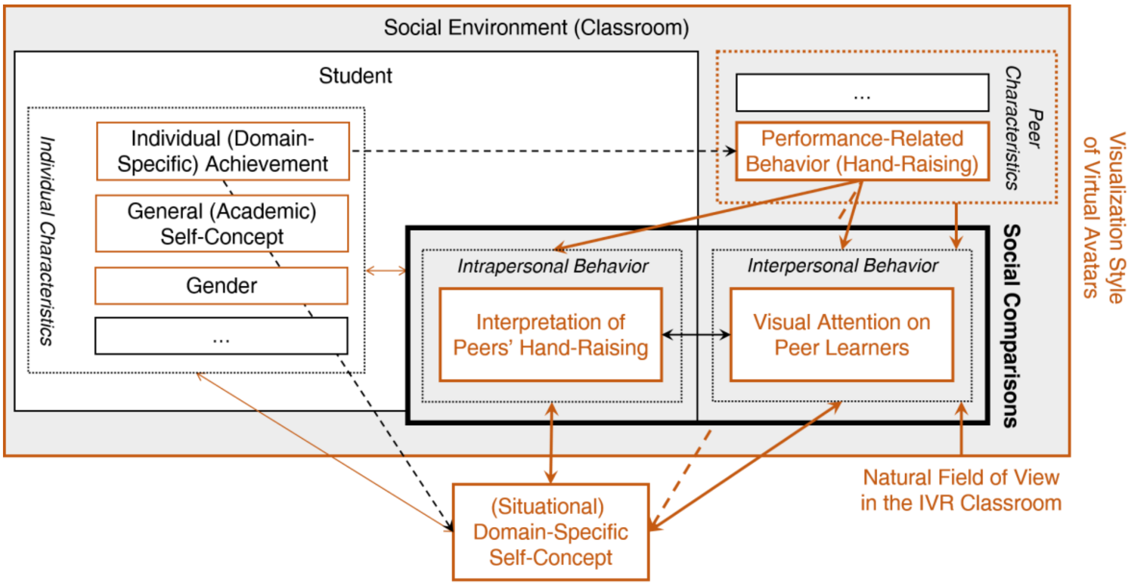


Figure 11. Relationships examined in the studies included in the present dissertation.

As Figure 11 shows, the present studies operationalized and manipulated social comparison information via peer learners' performance-related behavior in the classroom, specifically the proportion of students who raised their hands. By doing so, the present studies translated the—social-psychologically grounded—experimental approach of manipulating comparison information (see Chapter 1.2.2) in an authentic and yet controllable IVR classroom setting. Traditionally, the experimental approach of manipulating social comparison information has used so-called scenarios that outline a certain situation with experimentally manipulated comparison targets and the respective studies have examined participants' reactions to these fictitious scenarios (see, e.g., Mussweiler et al., 2004; Mussweiler & Strack, 2000). Using an IVR classroom as an experimental setting made it possible to apply this experimental paradigm in a more authentic manner; the study design in the present studies did not select one or two specific others and explain (i.e., interpret) their performance-related profiles. Instead, it placed participants in an authentic classroom setting and left the selection and interpretation of implicitly provided social comparison information (i.e., peer learners' hand-raising behavior) to the participants. Moreover, whereas traditional research designs have typically relied solely on self-reported self-evaluations as approximations of participants' reactions to the manipulated comparison information, the present dissertation put a major focus on the actual processing of social information to describe social comparisons in the classroom. Following the proposed theoretical model, the respective processes in the black box of social comparisons in the classroom were examined via students' *intra*personal and *inter*personal social comparison behaviors. As Figure 11 shows, the studies included in the present dissertation measured students' intrapersonal social comparison behavior (i.e., their covert cognitive responses to the social comparison information) via self-reports of their interpretations of their peers' hand-raising behavior, specifically the perceived performance level of the class. Students' interpersonal social comparison behavior (i.e., their overt behavioral responses to the social comparison information) was measured via eye-tracking data, specifically visual attention to peer learners. In this vein, the present studies are able to substantially contribute to (a) BFLPE research as they experimentally modeled and examined the mechanisms that are typically assumed to underlie social comparison effects as well as to (b) experimental social comparison research as they demonstrated how an established experimental approach, such as the one here, which involved the manipulation of comparison information, can be transformed into a more authentic and yet standardized setting by using IVR as an experimental tool.

In line with the BFLPE, the present studies showed an effect of classmates' performance-related behavior on students' situational self-concept (see Study 1), indicating that social comparisons also occur in an IVR classroom simulation. Importantly with regard to the authenticity of the IVR classroom simulation, the BFLPE could be reproduced in the IVR classrooms across different configurations; students' natural field of view and the style with which the avatars were represented had no impact on the effect from the manipulation of the class' performance level (i.e., peers' hand-raising behavior) on students' situational self-concept. Ultimately, reproducing a well-established finding, such as the BFLPE in the IVR classroom, provided grounds for examining the respective social comparison processes that were suggested by the theoretical model.

As discussed in detail in Chapter 6.1.1, the studies' results showed that the BFLPE on students' situational self-concept could be explained by students' individual responses to social comparison information (i.e., peers' hand-raising behavior) reflected in both intra- and interpersonal behavior. As suggested by the theoretical model that was proposed as a way to examine social comparisons in the classroom (see Figures 5 and 11), the findings of the present studies demonstrate that intra- and interpersonal social comparison behaviors are both simultaneously related to (a) situational characteristics of the social environment and (b) individual characteristics. Thereby, the insights gained from the present studies address one of the central criticisms of the BFLPE (see Dai & Rinn, 2008): To date, not only has BFLPE research been lacking the explicit examination of social comparisons, but moreover by default, it has also been assuming similar levels of engagement in social comparisons across different individuals and situations.

Contributing to a more differentiated understanding of students' social comparison behaviors, the present studies show that particularly students' observable social comparison behavior (i.e., visual attention to peers) differed considerably depending on the specifics of the IVR classroom situation, such as students' natural field of view of the classroom, the style with which the virtual avatars were represented, and the salience of the social comparison information that was presented (see Studies 2 and 3). Whereas the findings are limited to interindividual differences that are based on the different configurations of a specific IVR classroom situation,[25] they arguably reinforce the fact that students' engagement in social comparisons should not be assumed to be constant across different individuals and situations. Moreover, particularly the findings of Studies 1 and 2 highlight the importance of subjective perceptions and interpretations of social comparison information and emphasize the theoretical understanding of the self as an active entity that selectively processes social comparison information. The present studies did not explicitly examine potential moderators of the BFLPE and the respective social comparison behavior, but I argue that a theoretical focus on actual social information processing—combined with the possibility of isolating and manipulating single variables in an IVR classroom—provides the opportunity to address the discussion about potential moderators of the BFLPE from a new angle (see Chapter 1.2.1). Especially against the background that consistent evidence for substantial moderators of the BFLPE is missing (see, e.g., Marsh & Seaton, 2015; Marsh et al., 2021), considering that there are individual differences in how students respond to social comparison information (and potential moderators) could provide a worthwhile approach from which to revisit the overall mixed findings regarding moderators of the BFLPE. The respective implications for future research are discussed in Chapter 6.3.

[25] Chapter 6.3.1 discusses implications for future research on social comparisons in the classroom, including repeated-measures designs using IVR classroom scenarios. Such designs would allow additional insights into intraindividual differences across situations.

6.2 Strengths and Limitations

The three empirical studies discussed above were based on the same IVR experiment, which comes with strengths and limitations that need to be considered when interpreting the results. The fact that the three studies are based on the same IVR experiment has the advantage that the respective findings of the three studies can be related to each other and provide complementary insights into (a) the process of social comparisons and (b) the use of IVR as an experimental tool. At the same time, the strengths and limitations of the research design mostly apply to all three empirical studies in this dissertation. Before going into the respective details, I would like to highlight that in line with the so-called mapping principle suggested by Williams (2010), by which the extent to which human behaviors in IVR environments match those that naturally occur in real life can be examined, the present studies provide solid evidence for the validity of IVR classrooms by reproducing the well-established BFLPE in the experimental IVR classroom setting. Notably, whereas students' self-reported presence and their perceptions of the realism of the IVR lesson in the present studies did not indicate shortcomings or particular affordances of any of the implemented configurations, it is important to consider that the present studies examined only the effects of three configuration features (natural field of view from the position of the student's seat, the style with which the virtual avatars were represented, and performance-related peer behavior). These features have been argued to be particularly decisive for students' perceptions of the IVR classroom; however, given the plethora of design decisions that the configuration of an IVR classroom environment brings, I argue that the question about the best IVR classroom configuration for experimental research remains at least partially open (see the respective implications for future research in Chapter 6.3.2). This being said, I will use the following chapter to point out three central aspects that characterized the three studies in this dissertation and thereby must be considered when interpreting the results with respect to both theoretical and methodological contributions.

The first aspect that I would like to critically discuss is the sample used in the present studies. On the one hand, the sample of N = 381 sixth-graders is relatively large for this type of study, considering that most experimental studies on social comparisons—especially those using eye-tracking measures—work with less than half of this sample size. Moreover, the sample of sixth-graders is particularly informative for the IVR research community against the background that systematic IVR studies with children are scarce, and most findings on IVR experiences and perceptions of such IVR environments are based on samples of adolescents or adults (see, e.g., Bailey & Bailenson, 2017; Southgate et al., 2017). On the other hand, however, when interpreting the results of the present studies, one needs to consider that the sample is rather limited in the sense that only students in one grade level at academic track schools were included. As a consequence, the participating students showed comparably low variability in certain background characteristics (e.g., prior learning experiences, socioeconomic status). With regard to the IVR classroom experience, it is important to consider that perceptions of realism are presumed to differ significantly between different age groups and types of schools; in other words, what a sixth-grade student from an academic track school perceives as realistic in terms of classroom composition and classmates' behavior is probably very different from what a first-grader or a tenth-grade student from a community college in a disadvantaged area might find realistic. This aspect is also important to consider when aiming to replicate the present studies' findings in different countries: Not only are perceptions of realistic classroom situations likely to differ between cultures, but the role of the social environment for the formation of self-evaluations is also culturally different (e.g., Cross et al., 2011; Markus & Kitayama, 1991).

The second aspect that bears some important limitations concerns the measurements used in the present studies. With regard to the self-reports used to determine students' situational self-concepts after the IVR lesson, the present studies used an adaptation of one of the commonly used self-description questionnaires (specifically the SDQ III; Marsh, 1992; Schwanzer et al., 2005). After the 15-min IVR lesson, students were asked about their situational, criterion-oriented, and dispositional domain-specific self-concepts, whereby the situational self-concept measure included explicit references to the social environment (i.e., virtual peer learners). Only situational self-concept was affected by the experimental manipulation (see Study 1). Aside from general validity issues that always need to be considered when using self-reports of students' self-evaluations (see, e.g., the meta-analysis by Freund & Kasten, 2012), it is important to recognize that also in existing studies, the sizes of observed reference group effects on students' academic self-concepts have differed, depending on how much the wording of the respective measures was related to social comparisons compared with criterial standards or very general self-evaluations (see Marsh, Trautwein, et al., 2008). In line with the conceptualizations of self-concept adopted in the present dissertation (see Chapter 1), one would expect that the effect on situational (i.e., socially-oriented) self-concept would be reflected in more stable (i.e., criterion-oriented or dispositional domain-specific) self-conceptualizations after repeated experiences; however, as the present studies present only a one-time measure, the relationships between the situational measure (including the explicit reference to social counterparts in the classroom) and more general self-evaluations remain unclear. Respective implications for future studies are discussed in Chapter 6.3.1.

With regard to the use of eye-tracking data to gain insights into students' overt social comparison behaviors, the present studies used innovative approaches to analyze the gaze data to gain an in-depth understanding of how students process (social) information in the IVR classroom. The present studies thereby provided unprecedented insights into the use of gaze data to examine social comparisons by analyzing students' visual attention to peer learners (Study 2) and the overall distribution of attention across the IVR classroom (Study 3). However, both eye-tracking studies used aggregated gaze data over the 15-min lesson and therefore did not fully utilize the potential of the respective gaze data, for instance, to gain more detailed insights into the timeline of information processing throughout the IVR lesson or differential analyses of visual attention to single students (see, e.g., Kaakinen, 2021, and implications for future research in Chapter 6.3.1). In addition, the present dissertation used visual attention as an overt behavioral indicator of social comparisons (i.e., covert cognitive processes). Whereas theoretically and empirically speaking, there is a strong association between overt and covert processes (e.g., De Jaegher et al., 2010; Reichle et al., 2012), a major limitation of the present dissertation is that it did not examine that link in more detail. This is particularly crucial against the background that (a) overt behavior never fully approximates covert processes and (b) eye-tracking research cannot consider information that is perceived in the periphery or by other sensory inputs (e.g., auditory information). In other words, (a) students might seem like they are paying close attention to their classmates even when their minds are actually wandering and they are not really processing the information that their visual attention is focused on (see, e.g., Smallwood & Schooler, 2006); in addition, (b) even if students are in fact focusing their attention on what they are looking at, they might never be fully excluding what is happening in their peripheral field of view or what they are hearing, and the respective information might also have an impact even though it is not the focus of or reflected by students' visual attention (see, e.g., Schmitz et al., 2020, for an investigation of the role of the peripheral field of view in an IVR environment).[26] This is to say that whereas eye movements and visual attention are considered a relatively close and well-established approximation of covert cognitive processes (see the eye-mind link; Just & Carpenter, 1976; Rayner, 1998; Reichle et al., 2012), one needs to keep in mind that visual attention assessed via gaze data is only an approximation and is not a direct reflection of cognitive processes. The respective implications for future studies are discussed in Chapter 6.3.1.

[26] This is particularly important to consider with regard to the natural field of view, which greatly depends on the position of a student's seat in the classroom and differs significantly with regard to information in the periphery (i.e., significantly more peer learners when sitting in the back of the classroom).

The third aspect that I would like to critically discuss pertains to the experimental manipulation used in the present studies to examine social comparisons. Compared with traditional research designs, it is undoubtedly an advantage of the present IVR classroom that it was possible to manipulate a single aspect, such as classmate's performance-related behavior, while holding all other factors constant. The present studies varied virtual classmates' hand-raising behavior as an indicator of performance. Importantly, in previous research, students' hand-raising has been associated with achievement but has also been interpreted more generally as an indicator of behavioral engagement (Böheim, Knogler, et al., 2020; Böheim, Urdan, et al., 2020) and motivation (e.g., Südkamp et al., 2014). For instance, Südkamp et al. (2014) examined teachers' perceptions of students' performance-related behavior in a 2-D computer-based classroom simulation by using the proportion of correct answers as an indicator of achievement, and the participation rate (i.e., hand-raising) in class as an indicator of motivation. In the present studies, the answers given by the students who raised their hands were always correct so that the performance-related attributions of students' hand-raising behavior would be more salient. Whereas the present studies provided evidence that the experimental manipulation of hand-raising was indeed associated with participants' perceptions of the performance level of the class (Study 1), it is important to note that the manipulation of hand-raising behavior presents only one and even more so a rather "low level" of implicit performance-related information that students can use for their social comparisons. This might be a reason for the small effect that was found with respect to individual differences in situational self-concept (Study 1). At the same time, it calls for future research to investigate whether similar (or even more pronounced) effects occur with other or additional manipulations of classmates' performance-related behavior and—extending the focus of the present studies to classmates' behavior—manipulations of social comparison information beyond the information that can be gleaned from peer learners' behavior (see implications for future research in Chapter 6.3.1).

Moreover, the present studies used a novel classroom situation and an unknown lesson topic (i.e., computational thinking, which is not part of the curriculum in secondary schools until Grade 7) to examine students' social comparison behavior in an unbiased environment where one could somewhat observe "the emergence" of individual differences in self-concept in the specific novel subject and situation. The findings of the present studies provide insights into students' social comparison behaviors in a novel situation in an IVR classroom (see the Discussion in Chapter 6.1), specifically with regard to how students use and respond to their peer learners' hand-raising as an indicator of performance. However, compared with social comparisons in real-life classrooms, one needs to bear in mind that in real-life schooling, social comparisons (a) occur for longer than one 15-min lesson, and even more importantly, they also occur outside the classroom (e.g., in conversations with peers outside of class) and (b) are quickly affected by certain relationships among peers and with the teacher (e.g., after the first lesson, students may already have friends or rivals and a certain image of classmates that shapes their information processing). In this regard, I would like to point out that whereas social comparisons are widely acknowledged to be a major determinant of individual differences in students' academic self-concept, there are other types of comparisons and information that students use to evaluate themselves (see Chapter 1.1.3). Typically, these different sources of comparison information are difficult to disentangle as often illustrated by the example of grades: When students get a certain grade on a test, they will likely compare it with the class average or their best friend's or their biggest rival's grades (social comparison), they will remember their grades on recent tests and put them into relation with their previous grades (temporal comparison), they will also think about their grades in other subjects (dimensional comparison) and probably evaluate whether a certain grade is good enough to pass the class (criterial comparison). The present dissertation demonstrated that the use of an IVR classroom as an experimental tool makes it possible to isolate social comparisons in a novel environment without a specific focus on performance or other sources of comparison information. However, I do not want to claim that IVR research settings are able to fully filter out all other influences, such as comparison information that is acquired from the real world (e.g., low self-concept in STEM-related domains that influences experiences in an IVR classroom regardless of the novel content and environment). I therefore argue that the respective influences should be kept in mind and should be carefully considered when designing studies that will use IVR classrooms as an experimental tool.

Finally, I would like to highlight two strengths of the present studies that are particularly important with respect to implications for future research, namely, (a) interdisciplinary integration and (b) replicability. First, the present dissertation integrated different disciplines by drawing on (a) social and educational psychology's conceptualizations of self-concept and social comparisons, (b) cognitive psychology's approaches to (social) information processing, and (c) the technological affordances of IVR classrooms. As demonstrated in the preceding chapters of this Discussion, the findings of the present dissertation therefore yield innovative insights for researchers from different disciplines. Second, IVRs as an experimental tool provide great potential for replication studies (see, e.g., Blascovich et al., 2002). Once the IVR environment is programmed, it allows for direct and conceptual replications as well as the reproduction of similar experiments at relatively low cost and effort. As suggested by Irvine

(2021), I argue that respective replications—or cumulative studies that support theoretical and measurement developments as Irvine (2021) referred to them—are particularly important when it comes to IVR classroom studies: IVR technology allows novel insights and methodological approaches that—even if based on established theories and measures from real-world classrooms—have not explicitly been examined or employed in this manner before and therefore call for replications and cumulative support.

The following chapter focuses on the respective implications and future directions that are based on the aforementioned strengths and limitations of the present studies.

6.3 Implications and Future Directions

On the basis of the discussion of the present studies' results and the overall strengths and limitations of this dissertation, in the following chapter, I will discuss implications and future directions. More specifically, I will outline implications for research on social comparisons by integrating theoretical and methodological advancements from the present dissertation (Chapter 6.3.1) as well as future directions for the use of IVR classrooms as an experimental tool (Chapter 6.3.2). I will conclude with implications for educational policy and practice with regard to social comparison effects in the classroom and the use of IVR classrooms for effective virtual learning environments (Chapter 6.3.3).

6.3.1 Implications for Research on Social Comparisons in the Classroom

On the basis of the present dissertation's findings and guided by existing research, there are two aspects that I consider particularly promising with regard to future research on social comparisons in the classroom, namely, (a) the selection and analysis of process data to gain insights into social comparison processes and (b) the use of IVR classrooms for additional manipulations of social comparison information.

Use of Process Data to Gain Insights Into Social Comparison Processes

To examine both intra- and interpersonal social comparison processes, the present dissertation relied on one 15-min IVR classroom experience and used one-time measures of students' self-concept as well as aggregated gaze data. In order to extend these findings, future studies should consider temporal as well as semantic extensions to describe students' intra- and interpersonal social comparison behavior.

With regard to examining students' covert cognitive responses to social comparison information (i.e., *intra*personal social comparison behavior), future studies should employ a repeated-measures design in which students can experience multiple IVR lessons, and the subsequent effects on their self-concept can be examined on the basis of repeated experiences in the IVR classroom. Especially against the background of findings from Study 1 that showed significant effects of the 15-min IVR lesson only on students' situational (socially-oriented) self-concept but not on their criterion-oriented or dispositional domain-specific self-evaluations, the respective insights from a standardized and yet authentic classroom setting would contribute substantially to the understanding of the intraindividual stability and interindividual differences in students' self-concept. Moreover, to additionally account for other sources of comparison information except for classmates (see the Discussion in Chapter 6.2), I argue that particularly when investigating intrapersonal social comparison processes, the consideration of other types of internal comparisons (e.g., dimensional or temporal social comparisons) seems crucial. In other words, even though the IVR environment presents a novel classroom environment to students, it cannot be ruled out that their processing of social information in the classroom—and more specifically how they interpret it—is influenced by their experiences in other subjects or at previous timepoints. In this regard, particularly a research design that purposefully implements repeated IVR classroom experiences over a certain time as suggested before can present a novel approach that can be applied to systematically examine social comparison effects and the processes involved in internal comparisons simultaneously (see a call for respective studies, e.g., by Morina, 2021; Trautwein & Lüdtke, 2005; Wolff et al., 2018). In this vein, using IVR classrooms as a research tool provides, for instance, the opportunity to test the (reciprocal) internal/external-frame-of-reference model (Möller & Köller, 2001; Möller & Marsh, 2013; Möller et al., 2011; Möller et al., 2014) in an experimentally controlled and yet authentic classroom setting.

With regard to students' overt behavioral responses to social comparison information (i.e., *inter*personal social comparison behavior, e.g., their visual attention to peer learners), future studies should make use of more in-depth analyses of gaze data in terms of both (a) temporal aggregation and (b) semantic interpretability (Kaakinen, 2021). With respect to temporal aggregation, in addition to the repeated-measures designs mentioned before, future studies should aim to conduct a more fine-grained analysis of gaze data (and other behavioral measures) throughout the course of the IVR lesson. For instance, the use of crossed-lagged panel or continuous time models to examine the interrelationships and potential reciprocal dependencies of different types of measures (e.g., changes in pupil diameter as a reaction to visual attention to peers) and the development of certain behaviors over time (e.g., overall declining or increasing visual attention to peer learners or a focus on peer learners particularly while they are raising their hands) would extend the insights that can be gleaned from the present studies. With respect to the semantic interpretability of gaze markers, existing experimental research paradigms from social psychology provide valuable directions for the use of process data in future IVR classroom studies. The studies in the present dissertation applied the approach of manipulated comparison information (see Chapter 1.2.2) to an authentic setting by using an IVR classroom. In a similar vein, other experimental paradigms, such as social comparison choice, could be implemented in an IVR classroom setting. Implementing the general idea of comparison choice studies in an IVR classroom, future studies should examine, for instance, students' naturally occurring social comparison choices in an IVR classroom by examining visual attention to single students more distinctively. Importantly, the

IVR classroom setting as an experimentally controlled and yet authentic environment would allow researchers to simultaneously examine comparisons with specific others (as implied by comparison choice studies) and generalized others (as assumed by the BFLPE and implemented in the present studies).

Finally, as highlighted in Chapter 6.2, not only should future research further develop and extend measures of intra- and interpersonal social comparison behavior, but it should examine (a) the link between the covert and overt processes that are involved and (b) how they are shaped not only by situational characteristics but even more so by interindividual characteristics. Particularly when examining comparisons in an IVR classroom, whereby students have to make ad hoc inferences about relevant characteristics and social comparison information from their peers' behavior (compared with more "contextualized" comparisons in the real world where interactions with peers typically continue outside the classroom), it seems crucial to obtain an in-depth understanding of the link between overt and covert social comparison behaviors. A respective in-depth understanding combined with insights into interindividual influences in this specific setting would allow researchers to further advance the theoretical understanding of what happens in the black box of social information processing in the classroom.

Experimental Approaches and Manipulations of Social Comparison Information

As discussed in Chapter 6.2, the present dissertation worked with manipulations of social comparison information via virtual peer learners' hand-raising behavior. I argue that future research should use the potential of experimental manipulations in the IVR classroom to extend these insights and examine how other peer or teacher behaviors affect intra- and interpersonal social comparison behavior. At the same time, I argue that an IVR classroom presents not just a promising avenue for examining students' responses to social comparison information in the classroom, but it additionally allows for an ideal experimental design to systematically and yet authentically investigate the influence of moderating variables on the BFLPE. In this vein, other manipulations of social comparison information should build on existing findings that highlight the role of teacher-student relationships and interactions for individual differences in students' self-concepts and the extent of the BFLPE (e.g., Corpus et al., 2006; Leflot et al., 2010; Lüdtke et al., 2005; Roy et al., 2015; Schwabe et al., 2019). For instance, Lüdtke et al. (2005) used students' self-reports of their teacher's frame of reference to examine its role in determining self-concept formation and the BFLPE. The use of IVR classrooms as an experimental tool would allow researchers to examine a respective research question in a standardized and yet authentic environment by manipulating whether the virtual teacher uses a social or individual reference standard when giving feedback to students.

Moreover, manipulations of social comparison information in future IVR studies should implement additional variations of virtual peer learners' behavior. Extending the findings of the present studies, these manipulations could, for instance, vary the correctness of students' answers and other behaviors except for their hand-raising and examine the extent to which students infer performance levels from the respective behaviors. In addition, against the background of existing research that has highlighted the role of perceived (dis)similarities between individuals and their comparison targets, manipulations of peer learners in the IVR classroom should involve manipulations of the composition of the classroom to systematically examine how social categorizations (Turner et al., 1987) and perceived psychological closeness (Mussweiler, 2003) affect intra- and interpersonal social comparison behavior. For instance—whereas it is hardly feasible and ethically impossible in real-world settings—it is possible to manipulate the gender ratio or the ethnic composition of peer learners in an IVR classroom to examine how various classroom compositions affect students' cognitive (e.g., interpretation) and behavioral (e.g., attention distribution) responses to social comparison information.

6.3.2 The Future of IVR Classrooms as an Experimental Tool

The present dissertation has provided insights into how an IVR classroom can be used to gain insights into social comparisons and the respective processing of social information in the classroom. Building on the present dissertation's findings, I would like to highlight two important implications for the future use of IVR classrooms as an experimental tool, pertaining (a) to the question of validity and the role of different configurations for students' processing of social information in the IVR classroom and (b) to the further potential of IVR classrooms to examine social comparisons as well as other classroom phenomena that are relevant to educational and social psychology research in general.

Additional Features and IVR Classroom Configurations

As outlined in Chapter 6.2, the present dissertation provided solid evidence for the valid use of IVR classrooms by reproducing the well-established BFLPE. At the same time, the IVR classroom configurations examined in the present studies utilized only a small number of the many design and configuration features that an IVR classroom allows for. Against this background, I argue that future research should more closely investigate how different IVR classroom configurations affect students' perceptions of the IVR classroom as well as their cognitive and behavioral responses to the information provided in it. In order to obtain a comprehensive understanding and substantiated proof of concept for the use of an IVR classroom as an experimental tool for classroom research, there is a need for studies with different configurations of IVR classrooms and samples consisting of diverse groups of students (e.g., from different grade levels and types of schools).

In this regard, one needs to bear in mind that IVR technology develops at a fast pace, and therefore, what was used as a state-of-the-art IVR classroom environment in the present dissertation will most likely be somewhat outdated—or will at least be missing some significantly enhanced features from the latest developments—in a couple of years. Current research trends from the IVR research community have pointed to critical aspects in the configuration of IVR environments, some of which appear to be of great importance for the design and use of IVR classrooms as authentic research environments. I would like to highlight two of these research trends, namely, the role of users' own embodiment in an IVR environment and the implementation of real-time interactions (see, e.g., Bainbridge, 2007). With respect to the first, there is a substantive and rapidly growing body of research that has examined the effects of self-representation via avatars on users' behavior and experience in the IVR environment (see Ratan et al., 2020, for a recent meta-analysis). Such studies have found that the ways in which users perceive and respond to IVR environments depend on their own representation (e.g., Aseeri & Interrante, 2018; Jo et al., 2016; Joy et al., 2021; Ogawa et al., 2018; Villani et al., 2012; Wirth et al., 2021). Whereas in the present dissertation studies, participating students had no representation of their own body in the IVR classroom, these findings suggest that self-representations likely have a decisive effect on students' experiences in an IVR lesson. Thus, particularly with regard to research on self-relevant information, such as in the present dissertation, it would be interesting to conduct future research to determine how a certain choice of or predefined representation of participants in the IVR classroom affects their experience and behavior in the IVR classroom. With respect to the second aspect mentioned above, real-time interactions or interactive virtual agents driven by artificial intelligence are becoming more and more popular in IVR environments (Herrera et al., 2020; Krämer, 2017; Liao et al., 2019; Syrjämäki et al., 2020). Whereas the fully preprogrammed IVR classroom simulation used in the present studies provided a fully standardized research environment, it seems worthwhile for future studies to make use of the potential to implement real-time interactions in an IVR classroom. Especially against the background that social processes in the classroom in particular are never a one-way street but are based on reciprocal relationships (see, e.g., Richardson & Gobel, 2015), the implementation of experimentally controllable features that foster students' perceptions of a responsive social IVR environment should be explored further.

Application to Phenomena Beyond Social Comparisons in the Classroom

As Alexander (2018) pointed out in her treatise on research in educational psychology, to date, "there still remains a compelling search for causality afforded by experimental investigations" (p. 149). Considering the preconditions and threats to drawing causal inferences (see West & Thoemmes, 2010), the affordances of research with IVR classrooms as experimental tools clearly provide an unprecedented opportunity for educational researchers to add to existing correlational findings on various theories by using experimental designs with high ecological validity. The present dissertation utilized these affordances to gain systematic and authentic insights into social comparisons in the classroom and the mechanisms that underlie the BFLPE (see Figure 10). I argue that the present dissertation's findings should encourage educational and social psychology researchers to make use of IVR classrooms as ecologically valid and yet controllable research designs that can be applied to address the lack of experimental and causal evidence beyond social comparison processes in the classroom. Inspired by the present dissertation, particularly research on contextual and compositional effects in the classroom (see, e.g., Brophy & Good, 1974; Harker & Tymms, 2004; Marsh et al., 2012; Thrupp et al., 2002; Trautwein et al., 2015), which usually struggles to disentangle different influences in real-world settings, should be further explored by future research with IVR classrooms. Beyond effects on self-concept, future research could also systematically examine peer effects—or more broadly, contextual and compositional effects—on students' achievement (see, e.g., Gottfried, 2012; Hattie, 2002; Lavy et al., 2011), their emotions and motivation while learning (see, e.g., Frenzel et al., 2007; Hardré & Sullivan, 2008; Pekrun et al., 2019), as well as their perceptions of teaching quality (see, e.g., Decristan et al., 2017; Fauth et al., 2021; Fauth et al., 2020; Lüdtke & Robitzsch, 2020). By enabling systematic and yet authentic insights into underlying classroom processes,[27] the IVR classroom can additionally provide the opportunity to add to common topics in educational psychology research that—due to natural limitations in field-based research designs—typically rely on correlational analyses, such as teacher-student relationships and phenomena such as emotional contagion (see, e.g., Becker et al., 2014; Frenzel et al., 2018; Frisby, 2019).

[27] The respective processes need to be operationalized and described with the specific research interest in mind. For instance, the present dissertation was aimed at describing social comparison processes and therefore systematized these processes in intra- and interpersonal behaviors reflecting the individual cognitive and behavioral responses to social information provided by peer learners in the classroom.

As a final note and in an attempt to provide an outlook for the future of IVR classrooms as an experimental tool, I argue that the present dissertation provided initial insights into the potential that the increasing interdisciplinary integration of educational and social psychology with cognitive and computer sciences yields. In line with what Alexander (2018) and Harris (2018) elaborated on in their articles about the future of research in educational psychology, future studies with IVR classrooms should aim to further utilize and explore the unique potential of an interdisciplinary integration as such. As demonstrated by the present dissertation, the combination of theoretical foundations from educational and social psychology with the affordances of IVR technology (see, e.g., Bainbridge, 2007; Blascovich et al., 2002; Fox et al., 2009) provides great potential. Future research should additionally make use of technology-based and increasingly automated measures of students' behaviors on the basis of big and naturally occurring data (see Girard & Cohn, 2016; Miller et al., 2021; Paxton & Griffiths, 2017) and computational models of fundamental human behaviors (see, e.g., Fridman & Kaminka, 2011, for a computational model of social comparisons).

6.3.3 What the Present Dissertation Means for Educational Policy and Practice

As outlined in the present dissertation, beliefs that students hold about their own academic abilities, such as "I am good at mathematics," "I struggle to learn foreign languages," or "I can solve technical problems quickly," are shaped to a large extent by social comparisons with their classmates. Against the background that these beliefs (i.e., students' academic self-concept) have far-reaching effects on students' general well-being, interests, achievement, and overall academic trajectories way beyond traditional schooling (e.g., Arens et al., 2019; Cross et al., 2003; Göllner et al., 2018; Trautwein & Möller, 2016; Valentine et al., 2004), it has been of central interest to educational policy and practice to design educational environments that can help attenuate detrimental effects on students' self-evaluations. Moreover, educational practice and policy have been increasingly interested in leveraging the potential of innovative technologies, such as IVR, for educational purposes (e.g., Howard et al., 2021; Johnston et al., 2017; Karutz & Bailenson, 2015; Seidel & Chatelier, 2013).

I would like to highlight that the present dissertation's objectives were to theoretically and methodologically advance research on social comparisons in the classroom and the use of IVR classrooms. In this vein, the present dissertation predominantly yielded implications for the scientific community (see Chapters 6.3.1 and 6.3.2), and I argue that an adequate interpretation of the present dissertation's findings for educational policy and practice is limited to very cautious conclusions that are based on the specific experimental setting and sample used

(Bromme & Goldman, 2014; Tseng, 2012). This being said, I would like to outline some tentative implications for educational practitioners and policy-makers with regard to (a) social comparisons in the classroom and (b) the use of IVR classrooms in instructional practice.

First, with respect to social comparisons in the classroom, educational psychologists have long tried to prevent negative reference group effects on students' academic self-concept by using a variety of strategies, including redesigning grading practices (e.g., implications from Trautwein et al., 2006), changing the standards used for comparisons when giving feedback (e.g., Lüdtke et al., 2005), and modifying the composition of learning groups (e.g., Harker & Tymms, 2004). These measures ultimately follow the logic that learning environments can be redesigned in such a way that they can mitigate or eliminate negative effects. However, in line with much respected theoretical work by sociologists and economists who have suggested that social comparisons are universal and humans generally care more about their relative rather than absolute standing (Frank, 1985, 2013; Hyman, 1942), the findings of the present dissertation suggest that social comparison processes are ubiquitous in the classroom, even in situations that are clearly not designed to be performance-oriented and that emphasize learning rather than evaluation. At the same time, the dissertation provides indications of how to potentially address this issue in the future. On the one hand, the results showed effects of peer learners' performance-related behavior only on students' situational self-concept. In line with the conceptual framework of self-concept (see Chapters 1.1.2 and 1.1.3), repeated experiences in situations as in the IVR classroom can be expected to lead to situational self-evaluations manifesting themselves in a more stable self-concept. Whereas future research should investigate this distinction and relationship more closely in longer and repeated learning sequences (see the implications in Chapter 6.3.1), this finding highlights the importance of single situations and the respective experiences of success and failure that ultimately determine students' self-concepts. On the other hand, the results of the present studies indicate that there are interindividual differences in how students process and respond to social comparison information. Accordingly, the present dissertation's findings suggest that social comparisons are ubiquitous, but they additionally indicate that students do not all respond to certain kinds of social comparison information in the same way. Whereas the respective processes need to be examined in more depth in future research (see the implications in Chapter 6.3.1), these findings emphasize the role of interindividual differences in social learning processes. Taken together, the findings of the present dissertation encourage educational practitioners to be mindful of the consequences that single situations have on students' individual evaluations of themselves and their abilities, especially in novel teaching and learning situations without established social dynamics and individual self-perceptions.

Second, with regard to the use of IVR classrooms in instructional practice, it is first important to note that learners do not automatically learn better in IVR settings compared with traditional (media) formats; reviews and meta-analyses of respective studies have indicated that it is about much more than just innovative presentations of learning content in an IVR environment (Howard et al., 2021; Merchant et al., 2014; Mikropoulos & Natsis, 2011). The present dissertation offers some insights into how the transformation of traditional classrooms into IVR environments can be used to strategically implement certain classroom configurations for more effective learning (e.g., in remote learning scenarios). As Alexander (2018) put it, such strategic implementations lead to "varied educational contexts that have the potential to enhance (and, thus, alter) the process of learning in significant ways" (p. 156). The present studies, for instance, suggest that the natural field of view when sitting closer to the teacher is more beneficial for students' focus on instructional content presented in the front of the classroom. Moreover, in an attempt to strategically use virtual peer learners in an IVR classroom as pedagogical agents (see, e.g., Bailenson et al., 2008; Hudson & Hurter, 2016; Krämer, 2017; Makransky, Wismer, et al., 2019), the present dissertation indicates that the style with which virtual avatars are represented is less decisive than unambiguous (performance-related) behavior of the virtual peers with respect to whether students will attend to and respond to the peers. In collaboration with educational researchers, educational policy and practice should consider such findings when designing and aiming to leverage the potential of IVR classroom-based learning applications in schools.

6.4 Conclusion

Students' academic self-concept plays a central role in their learning and academic trajectories. It is widely acknowledged that social comparisons in the classroom are a major determinant of the respective beliefs that students hold about their own academic abilities. However, due to the natural limitations of research in real-world classrooms and lab-based experiments, a systematic and in-depth examination of social comparisons in the classroom has been missing. The present dissertation used an IVR classroom as an experimental tool that is able to combine the best of both worlds—authenticity from real-world classroom research and experimental control from lab settings—to revisit social comparisons in the classroom. Thereby, the present dissertation provided novel insights into social comparison processes in the classroom as the mechanisms that underlie the BFLPE. More specifically, the present dissertation identified covert and overt social comparison behaviors as a reflection of students' cognitive and behavioral responses to social comparison information. By relating respective social comparison behaviors to classmates' performance-related behavior as well as to individual differences in students' situational self-concept, the present dissertation contributed to an advanced theoretical understanding of social comparisons in the classroom. Moreover, the present dissertation utilized the affordances of state-of-the-art IVR technology to collect standardized process data, specifically students' gaze behavior, to gain insights into social comparisons and the respective processing of social information in the classroom.

In sum, the present dissertation successfully revisited students' social comparisons in the classroom and provided new insights into the underlying mechanisms that lead to individual differences in students' academic self-concept. Moreover, the present dissertation demonstrated how the potential of immersive virtual realities as an experimental tool can be used for theoretical and methodological advancements of research on social comparisons in the classroom and beyond.

7

REFERENCES

7 References

Adams, R., Finn, P., Moes, E., Flannery, K., & Rizzo, A. S. (2009). Distractibility in attention/deficit/hyperactivity disorder (ADHD): The virtual reality classroom. *Child Neuropsychology*, *15*(2), 120–135. https://doi.org/10.1080/09297040802169077

Albert, S. (1977). Temporal comparison theory. *Psychological Review*, *84*(6), 485–503. https://doi.org/10.1037/0033-295X.84.6.485

Alexander, P. A. (2018). Past as prologue: Educational psychology's legacy and progeny. *Journal of Educational Psychology*, *110*(2), 147–162. https://doi.org/10.1037/edu0000200

Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology*, *35*(11), 919–925. https://doi.org/10.1080/0144929x.2016.1212931

Anderson, N. C., Anderson, F., Kingstone, A., & Bischof, W. F. (2015). A comparison of scanpath comparison methods. *Behavior Research Methods*, *47*(4), 1377–1392. https://doi.org/10.3758/s13428-014-0550-3

Appel, T., Scharinger, C., Gerjets, P., & Kasneci, E. (2018). Cross-subject workload classification using pupil-related measures. *ETRA '18: 2018 Symposium on Eye Tracking Research and Applications*, Article 4. https://doi.org/10.1145/3204493.3204531

Arens, A. K., Schmidt, I., & Preckel, F. (2019). Longitudinal relations among self-concept, intrinsic value, and attainment value across secondary school years in three academic domains. *Journal of Educational Psychology*, *111*(4), 663–684. https://doi.org/10.1037/edu0000313

Arigo, D., Mogle, J. A., Brown, M. M., Pasko, K., Travers, L., Sweeder, L., & Smyth, J. M. (2019). Methods to assess social comparison processes within persons in daily life: A scoping review. *Frontiers in Psychology*, *10*, Article 2909. https://doi.org/10.3389/fpsyg.2019.02909

Aseeri, S. A., & Interrante, V. (2018). The influence of avatar representation and behavior on communication in social immersive virtual environments. *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 823–824. https://doi.org/10.1109/VR.2018.8446261

Aspinwall, L. G., & Taylor, S. E. (1993). Effects of social comparison direction, threat, and self-esteem on affect, self-evaluation, and expected succes. *Journal of Personality and Social Psychology*, *64*(5), 708–722. https://doi.org/10.1037/0022-3514.64.5.708

Attard-Johnson, J., C, O. C., & Bindemann, M. (2019). Comparing methods for the analysis of pupillary response. *Behavior Research Methods*, *51*(1), 83–95. https://doi.org/10.3758/s13428-018-1108-6

Bailenson, J., Aharoni, E., Beall, A. C., Guadagno, R. E., Dimov, A., & Blascovich, J. (2004). Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. In *Proceedings of the 7th Annual International Workshop on Presence* (pp. 1864–1105). International Society for Presence Research.

Bailenson, J., Beall, A. C., & Blascovich, J. (2002). Gaze and task performance in shared virtual environments. *The Journal of Visualization and Computer Animation, 13*(5), 313–320. https://doi.org/10.1002/vis.297

Bailenson, J., Blascovich, J., Beall, A. C., & Loomis, J. (2001). Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence, 10*(6), 583–598. https://doi.org/10.1162/105474601753272844

Bailenson, J., Swinth, K., Hoyt, C., Persky, S., Dimov, A., & Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence, 14*(4), 379–393. https://doi.org/10.1162/105474605774785235

Bailenson, J., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science, 16*(10), 814–819. https://doi.org/10.1111/j.1467-9280.2005.01619.x

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences, 17*(1), 102–141. https://doi.org/10.1080/10508400701793141

Bailey, J. O., & Bailenson, J. (2017). Immersive virtual reality and the developing child. In F. C. Blumberg & P. J. Brooks (Eds.), *Cognitive development in digital contexts* (pp. 181–200). Academic Press. https://doi.org/10.1016/b978-0-12-809481-5.00009-2

Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. *Science, 317,* 472–476. https://doi.org/10.1126/science.1146930

Baldwin, J. M. (1897). *Social and ethical interpretations*. MacMillan.

Bannister, B. D. (1986). Performance outcome feedback and attributional feedback: Interactive effects on recipient responses. *Journal of Applied Psychology, 71*(2), 203–210. https://doi.org/10.1037/0021-9010.71.2.203

Bauer, A., Schneider, S., Waldorf, M., Adolph, D., & Vocks, S. (2017). Familial transmission of a body-related attentional bias – An eye-tracking study in a nonclinical sample of female adolescents and their mothers. *PloS One, 12*(11), Article e0188186. https://doi.org/10.1371/journal.pone.0188186

Bauer, A., Schneider, S., Waldorf, M., Braks, K., Huber, T. J., Adolph, D., & Vocks, S. (2017). Selective visual attention towards oneself and associated state body eatisfaction: An eye-tracking study in adolescents with different types of eating disorders. *Journal of Abnormal Child Psychology*, *45*(8), 1647–1661. https://doi.org/10.1007/s10802-017-0263-z

Baumeister, R. F., & Bushman, B. J. (2011). The self. In R. F. Baumeister & B. J. Bushman (Eds.), *Social Psychology and Human Nature* (2nd ed., pp. 57–96). Wadsworth.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness or healthier lifestyles? *Psychological Science in the Public Interest*, *4*(1), 1–44. https://doi.org/10.1111/1529-1006.01431

Baumgartner, T., Speck, D., Wettstein, D., Masnari, O., Beeli, G., & Jancke, L. (2008). Feeling present in arousing virtual reality worlds: prefrontal brain regions differentially orchestrate presence experience in adults and children. *Frontiers in Human Neuroscience, 2*, Article 8. https://doi.org/10.3389/neuro.09.008.2008

Becker, E. S., Goetz, T., Morger, V., & Ranellucci, J. (2014). The importance of teachers' emotions and instructional behavior for their students' emotions – An experience sampling analysis. *Teaching and Teacher Education, 43*, 15–26. https://doi.org/10.1016/j.tate.2014.05.002

Becker, M., & Neumann, M. (2016). Context-related changes in academic self concept development: On the long-term persistence of Big-Fish-Little-Pond effects. *Learning and Instruction, 45*, 31–39. https://doi.org/10.1016/j.learninstruc.2016.06.003

Becker, M., & Neumann, M. (2018). Longitudinal Big-Fish-Little-Pond effects on academic self-concept development during the transition from elementary to secondary schooling. *Journal of Educational Psychology, 110*(6), 882–897. https://doi.org/10.1037/edu0000233

Bellini, H., Chen, W., Sugiyama, M., Shin, M., Alam, S., & Takayama, D. (2016). *Virtual & augmented reality: Understanding the race for the next computing platform* (Profiles in Innovation, Issue 1). The Goldman Sachs Group.

Binda, P., & Gamlin, P. D. (2017). Renewed attention on the pupil light reflex. *Trends in Neurosciences*, *40*(8), 455–457. https://doi.org/10.1016/j.tins.2017.06.007

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence*, *12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Bioulac, S., Lallemand, S., Rizzo, A., Philip, P., Fabrigoule, C., & Bouvard, M. P. (2012). Impact of time on task on ADHD patient's performances in a virtual classroom. *European Journal of Paediatric Neurology, 16*(5), 514–521. https://doi.org/10.1016/j.ejpn.2012.01.006

Blanton, H., Gibbons, F. X., Buunk, A. P., & Kuyper, H. (1999). When better-than-others compare upward: Choice of comparison and comparative evaluation as independent predictors of academic performance. *Journal of Personality and Social Psychology, 76*(3), 420–430. https://doi.org/10.1037/0022-3514.76.3.420

Blascovich, J. (2002). Social influence within immersive virtual environments. In R. Schroeder (Ed.), *The social life of avatars* (pp. 127–145). Springer. https://doi.org/10.1007/978-1-4471-0277-9_8

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Blume, F., Göllner, R., Moeller, K., Dresler, T., Ehlis, A.-C., & Gawrilow, C. (2019). Do students learn better when seated close to the teacher? A virtual classroom study considering individual levels of inattention and hyperactivity-impulsivity. *Learning and Instruction, 61*, 138–147. https://doi.org/10.1016/j.learninstruc.2018.10.004

Böheim, R., Knogler, M., Kosel, C., & Seidel, T. (2020). Exploring student hand-raising across two school subjects using mixed methods: An investigation of an everyday classroom behavior from a motivational perspective. *Learning and Instruction*, *65*, Article 101250. https://doi.org/10.1016/j.learninstruc.2019.101250

Böheim, R., Urdan, T., Knogler, M., & Seidel, T. (2020). Student hand-raising as an indicator of behavioral engagement and its role in classroom learning. *Contemporary Educational Psychology, 62*, Article 101894. https://doi.org/10.1016/j.cedpsych.2020.101894

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Bönsch, A., Wendt, J., Overath, H., Gurerk, O., Harbring, C., Grund, C., Kittsteiner, T., & Kuhlen, T. W. (2017). Peers at work: Economic real-effort experiments in the presence of virtual co-workers. In *2017 IEEE Virtual Reality (VR)* (pp. 301–302). IEEE. https://doi.org/10.1109/VR.2017.7892296

Bozkir, E., Stark, P., Gao, H., Hasenbein, L., Hahn, J.-U., Kasneci, E., & Göllner, R. (2021). Exploiting object-of-interest information to understand attention in VR classrooms. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 597–605). IEEE. https://doi.org/10.1109/VR50410.2021.00085

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Braun, L., Rieger, S., Spengler, M., Göllner, R., Rose, N., Trautwein, U., & Nagengast, B. (2020). Rethinking the elusive construct of global self-concept: A latent composite score as the apex of the Shavelson model. PsyArXiv. https://doi.org/10.31234/osf.io/dbkw6

Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology*, *66*(2), 268–275. https://doi.org/10.1037/0022-3514.66.2.268

Bric, J. D., Lumbard, D. C., Frelich, M. J., & Gould, J. C. (2016). Current state of virtual reality simulation in robotic surgery training: A review. *Surgical Endoscopy, 30*(6), 2169–2178. https://doi.org/10.1007/s00464-015-4517-y

Brisson, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal, 54*(6), 1048–1078. https://doi.org/10.3102/0002831217716084

Bromme, R., & Goldman, S. R. (2014). The public's bounded understanding of science. *Educational Psychologist, 49*(2), 59–69. https://doi.org/10.1080/00461520.2014.921572

Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. Holt, Rinehart and Winston.

Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology, 102*(4), 964–981. https://doi.org/10.1037/a0019644

Bulu, S. T. (2012). Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education, 58*(1), 154–161. https://doi.org/10.1016/j.compedu.2011.08.024

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*(4), 523–547. https://doi.org/10.1037/0033-295x.97.4.523

Bunge, M. (1963). A general black box theory. *Philosophy of Science, 30*(4), 346–358. https://doi.org/10.1086/287954

Burnside, R., & Ullsperger, M. (2020). Social comparison impacts stimulus evaluation in a competitive social learning task. *PloS One, 15*(6), Article e0234397. https://doi.org/10.1371/journal.pone.0234397

Buunk, A. P., Cohen-Schotanus, J., & van Nek, R. H. (2007). Why and how people engage in social comparison while learning social skills in groups. *Group Dynamics, 11*(3), 140–152. https://doi.org/10.1037/1089-2699.11.3.140

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes, 102*(1), 3–21. https://doi.org/10.1016/j.obhdp.2006.09.007

Buunk, A. P., & Mussweiler, T. (2001). New directions in social comparison theory. *European Journal of Social Psychology, 31*(5), 467–475. https://doi.org/10.1002/ejsp.77

Cairns, E., McWhirter, L., Duffy, U., & Barry, R. (1990). The stability of self-concept in late adolescence: Gender and situational effects. *Personality and Individual Differences, 11*(9), 937–944. https://doi.org/10.1016/0191-8869(90)90275-V

Cambria, J., Brandt, H., Nagengast, B., & Trautwein, U. (2017). Frame of reference effects on values in mathematics: evidence from German secondary school students. *ZDM Mathematics Education, 49*(3), 435–447. https://doi.org/10.1007/s11858-017-0841-0

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Cebeci, B., Celikcan, U., & Capin, T. K. (2019). A comprehensive study of the affective and physiological responses induced by dynamic virtual reality environments. *Computer Animation and Virtual Worlds, 30*(3-4), Article e1893. https://doi.org/10.1002/cav.1893

Cheng, K.-H., & Tsai, C.-C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education*, *140*, Article 103600. https://doi.org/10.1016/j.compedu.2019.103600

Cheng, R. W., McInerney, D. M., & Mok, M. M. C. (2014). Does Big-Fish-Little-Pond effect always exist? Investigation of goal orientations as moderators in the Hong Kong context. *Educational Psychology*, *34*(5), 561–580. https://doi.org/10.1080/01443410.2014.898740

Cheryan, S., Meltzoff, A. N., & Kim, S. (2011). Classrooms matter: The design of virtual classrooms influences gender disparities in computer science classes. *Computers & Education*, *57*(2), 1825–1835. https://doi.org/10.1016/j.compedu.2011.02.004

Chiesi, A. M. (2001). Network analysis. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 10499–10502). Pergamon. https://doi.org/10.1016/B0-08-043076-7/04211-X

Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education*, *78*, 66–77. https://doi.org/10.1016/j.compedu.2014.05.005

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/s1364-6613(00)01476-5

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Cialdini, R. B., & Richardson, K. D. (1980). Two indirect tactics of image management: Basking and blasting. *Journal of Personality and Social Psychology*, *39*(3), 406–415. https://doi.org/10.1037/0022-3514.39.3.406

Collaco, E., Kira, E., Sallaberry, L. H., Queiroz, A. C. M., Machado, M., Crivello, O., Jr., & Tori, R. (2021). Immersion and haptic feedback impacts on dental anesthesia technical skills virtual reality training. *Journal of Dental Education*, *85*(4), 589–598. https://doi.org/10.1002/jdd.12503

Collins, R. (2000). Among the better ones: Upward assimilation in social comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison* (pp. 159–171). Springer. https://doi.org/10.1007/978-1-4615-4237-7_9

Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin*, *119*(1), 51–69. https://doi.org/10.1037/0033-2909.119.1.51

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*(4), 594–628. https://doi.org/10.1016/j.jml.2005.08.005

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*(2), 261–288. https://doi.org/10.1037//0033-295X

Cooley, D. H. (1902). *Human nature and the social order*. Scribners.

Corpus, J. H., Ogle, C. M., & Love-Geiger, K. E. (2006). The effects of social-comparison versus mastery praise on children’s intrinsic motivation. *Motivation and Emotion, 30*(4), 333–343. https://doi.org/10.1007/s11031-006-9039-4

Coyne, J. T., Foroughi, C., & Sibley, C. (2017). Pupil diameter and performance in a supervisory control task: A measure of effort or individual differences? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61*(1), 865–869. https://doi.org/10.1177/1541931213601689

Cross, S. E., Gore, J. S., & Morris, M. L. (2003). The relational-interdependent self-construal, self-concept consistency, and well-being. *Journal of Personality and Social Psychology, 85*(5), 933–944. https://doi.org/10.1037/0022-3514.85.5.933

Cross, S. E., Hardin, E. E., & Gercek-Swing, B. (2011). The what, how, why, and where of self-construal. *Personality and Social Psychology Review, 15*(2), 142–179. https://doi.org/10.1177/1088868310373752

Cummings, J. J., & Bailenson, J. (2016). How immersive is enough? A meta-analysis of the affect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Cummings, J. J., Tsay-Vogel, M., Cahill, T. J., & Zhang, L. (2021). Effects of immersive storytelling on affective, cognitive, and associative empathy: The mediating role of presence. *New Media & Society*, Online first publication. https://doi.org/10.1177/1461444820986816

Curtin, K. M. (2018). Network analysis. In B. Huang (Ed.), *Comprehensive geographic information systems* (pp. 153–161). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09599-3

Dai, D. Y., & Rinn, A. N. (2008). The Big-Fish-Little-Pond effect: What do we know and where do we go from here? *Educational Psychology Review, 20*(3), 283–317. https://doi.org/10.1007/s10648-008-9071-x

Dai, D. Y., Rinn, A. N., & Tan, X. (2013). When the big fish turns small: Effects of participating in gifted summer programs on academic self-concepts. *Journal of Advanced Academics, 24*(1), 4–26. https://doi.org/10.1177/1932202x12473425

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences, 14*(10), 441–447. https://doi.org/10.1016/j.tics.2010.06.009

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*(6130), 305–308. https://doi.org/10.1126/science.1230579

Decristan, J., Fauth, B., Kunter, M., Büttner, G., & Klieme, E. (2017). The interplay between class heterogeneity and teaching quality in primary school. *International Journal of Educational Research, 86*, 109–121. https://doi.org/10.1016/j.ijer.2017.09.004

Demo, D. H. (1992). The self-concept over time: Research issues and directions. *Annual Review of Sociology, 18*(1), 303–326. https://doi.org/10.1146/annurev.so.18.080192.001511

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology, 110*(8), 1112–1126. https://doi.org/10.1037/edu0000259

Dickhauser, O., & Reinhard, M. A. (2006). Factors underlying expectancies of success and achievement: The influential roles of need for cognition and general or specific self-concepts. *Journal of Personality and Social Psychology, 90*(3), 490–500. https://doi.org/10.1037/0022-3514.90.3.490

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Muhlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. *Frontiers in Psychology, 6*, Article 26. https://doi.org/10.3389/fpsyg.2015.00026

Diestel, R. (2017). *Graph theory* (5th ed.). Springer. https://doi.org/10.1007/978-3-662-53622-3

Dijkstra, P., Kuyper, H., van der Werf, G., Buunk, A. P., & van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research, 78*(4), 828–879. https://doi.org/10.3102/0034654308321210

Doyle, W. (2006). Ecological approaches to classroom management. In C. Evertson & C. Weinstein (Eds.), *Handbook of classroom management: Research, practice and contemporary issues* (pp. 97–125). Erlbaum.

Dumas, F., Huguet, P., Monteil, J. M., Rastoul, C., & Nezlek, J. (2005). Social comparison in the classroom: Is there a tendency to compare upward in elementary school? *Current Research in Social Psychology, 10*(12), 166–187.

Eccles, J. S. (1983). Expectancies, values, and academic choice: Origins and changes. In J. Spence (Ed.), *Achievement and Achievement Motivation* (pp. 87–134). W. H. Freeman.

Erciyes, K. (2021). *Discrete mathematics and graph theory*. Springer. https://doi.org/10.1007/978-3-030-61115-6

Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., & Yuan, Q. (2018). The Big-Fish-Little-Pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology, 9*, Article 1569. https://doi.org/10.3389/fpsyg.2018.01569

Fauth, B., Atlay, C., Dumont, H., & Decristan, J. (2021). Does what you get depend on who you are with? Effects of student composition on teaching quality. *Learning and Instruction, 71*, Article 101355. https://doi.org/10.1016/j.learninstruc.2020.101355

Fauth, B., Wagner, W., Bertram, C., Göllner, R., Roloff, J., Lüdtke, O., Polikoff, M. S., Klusmann, U., & Trautwein, U. (2020). Don't blame the teacher? The need to account for classroom characteristics in evaluations of teaching quality. *Journal of Educational Psychology, 112*(6), 1284–1302. https://doi.org/10.1037/edu0000416

Fauville, G., Queiroz, A. C. M., Hambrick, L., Brown, B. A., & Bailenson, J. N. (2020). Participatory research on using virtual reality to teach ocean acidification: A study in the marine education community. *Environmental Education Research*, *27*(2), 254–278. https://doi.org/10.1080/13504622.2020.1803797

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Filipp, S.-H. (1979). Entwurf eines heuristischen Bezugsrahmens für Selbstkonzept-Forschung: Menschliche Informationsverarbeitung und naive Handlungstheorie. In S.-H. Filipp (Ed.), *Selbstkonzeptforschung: Probleme, Befunde, Perspektiven* (pp. 129–152). Klett-Cotta.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524793.001.0001

Fiske, S. T. (1993). Social cognition and social perception. *Annual Review of Psychology, 44*(1), 155–194. https://doi.org/10.1146/annurev.ps.44.020193.001103

Folk, C. L. (2015). Controlling spatial attention: Lessons from the lab and implications for everyday life. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 3–25). MIT Press. https://doi.org/10.7551/mitpress/10033.001.0001

Forgas, J. P. (1983). What is social about social cognition? *British Journal of Social Psychology, 22*, 129–144. https://doi.org/10.1111/j.2044-8309.1983.tb00574.x

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology, 21*(3), 95–113. https://doi.org/10.1027/1864-1105.21.3.95

Frank, R. H. (1985). *Choosing the right pond: Human behavior and the quest for status*. Oxford University Press.

Frank, R. H. (2013). *The Darwin economy: Liberty, competition, and the common good*. Princeton University Press.

Frenzel, A. C., Becker-Kurz, B., Pekrun, R., Goetz, T., & Lüdtke, O. (2018). Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology, 110*(5), 628–639. https://doi.org/10.1037/edu0000228

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Perceived learning environment and students' emotional experiences: A multilevel analysis of mathematics classrooms. *Learning and Instruction, 5*, 478–493. https://doi.org/10.1016/j.learninstruc.2007.09.001

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin, 138*(2), 296–321. https://doi.org/10.1037/a0026556

Frey, K. S., & Ruble, D. N. (1990). Strategies for comparative evaluation: Maintaining a sense of competence across the lifespan. In R. J. Sternberg & J. Kolligian Jr. (Eds.), *Competence considered* (pp. 167–189). Yale University Press.

Fridman, N., & Kaminka, G. A. (2011). Towards a computational model of social comparison: Some implications for the cognitive architecture. *Cognitive Systems Research, 12*(2), 186–197. https://doi.org/10.1016/j.cogsys.2010.08.002

Friedrich, A., Flunger, B., Nagengast, B., Jonkmann, K., & Trautwein, U. (2015). Pygmalion effects in the classroom: Teacher expectancy effects on students' math achievement. *Contemporary Educational Psychology, 41*, 1–12. https://doi.org/10.1016/j.cedpsych.2014.10.006

Frisby, B. N. (2019). The influence of emotional contagion on student perceptions of instructor rapport, emotional support, emotion work, valence, and cognitive learning. *Communication Studies, 70*(4), 492–506. https://doi.org/10.1080/10510974.2019.1622584

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology, 74*(2), 435–452. https://doi.org/10.1037/0022-3514.74.2.435

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8*(9), 396–403. https://doi.org/10.1016/j.tics.2004.07.002

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology, 44*(3), 625–636. https://doi.org/10.1037/0012-1649.44.3.625

Gibbons, F. X. (1986). Social comparison and depression: Company's effect on misery. *Journal of Personality and Social Psychology, 51*(1), 140–148. https://doi.org/10.1037/0022-3514.51.1.140

Gibbons, F. X., & Buunk, A. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76*(1), 129–142. https://doi.org/10.1037/0022-3514.76.1.129

Girard, J. M., & Cohn, J. F. (2016). A primer on observational measurement. *Assessment, 23*(4), 404–413. https://doi.org/10.1177/1073191116635807

Goethals, G., & Darley, J. (1977). Social comparison theory: An attributional approach. In J. M. Suls & R. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives*. Halsted/Wiley.

Goethals, G. R. (1986). Social comparison theory: Psychology from the lost and found. *Personality and Social Psychology Bulletin, 12*(3), 261–278. https://doi.org/10.1177/0146167286123001

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics, 24*, 631–645. https://doi.org/10.1016/S0169-8141(98)00068-7

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science, 21*(2), 90–95. https://doi.org/10.1177/0963721412436811

Göllner, R., Damian, R. I., Nagengast, B., Roberts, B. W., & Trautwein, U. (2018). It's not only who you are but who you are with: High school composition and individuals' attainment over the life course. *Psychological Science, 29*(11), 1785–1796. https://doi.org/10.1177/0956797618794454

Gottfried, M. A. (2012). Peer effects in urban schools: Assessing the impact of classroom composition on student achievement. *Educational Policy, 28*(5), 607–647. https://doi.org/10.1177/0895904812467082

Granholm, R. F., Asarnow, A. J., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*, 457–461. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Gruder, C. L. (1977). Choice of comparison persons in evaluating oneself. In J. Suls & R. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives* (pp. 21–42). Halsted/Wiley.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*(1), 124–136. https://doi.org/10.1037/0022-0663.95.1.124

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences, 20*(6), 644–653. https://doi.org/10.1016/j.lindif.2010.08.001

Guillon, Q., Afzali, M. H., Roge, B., Baduel, S., Kruck, J., & Hadjikhani, N. (2015). The importance of networking in autism gaze analysis. *PloS One, 10*(10), Article e0141191. https://doi.org/10.1371/journal.pone.0141191

Gürerk, Ö., Bönsch, A., Kittsteiner, T., & Staffeldt, A. (2019). Virtual humans as co-workers: A novel methodology to study peer effects. *Journal of Behavioral and Experimental Economics, 78*, 17–29. https://doi.org/10.1016/j.socec.2018.11.003

Gutiérrez-Maldonado, J., Letosa-Porta, À., Rus-Calafell, M., & Peñaloza-Salazar, C. (2009). The assessment of attention deficit hyperactivity disorder in children using continous performance tasks in virtual environments. *Anuario de Psicología, 40*(2), 211–222.

Hannover, B. (1997). *Das dynamische Selbst: Die Kontextabhängigkeit selbstbezogenen Wissens*. Huber.

Hardré, P. L., & Sullivan, D. W. (2008). Student differences and environment perceptions: How they contribute to student motivation in rural high schools. *Learning and Individual Differences, 18*(4), 471–485. https://doi.org/10.1016/j.lindif.2007.11.010

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15*(2), 177–199. https://doi.org/10.1076/sesi.15.2.177.30432

Harris, K. R. (2018). Educational psychology: A future retrospective. *Journal of Educational Psychology, 110*(2), 163–173. https://doi.org/10.1037/edu0000267

Harter, S. (1986). Cognitive-developmental processes in the integration of concepts about emotions and the self. *Social Cognition, 4*(2), 119–151. https://doi.org/10.1521/soco.1986.4.2.119

Harter, S. (1998). The development of self-representations. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 553–617). John Wiley & Sons, Inc.

Harter, S. (1999). *The construction of the self: A developmental perspective*. Guilford Press.

Hattie, J. A. C. (1992). *Self-concept*. Lawrence Erlbaum.

Hattie, J. A. C. (2002). Classroom composition and peer effects. *International Journal of Educational Research, 37*(5), 449–481. https://doi.org/10.1016/s0883-0355(03)00015-6

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences, 9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments, 1*(2), 262–271. https://doi.org/10.1162/pres.1992.1.2.262

Heidicker, P., Langbehn, E., & Steinicke, F. (2017). Influence of avatar appearance on presence in social VR. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE. https://doi.org/10.1109/3DUI.2017.7893357

Helgeson, V. S., & Mickelson, K. D. (1995). Motives for social comparison. *Personality and Social Psychology Bulletin, 21*(11), 1200–1209. https://doi.org/10.1177/01461672952111008

Helmke, A. (1992). *Selbstvertrauen und schulische Leistungen*. Hogrefe.

Henderson, R. R., Bradley, M. M., & Lang, P. J. (2018). Emotional imagery and pupil diameter. *Psychophysiology, 55*(6), Article e13050. https://doi.org/10.1111/psyp.13050

Herrera, F., Oh, S. Y., & Bailenson, J. N. (2020). Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence, 27*(2), 163–182. https://doi.org/10.1162/pres_a_00324

Herrmann, J., Schmidt, I., Kessels, U., & Preckel, F. (2016). Big fish in big ponds: Contrast and assimilation effects on math and verbal self-concepts of students in within-school gifted tracks. *British Journal of Educational Psychology, 86*(2), 222–240. https://doi.org/10.1111/bjep.12100

Higgins, F. T., & Bargh, J. A. (1987). Social cognition and social perception. *Annual Review of Psychology, 38*(1), 369–425. https://doi.org/10.1146/annurev.ps.38.020187.002101

Hirt, C., Eckard, M., & Kunz, A. (2020). Stress generation and non-intrusive measurement in virtual environments using eye tracking. *Journal of Ambient Intelligence and Humanized Computing, 11*, 5977–5989. https://doi.org/10.1007/s12652-020-01845-y

Hite, R. L., Jones, M. G., Childers, G. M., Ennes, M., Chesnutt, K., Pereyra, M., & Cayton, E. (2019). Investigating potential relationships between adolescents' cognitive development and perceptions of presence in 3-D, haptic-enabled, virtual reality science instruction. *Journal of Science Education and Technology, 28*(3), 265–284. https://doi.org/10.1007/s10956-018-9764-y

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior, 26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Horvath, K., & Lombard, M. (2010). Social and spatial presence: An application to optimize human-computer interaction. *PsychNology Journal, 8*(1), 85–114.

Howard, M. C., Gutworth, M. B., & Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior, 121*, Article 106808. https://doi.org/https://doi.org/10.1016/j.chb.2021.106808

Hoyt, C. L., Blascovich, J., & Swinth, K. R. (2003). Social inhibition in immersive virtual environments. *Presence, 12*(2), 183–195. https://doi.org/10.1162/105474603321640932

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology, 49*(5), 505–528. https://doi.org/10.1016/j.jsp.2011.07.001

Huang, S. A., & Bailenson, J. (2019). Close relationships and virtual reality. In T. D. Parsons, L. Lin, & D. Cockerham (Eds.), *Mind, Brain and Technology* (pp. 49–65). Springer. https://doi.org/10.1007/978-3-030-02631-8_4

Hudson, I., & Hurter, J. (2016). Avatar types matter: Review of avatar literature for performance purposes. In S. Lackey & R. Shumaker (Eds.), *VAMR 2016: International conference on virtual, augmented and mixed reality* (pp. 14–21). Springer. https://doi.org/10.1007/978-3-319-39907-2_2

Huguet, P., Dumas, F., Marsh, H. W., Wheeler, L., Seaton, M., Nezlek, J., Suls, J., & Régner, I. (2009). Clarifying the role of social comparison in the Big-Fish-Little-Pond effect (BFLPE): An integrative study. *Journal of Personality and Social Psychology, 97*(1), 156–170. https://doi.org/10.1037/a0015558

Huguet, P., Dumas, F., Monteil, J. M., & Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its benficial impact on performance. *European Journal of Social Psychology*, *31*, 557–578. https://doi.org/10.1002/ejsp.81

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology, 10*, Article 2246. https://doi.org/10.3389/fpsyg.2019.02246

Hyman, H. H. (1942). The psychology of status. *Archives of Psychology (Columbia University), 269*, 94–102.

Ilies, R., & Judge, T. A. (2005). Goal regulation across time: The effects of feedback and affect. *Journal of Applied Psychology, 90*(3), 453–467. https://doi.org/10.1037/0021-9010.90.3.453

International Society for Presence Research. (2000). *The concept of presence: Explication statement*. https://smcsites.com/ispr/

Irvine, E. (2021). The role of replication studies in theory building. *Perspectives on Psychological Science, 16*(4), 844–853. https://doi.org/10.1177/1745691620970558

Jainta, S., & Baccino, T. (2010). Analyzing the pupil response due to increased cognitive demand: an independent component analysis study. *International Journal of Psychophysiology*, *77*(1), 1–7. https://doi.org/10.1016/j.ijpsycho.2010.03.008

James, W. (1892/1999). The self. In R. F. Baumeister (Ed.), *The Self in Social Psychology* (pp. 69–77). Psychology Press.

Jarodzka, H., Homqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agenda. *Journal of Eye Movement Research, 10*(1), 1–18. https://doi.org/10.16910/jemr.10.1.3

Jarodzka, H., Skuballa, I., & Gruber, H. (2021). Eye-tracking in educational practice: Investigating visual perception underlying teaching and learning in the classroom. *Educational Psychology Review, 33*, 1–10. https://doi.org/10.1007/s10648-020-09565-7

Jo, D., Kim, K.-H., & Kim, G. J. (2016). Effects of avatar and background representation forms to co-presence in mixed reality (MR) tele-conference systems. *SIGGRAPH ASIA 2016 Virtual reality meets physical reality: Modelling and simulating virtual humans and environments*, Article 12. https://doi.org/10.1145/2992138.2992146

Johnston, E., Olivas, G., Steele, P., Smith, C., & Bailey, L. (2017). Exploring pedagogical foundations of existing virtual reality educational applications: A content analysis study. *Journal of Educational Technology Systems, 46*(4), 414–439. https://doi.org/10.1177/0047239517745560

Jonkmann, K., Becker, M., Marsh, H. W., Lüdtke, O., & Trautwein, U. (2012). Personality traits moderate the Big-Fish-Little-Pond effect of academic self-concept. *Learning and Individual Differences, 22*(6), 736–746. https://doi.org/10.1016/j.lindif.2012.07.020

Joy, T., Ugur, E., & Ayhan, I. (2021). Trick the body trick the mind: Avatar representation affects the perception of available action possibilities in virtual reality. *Virtual Reality*, Online first publication. https://doi.org/10.1007/s10055-021-00511-8

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kaakinen, J. K. (2021). What can eye movements tell us about visual perception processes in classroom contexts? Commentary on a special issue. *Educational Psychology Review, 33*, 169–179. https://doi.org/10.1007/s10648-020-09573-7

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*(2), 136–153. https://doi.org/10.1037/0033-295X.93.2.136

Karutz, C. O., & Bailenson, J. (2015). Immersive virtual environments and the classrooms of tomorrow. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 290–310). John Wiley & Sons. https://doi.org/10.1002/9781118426456.ch13

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. *Neuroscientist, 20*(5), 509–521. https://doi.org/10.1177/1073858413514136

Kedia, G., Mussweiler, T., & Linden, D. E. (2014). Brain mechanisms of social comparison and their influence on the reward system. *Neuroreport, 25*(16), 1255–1265. https://doi.org/10.1097/WNR.0000000000000255

Kenrick, D. T., Li, N. P., & Butner, J. (2003). Dynamical evolutionary psychology: Individual decision rules and emergent social norms. *Psychological Review, 110*(1), 3–28. https://doi.org/10.1037/0033-295x.110.1.3

Kessels, U., & Hannover, B. (2004). Empfundene “Selbstnähe“ als Mediator zwischen Fähigkeitsselbstkonzept und Leistungskurswahlintentionen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36*(3), 130–138. https://doi.org/10.1026/0049-8637.36.3.130

Kiviniemi, M. T., Snyder, M., & Johnson, B. C. (2008). Motivated dimension manipulation in the processing of social comparison information. *Self and Identity, 7*(3), 225–242. https://doi.org/10.1080/15298860701275873

Klein, W. M. (1997). Objective standards are not enough: Affective, self-evaluative, and behavioral responses to social comparison information. *Journal of Personality and Social Psychology, 72*(4), 763–774. https://doi.org/10.1037/0022-3514.72.4.763

Klein, W. M. (2003). Effects of objective feedback and "single other" or "average other" social comparison feedback on performance judgments and helping behavior. *Personality and Social Psychology Bulletin, 29*(3), 418–429. https://doi.org/10.1177/0146167203251195

Köller, O., Daniels, Z., Schnabel, K. U., & Baumert, J. (2000). Kurswahlen von Mädchen und Jungen im Fach Mathematik: Zur Rolle von fachspezifischem Selbstkonzept und Interesse. *Zeitschrift für Pädagogische Psychologie, 14*(1), 26–37. https://doi.org/10.1024//1010-0652.14.1.26

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research, 51*(13), 1457–1483. https://doi.org/10.1016/j.visres.2010.12.014

Krämer, N. C. (2017). The immersive power of social interaction. In D. Liu, C. Dede, R. Huang, & J. Richards (Eds.), *Virtual, augmented, and mixed realities in education* (pp. 55–70). Springer. https://doi.org/10.1007/978-981-10-5490-7

Kübler, T. C., Rothe, C., Schiefer, U., Rosenstiel, W., & Kasneci, E. (2017). SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies. *Behavior Research Methods, 49*(3), 1048–1064. https://doi.org/10.3758/s13428-016-0765-6

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10,* 90–115. https://doi.org/10.1016/j.edurev.2013.10.001

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological Bulletin, 136*(6), 1045–1067. https://doi.org/10.1037/a0020970

Lanier, J. (1989). Virtual environments and interactivity: Windows to the future. In C. Conn, J. Lanier, M. Minsky, S. Fisher, & A. Druin (Eds.), *ACM SIGGRAPH Computer Graphics Vol. 23 Panel Proceedings* (pp. 7–10). ACM. https://doi.org/10.1145/77277.77278

Lavy, V., Paserman, M. D., & Schlosser, A. (2011). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal, 122,* 208–237. https://doi.org/10.1111/j.1468-0297.2011.02463.x

Le Meur, O., & Baccino, T. (2013, Mar). Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods, 45*(1), 251–266. https://doi.org/10.3758/s13428-012-0226-9

Lee, K. M. (2004). Presence, explicated. *Communication Theory, 14*(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Leflot, G., Onghena, P., & Colpin, H. (2010). Teacher-child interactions: Relations with children's self-concept in second grade. *Infant and Child Development, 19,* 385–405. https://doi.org/10.1002/icd.672

Liao, M. Y., Sunq, C. Y., Wang, H. C., & Lin, W. C. (2019). Virtual classmates: Embodying historical learners' messages as learning companions in a VR classroom through comment mapping. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 163–171). IEEE. https://doi.org/10.1109/VR.2019.8797708

Lieberman, M. D. (2005). Principles, processes, and puzzles of social cognition: an introduction for the special issue on social cognitive neuroscience. *Neuroimage, 28*(4), 745–756. https://doi.org/10.1016/j.neuroimage.2005.07.028

Lim, S., & Reeves, B. (2009). Being in the game: Effects of avatar choice and point of view on psychophysiological responses during play. *Media Psychology, 12*(4), 348–370. https://doi.org/10.1080/15213260903287242

Lodge, J., & Harrison, W. (2019). The role of attention in learning in the digital age. *The Yale Journal of Biology and Medicine, 92*(1), 21–28.

Lohbeck, A., & Möller, J. (2017). Social and dimensional comparison effects on math and reading self-concepts of elementary school children. *Learning and Individual Differences, 54*, 73–81. https://doi.org/10.1016/j.lindif.2017.01.013

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2). https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). Measuring telepresence: The temple presence inventory. In *12th Annual International Workshop on Presence* (pp. 1–15). International Society for Presence Research.

Lou, Y., Cai, H., Liu, X., & Li, X. (2019). Effects of self-enhancement on eye movements during reading. *Frontiers in Psychology, 10*, Article 343. https://doi.org/10.3389/fpsyg.2019.00343

Loyalka, P., Zakharov, A., & Kuzmina, Y. (2018). Catching the big fish in the little pond effect: Evidence from 33 countries and regions. *Comparative Education Review, 62*(4), 542–564. https://doi.org/10.1086/699672

Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the Big-Fish-Little-Pond effect. *Contemporary Educational Psychology, 30*(3), 263–285. https://doi.org/10.1016/j.cedpsych.2004.10.002

Lüdtke, O., & Robitzsch, A. (2020). Methodological challenges in assessing the causal effects of teaching. *Zeitschrift für Pädagogik, 66*(1), 210–222. https://doi.org/10.31234/osf.io/bpk4a

Lyubomirsky, S., & Ross, L. (1997). Hedonic consequences of social comparison: A contrast of happy and unhappy people. *Journal of Personality and Social Psychology, 73*(6), 1141–1157. https://doi.org/10.1037/0022-3514.73.6.1141

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 long symposium: Toward social mechanisms of android science* (pp. 25–29).

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior, 25*(3), 695–710. https://doi.org/10.1016/j.chb.2008.12.026

Maier, S. U., & Grueschow, M. (2021). Pupil dilation predicts individual self-regulation success across domains. *Nature Scientific Reports, 11*(1), Article 14342. https://doi.org/10.1038/s41598-021-93121-y

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2020). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology, 113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Makransky, G., Wismer, P., & Mayer, R. E. (2019). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning, 35*(3), 349–358. https://doi.org/10.1111/jcal.12335

Mangalmurti, A., Kistler, W. D., Quarrie, B., Sharp, W., Persky, S., & Shaw, P. (2020). Using virtual reality to define the mechanisms linking symptoms with cognitive deficits in attention deficit hyperactivity disorder. *Nature Scientific Reports, 10*, Article 529. https://doi.org/10.1038/s41598-019-56936-4

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Markus, H., & Kunda, Z. (1986). Stability and malleability of self-concept. *Journal of Personality and Social Psychology, 51*(4), 858–866. https://doi.org/10.1037/0022-3514.51.4.858

Markus, H., & Wurf, E. (1987). The dymamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38*(1), 299–337. https://doi.org/10.1146/annurev.ps.38.020187.001503

Marsh, H., & Parker, P. D. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47*(1), 213–231. https://doi.org/10.1037/0022-3514.47.1.213

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education*, *28*(2), 165–181. https://doi.org/10.1177/000494418402800207

Marsh, H. W. (1987). The Big-Fish-Little-Pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280

Marsh, H. W. (1990a). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82*(4), 646–656. https://doi.org/10.1037/0022-0663.82.4.646

Marsh, H. W. (1990b). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*(2), 77–172. https://doi.org/10.1007/BF01322177

Marsh, H. W. (1990c). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology, 82*(4), 623–636. https://doi.org/10.1037/0022-0663.82.4.623

Marsh, H. W. (1992). *Self-description questionnaire (SDQ) III: A theoretical and empirical basis for the measurement of multiple dimensions of late adolescent self-concept: A test manual and a research monograph*. University of Western Sydney.

Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. D. Phye (Ed.), *Handbook of classroom assessment. Learning, achievement, and adjustment* (pp. 131–198). Academic Press. https://doi.org/10.1016/B978-012554155-8/50008-9

Marsh, H. W., Craven, R., & Debus, R. L. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology, 83*(3), 377–192. https://doi.org/10.1037/0022-0663.83.3.377

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. *Perspectives on Psychological Science, 1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., & Hau, K. T. (2003). Big-Fish-Little-Pond effect on academic self-concept. A cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist, 58*(5), 364–376. https://doi.org/10.1037/0003-066x.58.5.364

Marsh, H. W., Köller, O., & Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the Big-Fish-Little-Pond effect on academic self-concept. *American Educational Research Journal, 38*(2), 321–350. https://doi.org/10.3102/00028312038002321

Marsh, H. W., Kong, C. K., & Hau, K. T. (2000). Longitudinal multilevel models of the Big-Fish-Little-Pond effect on academic self-concept: Counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, *78*(2), 337–349. https://doi.org/10.1037//0022-3514.78.2.337

Marsh, H. W., Kuyper, H., Morin, A. J. S., Parker, P. D., & Seaton, M. (2014). Big-Fish-Little-Pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, *33*, 50–66. https://doi.org/10.1016/j.learninstruc.2014.04.002

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J. S., Abduljabbar, A. S., & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106–124. https://doi.org/10.1080/00461520.2012.670488

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*, 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Martin, A. J., Yeung, A. S., & Craven, R. (2017). Competence self-perceptions. In A. J. Elliot, C. Dweck, & D. Yeager (Eds.), *Handbook of competence and motivation* (2nd ed., pp. 85–115). Guilford Press.

Marsh, H. W., Parker, P. D., & Pekrun, R. (2019). Three paradoxical effects on academic self-concept across countries, schools, and students. *European Psychologist*, *24*(3), 231–242. https://doi.org/10.1027/1016-9040/a000332

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2019). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology, 111*(2), 331–353. https://doi.org/10.1037/edu0000281

Marsh, H. W., & Seaton, M. (2015). The Big-Fish-Little-Pond effect, competence self-perceptions, and relativity: Substantive advances and methodological innovation. *Advances in Motivation Science, 2*, 127–184. https://doi.org/10.1016/bs.adms.2015.05.002

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The Big-Fish-Little-Pond effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*(3), 319–350. https://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1

Marsh, H. W., Trautwein, U., Lüdtke, O., & Köller, O. (2008). Social comparison and Big-Fish-Little-Pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*(3), 510–524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., Xu, K. M., Parker, P. D., Hau, K.-T., Pekrun, R., Elliot, A., Guo, J., Dicke, T., & Basarkod, G. (2021). Moderation of the Big-Fish-Little-Pond effect: Juxtaposition of evolutionary (Darwinian-economic) and achievement motivation theory predictions based on a Delphi approach. *Educational Psychology Review*, Advance online publication. https://doi.org/10.1007/s10648-020-09583-5

Marsh, H. W., & Yeung, A. S. (1997). Coursework selection: Relations to academic self-concept and achievement. *American Educational Research Journal, 34*(4), 691–720. https://doi.org/10.3102/00028312034004691

Massaro, D. W., & Anderson, N. H. (1971). Judgemental model of the Ebbinghaus illusion. *Journal of Experimental Psychology, 89*(1), 147–151. https://doi.org/10.1037/h0031158

Mathur, M. B., & Reichling, D. B. (2016). Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition, 146*, 22–32. https://doi.org/10.1016/j.cognition.2015.09.008

Mead, G. H. (1934). Mind, self, and society. In J. Margolis & J. Catudal (Eds.), *The quarrel between invariance and flux. A guide for philosophers and other players* (pp. 223–234). Chicago University Press.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*(1), 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Michinov, N., Jamet, E., Métayer, N., & Le Hénaff, B. (2015). The eyes of creativity: Impact of social comparison and individual creativity on performance and attention to others' ideas during electronic brainstorming. *Computers in Human Behavior*, *42*, 57–67. https://doi.org/10.1016/j.chb.2014.04.037

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999-2009). *Computers & Education*, *56*(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371–378. https://doi.org/10.1037/h0040525

Miller, M. R., Jun, H., & Bailenson, J. (2021). Motion and meaning: Sample-level nonlinear analyses of virtual reality tracking data. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE. https://doi.org/10.1109/54149.2021.00039

Mischel, W., & Morf, C. C. (2003). The self as a psycho-social dynamic processing system: A meta-perspective on a century of the self in psychology. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 15–43). The Guilford Press.

Mitchell, M. A., & Schmidt, N. B. (2014). An experimental manipulation of social comparison in social anxiety. *Cognitive Behaviour Therapy, 43*(3), 221–229. https://doi.org/10.1080/16506073.2014.914078

Möller, J., & Köller, O. (2001). Dimensional comparisons: An experimental approach to the internal/external frame of reference model. *Journal of Educational Psychology, 93*(4), 826–835. https://doi.org/10.1037/0022-0663.93.4.826

Möller, J., & Marsh, H. W. (2013). Dimensional comparison theory. *Psychological Review*, *120*(3), 544–560. https://doi.org/10.1037/a0032459

Möller, J., Retelsdorf, J., Köller, O., & Marsh, H. W. (2011). The reciprocal internal/external frame of reference model. *American Educational Research Journal, 48*(6), 1315–1346. https://doi.org/10.3102/0002831211419649

Möller, J., Zimmermann, F., & Köller, O. (2014). The reciprocal internal/external frame of reference model using grades and test scores. *British Journal of Educational Psychology, 84*(4), 591–611. https://doi.org/10.1111/bjep.12047

Moreau, G., Guay, M. C., Achim, A., Rizzo, A., & Lageix, P. (2006). The virtual classroom: An ecological version of the continuous performance test – A pilot study. In B. K. Wiederhold & S. Bouchard (Eds.), *Annual review of cybertherapy and telemedicine* (Vol. 4, pp. 59–66). Interactive Media Institute.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Morina, N. (2021). Comparisons inform me who I am: A general comparative-processing model of self-perception. *Perspectives on Psychological Science*, Online first publication. https://doi.org/10.1177/1745691620966788

Moro, C., Stromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education, 10*(6), 549–559. https://doi.org/10.1002/ase.1696

Mudrick, N. V., Azevedo, R., & Taub, M. (2019). Integrating metacognitive judgments and eye movements using sequential pattern mining to understand processes underlying multimedia learning. *Computers in Human Behavior, 96*, 223–234. https://doi.org/10.1016/j.chb.2018.06.028

Muller, D., & Butera, F. (2007). The focusing effect of self-evaluation threat in coaction and social comparison. *Journal of Personality and Social Psychology, 93*(2), 194–211. https://doi.org/10.1037/0022-3514.93.2.194

Mussweiler, T. (2001a). Focus of comparison as a determinant of assimilation versus contrast in social comparison. *Personality and Social Psychology Bulletin, 27*(1), 38–47. https://doi.org/10.1177/0146167201271004

Mussweiler, T. (2001b). 'Seek and ye shall find': Antecedents of assimilation and contrast in social comparison. *European Journal of Social Psychology, 31*(5), 499–509. https://doi.org/10.1002/ejsp.75

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110*(3), 472–489. https://doi.org/10.1037/0033-295X.110.3.472

Mussweiler, T., Rüter, K., & Epstude, K. (2004). The ups and downs of social comparison: Mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology, 87*(6), 832–844. https://doi.org/10.1037/0022-3514.87.6.832

Mussweiler, T., & Strack, F. (2000). The "relative self": Informational and judgmental consequences of comparative self-evaluation. *Journal of Personality and Social Psychology, 79*(1), 23–38. https://doi.org/10.1037/0022-3514.79.1.23

Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology, 104*(4), 1033–1053. https://doi.org/10.1037/a0027697

Nash, R. (2003). Is the school composition effect real?: A discussion with evidence from the UK PISA data. *School Effectiveness and School Improvement*, *14*(4), 441–457. https://doi.org/10.1076/sesi.14.4.441.17153

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues, 56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice Hall.

Nolin, P., Stipanicic, A., Henry, M., Lachapelle, Y., Lussier-Desrochers, D., Rizzo, A. S., & Allain, P. (2016). ClinicaVR: Classroom-CPT: A virtual reality tool for assessing attention and inhibition in children and adolescents. *Computers in Human Behavior, 59*, 327–333. https://doi.org/10.1016/j.chb.2016.02.023

Nowak, A., Vallacher, R. R., Tesser, A., & Borkowski, W. (2000). Society of self: The emergence of collective properties in self-structure. *Psychological Review, 107*(1), 39–61. https://doi.org/10.1037/0033-295X.107.1.39

O'Connell, A. N. (1980). Effects of manipulated status on performance, goal setting, achievement motivation, anxiety, and fear of success. *The Journal of Social Psychology, 112*(1), 75–89. https://doi.org/10.1080/00224545.1980.9924299

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41*(3), 181–206. https://doi.org/10.1207/s15326985ep4103

Ogawa, N., Narumi, T., & Hirose, M. (2018). Object size perception in immersive virtual reality: Avatar realism affects the way we perceive. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 647–648). IEEE. https://doi.org/10.1109/VR.2018.8446318

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI, 5*, Article 114. https://doi.org/10.3389/frobt.2018.00114

Olney, A. M., Risko, E. F., D'Mello, S. K., & Graesser, A. C. (2015). Attention in educational contexts: The role of the learning task in guiding attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 623–641). MIT Press. https://doi.org/10.7551/mitpress/10033.001.0001

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*(10), 1346–1353. https://doi.org/10.1111/j.1469-8986.2011.01217.x

Oyserman, D., & Lee, S. W. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological Bulletin, 134*(2), 311–342. https://doi.org/10.1037/0033-2909.134.2.311

Parker, P. D., Marsh, H. W., Jerrim, J. P., Guo, J., & Dicke, T. (2018). Inequity and excellence in academic performance: Evidence from 27 countries. *American Educational Research Journal, 55*(4), 836–858. https://doi.org/10.3102/0002831218760213

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*, Article 660. https://doi.org/10.3389/fnhum.2015.00660

Parsons, T. D., Bowerly, T., Buckwalter, J. G., & Rizzo, A. A. (2007). A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology*, *13*(4), 363–381. https://doi.org/10.1080/13825580600943473

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59,* 185–198. https://doi.org/10.1016/s1071-5819(03)00017-x

Paxton, A., & Griffiths, T. L. (2017). Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods, 49*(5), 1630–1638. https://doi.org/10.3758/s13428-017-0874-x

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., & Frenzel, A. C. (2019). Happy fish in little ponds: Testing a reference group model of achievement and emotion. *Journal of Personality and Social Psychology, 117*(1), 166–185. https://doi.org/10.1037/pspp0000230

Pelham, B. W., & Wachsmuth, J. O. (1995). The waxing and waning of the social self: Assimilation and contrast in social comparison. *Journal of Personality and Social Psychology, 69*(5), 825–838. https://doi.org/10.1037/0022-3514.69.5.825

Piaget, J. (1960). *The psychology of intelligence*. Littlefield, Adams.

Plieninger, H., & Dickhäuser, O. (2013). The female fish is more responsive: Gender moderates the BFLPE in the domain of science. *Educational Psychology, 35*(2), 213–227. https://doi.org/10.1080/01443410.2013.814197

Poh, M. Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering, 57*(5), 1243–1252. https://doi.org/10.1109/TBME.2009.2038487

Pollak, Y., Shomaly, H. B., Weiss, P. L., Rizzo, A. A., & Gross-Tsur, V. (2010). Methylphenidate effect in children with ADHD can be measured by an ecologically valid continuous performance test embedded in virtual reality. *CNS Spectrum*, *15*(2), 125–130. https://doi.org/10.1017/s109285290002736x

Pollak, Y., Weiss, P. L., Rizzo, A. A., Weizer, M., Shriki, L., Shalev, R. S., & Gross-Tsur, V. (2009). The utility of a continuous performance test embedded in virtual reality in measuring ADHD-related deficits. *Journal of Developmental and Behavioral Pediatrics, 30*(1), 2–6. https://doi.org/10.1097/DBP.0b013e3181969b22

Posner, M. I. (1988). Structures and functions of selective attention. In T. Boll & B. K. Bryant (Eds.), *Clinical neuropsychology and brain function: Research, measurement, and practice* (pp. 173–202). APA. https://doi.org/10.1037/10063-005

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78*(5), 391–408. https://doi.org/10.1037/h0031333

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology, 58*, 1–23. https://doi.org/10.1146/annurev.psych.58.110405.085516

Preckel, F., & Brüll, M. (2008). Grouping the gifted and talented: Are gifted girls most likely to suffer the consequences? *Journal for the Education of the Gifted, 32*(1), 54–85. https://doi.org/10.4219/jeg-2008-822

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision, 10*(10), 1–14. https://doi.org/10.1167/10.10.3

Pyszczynski, T., Greenberg, J., & LaPrelle, J. (1985). Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. *Journal of Experimental Social Psychology, 29*(1), 195–211. https://doi.org/10.1016/0022-1031(85)90015-0

Queiroz, A. C. M., Nascimento, A. M., Tori, R., & Da Silva Leme, M. I. (2018). Using HMD-based immersive virtual environments in primary/K-12 education. In D. Beck, C. A. L. Morgado, J. Pirker, A. Peña-Rios, T. Ogl, J. Richter, & C. Gütl (Eds.), *Immersive Learning Research Network: 4th International Conference iLRN 2018* (pp. 160–173). Springer. https://doi.org/10.1007/978-3-319-93596-6

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, Article 103778. https://doi.org/10.1016/j.compedu.2019.103778

Ratan, R., Beyea, D., Li, B. J., & Graciano, L. (2020). Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the proteus effect. *Media Psychology, 23*(5), 651–675. https://doi.org/10.1080/15213269.2019.1623698

Ray, D. G., Neugebauer, J., & Sassenberg, K. (2017). Learners' habitual social comparisons can hinder effective learning partner choice. *Learning and Individual Differences, 58*, 83–89. https://doi.org/10.1016/j.lindif.2017.08.003

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Reichle, E. D., Pollatsek, A., & Rayner, K. (2012). Using E-Z reader to simulate eye movements in nonreading tasks: A unified framework for understanding the eye-mind link. *Psychological Review, 119*(1), 155–185. https://doi.org/10.1037/a0026473

Richardson, D. C., & Gobel, M. S. (2015). Social attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 349–367). MIT Press. https://doi.org/10.7551/mitpress/10033.001.0001

Rijsman, J. B. (1974). Factors in social comparison of performance influencing actual performance. *European Journal of Social Psychology, 4*(3), 279–311. https://doi.org/10.1002/ejsp.2420040303

Rizzo, A. A., Bowerly, T., Buckwalter, J. G., Klimchuk, D., Mitura, R., & Parsons, T. D. (2006). A virtual reality scenario for all seasons: The virtual classroom. *CNS Spectrum, 11*(1), 35–44. https://doi.org/10.1017/S1092852900024196

Rizzo, A. A., Buckwalter, J. G., Bowerly, T., Van Der Zaag, C., Humphrey, L., Neumann, U., Chua, C., Kyriakakis, C., Van Rooyen, A., & Sisemore, D. (2000). The virtual classroom: A virtual reality environment for the assessment and rehabilitation of attention deficits. *Cyberpsychology & Behavior*, *3*(3), 483–499. https://doi.org/10.1089/10949310050078940

Roy, A., Guay, F., & Valois, P. (2015). The Big-Fish-Little-Pond effect on academic self-concept: The moderating role of differentiated instruction and individual achievement. *Learning and Individual Differences*, *42*, 110–116. https://doi.org/10.1016/j.lindif.2015.07.009

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. (1980). Developmental analysis of the role of social comparison in self-evaluation. *Developmental Psychology, 16*(2), 105–115. https://doi.org/10.1037/0012-1649.16.2.105

Rullo, M., Presaghi, F., & Livi, S. (2015). Reactions to ingroup and outgroup deviants: An experimental group paradigm for black sheep effect. *PloS One, 10*(5), Article e0125605. https://doi.org/10.1371/journal.pone.0125605

Sadria, M., Karimi, S., & Layton, A. T. (2019). Network centrality analysis of eye-gaze data in autism spectrum disorder. *Computers in Biology and Medicine, 111*, Article 103332. https://doi.org/10.1016/j.compbiomed.2019.103332

Salchegger, S. (2016). Selective school systems and academic self-concept: How explicit and implicit school-level tracking relate to the Big-Fish-Little-Pond effect across cultures. *Journal of Educational Psychology*, *108*(3), 405–423. https://doi.org/10.1037/edu0000063

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *ETRA '00: Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (pp. 71–78). ACM. https://doi.org/10.1145/355017.355028

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews: Neuroscience*, *6*(4), 332–339. https://doi.org/10.1038/nrn1651

Scheepers, D. (2009). Turning social identity threat into challenge: Status stability and cardiovascular reactivity during inter-group competition. *Journal of Experimental Social Psychology, 45*(1), 228–233. https://doi.org/10.1016/j.jesp.2008.09.011

Scheepers, D., & Ellemers, N. (2005). When the pressure is up: The assessment of social identity threat in low and high status groups. *Journal of Experimental Social Psychology*, *41*(2), 192–200. https://doi.org/10.1016/j.jesp.2004.06.002

Scheiter, K., & Eitel, A. (2017). The use of eye tracking as a research and instructional tool in multimedia learning. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-tracking technology applications in educational research* (pp. 143–164). IGI Global. https://doi.org/10.4018/978-1-5225-1005-5.ch008

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Consciousness and Cognition, 17*(2), 457–467. https://doi.org/10.1016/j.concog.2008.03.013

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, *44*(5), 718–730. https://doi.org/10.1016/j.neuropsychologia.2005.07.017

Schmitz, A., MacQuarrie, A., Julier, S., Binetti, N., & Steed, A. (2020). Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 63–72). IEEE. https://doi.org/10.1109/VR46266.2020.00024

Schneider, B., Abu-El-Haija, S., Reesman, J., & Pea, R. (2013). Toward collaboration sensing: applying network analysis techniques to collaborative eye-tracking data. In *LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 107–111). ACM. https://doi.org/10.1145/2460296.2460317

Schnotz, W. (2016). Learning and Instruction: A review of main research lines during recent decades. *Zeitschrift für Erziehungswissenschaft, 19*(1), 101–119. https://doi.org/10.1007/s11618-015-0663-1

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence, 10*(3), 266–281. https://doi.org/10.1162/105474601300343603

Schuemie, M. J., Van der Straaten, P., Krijn, M., & Van der Mast, C. A. P. G. (2001). Research on presence in virtual reality: A survey. *Cyberpsychology & Behavior, 4*(2), 183–201. https://doi.org/10.1089/109493101300117884

Schunk, D. H., & Pajares, F. (2005). Competence perceptions and academic functioning. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85–104). Guilford Press.

Schwabe, F., Korthals, R., & Schils, T. (2019). Positive social relationships with peers and teachers as moderators of the Big-Fish-Little-Pond effect. *Learning and Individual Differences, 70*, 21–29. https://doi.org/10.1016/j.lindif.2018.12.006

Schwanzer, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica*, *51*(4), 183–194. https://doi.org/10.1026/0012-1924.51.4.183

Seaton, M., Marsh, H., Dumas, F., Huguet, P., Monteil, J. M., Regner, I., Blanton, H., Buunk, A. P., Gibbons, F. X., Kuyper, H., Suls, J. M., & Wheeler, L. (2008). In search of the big fish: Investigating the coexistence of the Big-Fish-Little-Pond effect with the positive effects of upward comparisons. *British Journal of Social Psychology, 47*, 73–103. https://doi.org/10.1348/014466607X202309

Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the Big-Fish-Little-Pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology, 101*(2), 403–419. https://doi.org/10.1037/a0013838

Seaton, M., Marsh, H. W., & Craven, R. G. (2010). Big-Fish-Little-Pond effect: Generalizability and moderation – two sides of the same coin. *American Educational Research Journal, 47*(2), 390–433. https://doi.org/10.3102/0002831209350493

Seaton, M., Marsh, H. W., Yeung, A. S., & Craven, R. (2011). The big fish down under: Examining moderators of the 'Big-Fish-Little-Pond' effect for Australia's high achievers. *Australian Journal of Education*, *55*(2), 93–114. https://doi.org/10.1177/000494411105500202

Sedikides, C., & Strube, M. J. (1997). Self evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 29, pp. 209–269). Academic Press. https://doi.org/10.1016/S0065-2601(08)60018-0

Seidel, R. J., & Chatelier, P. R. (2013). *Virtual reality, training's future? Perspectives on virtual reality and related emerging technologies*. Springer. https://doi.org/10.1007/978-1-4899-0038-8

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, *55*(1), 5–14. https://doi.org/10.1037/0003-066X.55.1.5

Sharar, S. R., Carrougher, G. J., Nakamura, D., Hoffman, H. G., Blough, D. K., & Patterson, D. R. (2007). Factors influencing the efficacy of virtual reality distraction analgesia during postburn physical therapy: Preliminary results from 3 ongoing studies. *Archives of Physical Medicine and Rehabilitation*, *88*(2), 43–49. https://doi.org/10.1016/j.apmr.2007.09.004

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441. https://doi.org/10.2307/1170010

Sherman, S. J., Judd, C. M., & Park, B. (1989). Social cognition. *Annual Review of Psychology, 40*(1), 281–326. https://doi.org/10.1146/annurev.ps.40.020189.001433

Skinner, B. F. (1985). Cognitive science and behaviourism. *British Journal of Psychology, 76*(3), 291–301. https://doi.org/10.1111/j.2044-8295.1985.tb01953.x

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence, 8*(5), 560–565. https://doi.org/10.1162/105474699566477

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B, 364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS One, 1,* Article e39. https://doi.org/10.1371/journal.pone.0000039

Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3,* Article 74. https://doi.org/10.3389/frobt.2016.00074

Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence, 9*(5), 413–434. https://doi.org/10.1162/105474600566925

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence, 6*(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin, 132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Smith, E. R., & Semin, G. R. (2007). Situated social cognition. *Current Directions in Psychological Science, 16*(3), 132–135. https://doi.org/10.1111/j.1467-8721.2007.00490.x

Smith, R. H., & Insko, C. A. (1987). Social comparison choice during ability evaluation: The effects of comparison publicity, performance feedback, and self-esteem. *Personality and Social Psychology Bulletin*, *13*(1), 111–122. https://doi.org/10.1177/0146167287131011

Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L. (2021). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human–Computer Interaction*, Article 23. https://doi.org/10.1080/10447318.2021.1976509

Southgate, E., Smith, S. P., & Scevak, J. (2017). Asking ethical questions in research using immersive virtual and augmented reality technologies with children and youth. In *2017 IEEE Virtual Reality (VR)* (pp. 12–18). IEEE. https://doi.org/10.1109/VR.2017.7892226

Stanton, A. L., Danoff-Burg, S., Cameron, C. L., Snider, P. R., & Kirk, S. B. (1999). Social comparison and adjustment to breast cancer: An experimental examination of upward affiliation and downward evaluation. *Health Psychology, 18*(2), 151–158. https://doi.org/10.1037/0278-6133.18.2.151

Stavropoulos, V., Wilson, P., Kuss, D., Griffiths, M., & Gentile, D. (2017). A multilevel longitudinal study of experiencing virtual presence in adolescence: the role of anxiety and openness to experience in the classroom. *Behaviour and Information Technology, 36*(5), 524–539. https://doi.org/10.1080/0144929X.2016.1262900

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*(4), 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015). Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers. In *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3593–3602). ACM. https://doi.org/10.1145/2702123.2702415

Strohmaier, A. R., MacKay, K. J., Obersteiner, A., & Reiss, K. M. (2020). Eye-tracking methodology in mathematics education research: A systematic literature review. *Educational Studies in Mathematics, 104*(2), 147–200. https://doi.org/10.1007/s10649-020-09948-1

Südkamp, A., Kaiser, J., & Möller, J. (2014). Teachers' judgements of students' academic achievement: Results from field and experimental studies. In S. Krolak-Schwerdt, S. Glock, & M. Böhmer (Eds.), *Teachers' professional development: Assessment, training, and learning* (pp. 5–25). Sense. https://doi.org/10.1007/978-94-6209-536-6_2

Suls, J. (1986). Notes on the occasion of social comparison theory's thirtieth birthday. *Personality and Social Psychology Bulletin, 12*(3), 289–296. https://doi.org/10.1177/0146167286123004

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect? *Current Directions in Psychological Science, 11*(5), 159–163. https://doi.org/10.1111/1467-8721.00191

Suls, J., & Wheeler, L. (2000). A selective history of classic and neo-social comparison theory. In J. M. Suls & L. Wheeler (Eds.), *Handbook of social comparison* (pp. 3–19). Springer. https://doi.org/10.1007/978-1-4615-4237-7_1

Suls, J. M., & Mullen, B. (1982). From the cradle to the grave: Comparison of self-evaluation across the lifespan. In J. M. Suls & A. G. Greenwald (Eds.), *Psychological perspectives on the self* (Vol. 2, pp. 97–125). Lawrence Erlbaum.

Swann, W. B. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology, 53*(6), 1038–1051. https://doi.org/10.1037/0022-3514.53.6.1038

Syrjämäki, A. H., Isokoski, P., Surakka, V., Pasanen, T. P., & Hietanen, J. K. (2020). Eye contact in virtual reality – A psychophysiological study. *Computers in Human Behavior, 112*, Article 106454. https://doi.org/10.1016/j.chb.2020.106454

Tatler, B. W., & Land, M. F. (2015). Everyday visual attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 391–421). MIT Press. https://doi.org/10.7551/mitpress/10033.001.0001

Televantou, I., Marsh, H. W., Kyriakides, L., Nagengast, B., Fletcher, J., & Malmberg, L.-E. (2015). Phantom effects in school composition research: consequences of failure to control biases due to measurement error in traditional multilevel models. *School Effectiveness and School Improvement, 26*(1), 75–101. https://doi.org/10.1080/09243453.2013.871302

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. *Advances in Experimental Social Psychology, 21*, 181–227. https://doi.org/10.1016/s0065-2601(08)60227-0

Tesser, A., Millar, M., & Moore, J. (1988). Some affective consequences of social comparison and reflection processes: The pain and pleasure of being close. *Journal of Personality and Social Psychology, 54*(1), 49–61. https://doi.org/10.1037//0022-3514.54.1.49

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 104–124). The International Association for the Study of Attention and Performance.

Thijs, J., Verkuyten, M., & Helmond, P. (2010). A further examination of the Big-Fish-Little-Pond effect: Perceived position in class, class size, and gender comparisons. *Sociology of Education, 83*(4), 333–345. https://doi.org/10.1177/0038040710383521

Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research, 37*(5), 483–504. https://doi.org/10.1016/S0883-0355(03)00016-8

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92*(1), 144–151. https://doi.org/10.1037//0022-0663.92.U44

Trautwein, U., & Baeriswyl, F. (2007). Wenn leistungsstarke Klassenkameraden ein Nachteil sind: Referenzgruppeneffekte bei Übertrittsentscheidungen. *Zeitschrift für Pädagogische Psychologie, 21*(2), 119–133. https://doi.org/10.1024/1010-0652.21.2.119

Trautwein, U., Dumont, H., & Dicke, A.-L. (2015). Schooling: Impact on cognitive and motivational development. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., Vol. 21, pp. 119–124). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26056-X

Trautwein, U., & Lüdtke, O. (2005). The Big-Fish-Little-Pond effect: Future research questions and educational implications. *Zeitschrift für Pädagogische Psychologie, 19*(3), 137–140. https://doi.org/10.1024/1010-0652.19.3.137

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology, 98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology, 101*(4), 853–866. https://doi.org/10.1037/a0016306

Trautwein, U., & Möller, J. (2016). Self-concept: Determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel, & R. Roberts (Eds.), *Psychosocial skills and school systems in the twenty-fi rst century: Theory, research, and applications* (pp. 187–214). Springer. https://doi.org/10.1007/978-3-319-28606-8_8

Tseng, V. (2012). Social policy report: The uses of research in policy and practice. *Sharing Child and Youth Development Knowledge, 26*(2), 3–16.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Blackwell.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Villani, D., Gatti, E., Confalonieri, E., & Riva, G. (2012). Am I my avatar? A tool to investigate virtual body image representation in adolescence. *Cyberpsychology, Behavior, and Social Networking, 15*(8), 435–440. https://doi.org/10.1089/cyber.2012.0057

von Keyserlingk, L., Becker, M., Jansen, M., & Maaz, K. (2020). Leaving the pond—Choosing an ocean: Effects of student composition on STEM major choices at university. *Journal of Educational Psychology, 112*(4), 751–764. https://doi.org/10.1037/edu0000378

Webster, R. (2015). Declarative knowledge acquisition in immersive virtual learning environments. *Interactive Learning Environments, 24*(6), 1319–1333. https://doi.org/10.1080/10494820.2014.994533

Weinstein, C. S. (1991). The classroom as a social context for learning. *Annual Review of Psychology, 42*(1), 493–525. https://doi.org/10.1146/annurev.ps.42.020191.002425

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods, 15*(1), 18–37. https://doi.org/10.1037/a0015917

Wheeler, L. (1966). Motivation as a determinant of upward comparison. *Journal of Experimental Social Psychology, 1*, 27–31. https://doi.org/10.1016/0022-1031(66)90062-X

Wheeler, L., Martin, R., & Suls, J. (1997). The proxy model of social comparison for self-assessment of ability. *Personality and Social Psychology Review, 1*(1), 54–61. https://doi.org/10.1207/s15327957pspr0101_4

Wheeler, L., Shaver, K. G., Jones, R. A., Goethals, G. R., & Cooper, J. (1969). Factors determining choice of a comparison other. *Journal of Experimental Social Psychology, 5*, 219–232. https://doi.org/10.1016/0022-1031(69)90048-1

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*(3), 265–310. https://doi.org/10.1016/0273-2297(92)90011-P

Williams, D. (2010). The mapping principle, and a research framework for virtual worlds. *Communication Theory, 20*(4), 451–470. https://doi.org/10.1111/j.1468-2885.2010.01371.x

Wirth, M., Gradl, S., Prosinger, G., Kluge, F., Roth, D., & Eskofier, B. M. (2021). The impact of avatar appearance, perspective and context on gait variability and user experience in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)* (pp. 326–335). IEEE. https://doi.org/10.1109/vr50410.2021.00055

Wirth, W., Hartmann, T., Böcking, S., Vorderer, P., Klimmt, C., Schramm, H., Saari, T., Laarni, J., Ravaja, N., Gouveia, F. R., Biocca, F., Sacau, A., Jäncke, L., Baumgartner, T., & Jäncke, P. (2007). A process model of the formation of spatial presence experiences. *Media Psychology, 9*(3), 493–525. https://doi.org/10.1080/15213260701283079

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wolff, F., Helm, F., Zimmermann, F., Nagy, G., & Möller, J. (2018). On the effects of social, temporal, and dimensional comparisons on academic self-concept. *Journal of Educational Psychology, 110*(7), 1005–1025. https://doi.org/10.1037/edu0000248

Wood, J. V. (1996). What is social comparison and how should we study it? *Personality and Social Psychology Bulletin, 22*(5), 520–537. https://doi.org/10.1177/0146167296225009

Wouters, S., Colpin, H., Van Damme, J., & Verschueren, K. (2013). Endorsing achievement goals exacerbates the Big-Fish-Little-Pond effect on academic self-concept. *Educational Psychology, 35*(2), 252–270. https://doi.org/10.1080/01443410.2013.822963

Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. *Personality and Social Psychology Review, 21*(3), 253–277. https://doi.org/10.1177/1088868316652279

Wylie, R. (1979). *The self-concept* (2nd ed.). University of Nebraska Press.

Yazdan-Shahmorad, P., Sammaknejad, N., & Bakouie, F. (2020). Graph-based analysis of visual scanning patterns: A developmental study on green and normal images. *Nature Scientific Reports, 10*(1), Article 7791. https://doi.org/10.1038/s41598-020-63951-3

Zahorik, P., & Jenison, R. L. (1998). Presence as being-in-the-world. *Presence, 7*(1), 78–89. https://doi.org/10.1162/105474698565541

Zanbaka, C., Goolkasian, P., & Hodges, L. (2006). Can a virtual cat persuade you? The role of gender and realism in speaker persuasiveness. In *CHI '06: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1153–1162). ACM. https://doi.org/10.1145/1124772.1124945

Zell, E., & Alicke, M. D. (2009a). Contextual neglect, self-evaluation, and the frog-pond effect. *Journal of Personality and Social Psychology, 97*(3), 467–482. https://doi.org/10.1037/a0015453

Zell, E., & Alicke, M. D. (2009b). Self-evaluative effects of temporal and social comparison. *Journal of Experimental Social Psychology, 45*(1), 223–227. https://doi.org/10.1016/j.jesp.2008.09.007

Zell, E., & Alicke, M. D. (2010). The local dominance effect in self-evaluation: Evidence and explanations. *Personality and Social Psychology Review, 14*(4), 368–384. https://doi.org/10.1177/1088868310366144

Zu, T., Hutson, J., Loschky, L. C., & Rebello, N. S. (2020). Using eye movements to measure intrinsic, extraneous, and germane load in a multimedia learning environment. *Journal of Educational Psychology*, *112*(7), 1338–1352. https://doi.org/10.1037/edu0000441

8

APPENDIX

Script of the IVR Classroom Scenario

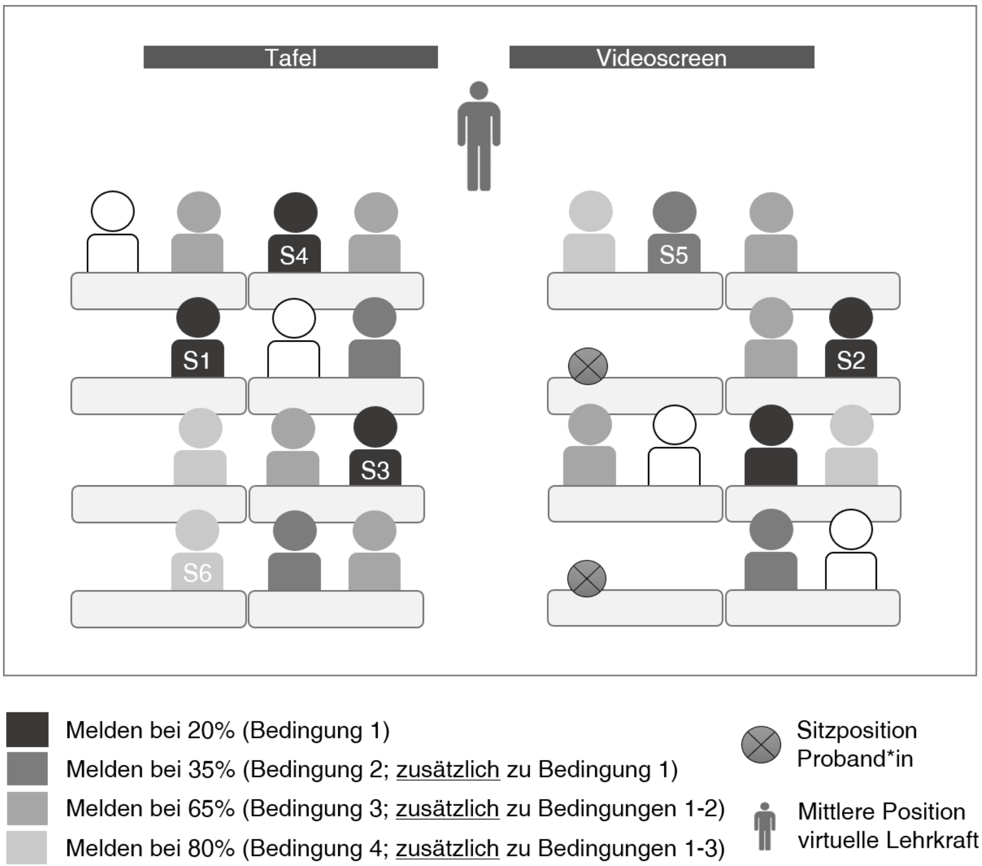
(German Original Version)

Allgemeine Beschreibung der Unterrichtssituation im IVR Klassenzimmer

Die virtuelle Lehrkraft steht vorne in der Mitte. Sie steht meist ruhig an einem Fleck und verlagert nur ab und an das Gewicht von einem Fuß auf den anderen während sie spricht. Wenn sie etwas auf dem Videoscreen zeigt, bewegt sie sich Richtung Videoscreen und bleibt links davon stehen.

Die virtuellen Mitschüler/innen sitzen mit Blick nach vorne da. Ihr Blick folgt immer der Lehrkraft (variiert jedoch auf natürliche Art zwischen Lernenden). Wenn ein/e virtuelle Mitschüler/in spricht, schauen einige der anderen Mitschüler/innen auf die/den Sprechenden. Reihen 1-2 und Reihen 3-4 verhalten sich dabei nahezu identisch. So wird sichergestellt, dass die Versuchspersonen unabhängig von der Sitzposition vorne/hinten im Klassenzimmer (siehe Darstellung unten) dieselbe Vergleichsgruppe haben, insbesondere auch in der Sitzposition vorne, (a) wenn sie sich umdrehen und die ganze Klasse in Betracht ziehen genauso wie (b) wenn sie sich nicht umdrehen und die letzten beiden Reihen kaum betrachten.

Die Unterrichtssituation gibt es in vier unterschiedlichen Bedingungen was das Leistungsniveau der virtuellen Mitschüler/innen anbelangt. Das Leistungsniveau wird über das Meldeverhalten manipuliert (d.h. der Anteil der sich meldenden Mitschüler/innen auf Fragen der Lehrkraft hin bzw. der Anteil der Mitschüler/innen, die durch Melden anzeigen die richtige Lösung einer Aufgabe zu kennen). Meldesituationen, in welchen die vier Bedingungen jeweils umgesetzt werden, sind im Skript in fett markiert: ***Meldesituation mit Variation in 4 Bedingungen;*** *Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich.*



Insgesamt lässt sich die Unterrichtssituation in vier Phasen (Szenen) unterteilen:

Minuten	**00:00 – 03:03**	**Szene 1 – Einführung** (inkl. Gewöhnungsphase an IVR Klassenzimmer)
Minuten	**03:03 – 07:30**	**Szene 2 – Input** (Lernen neuer Begriffe und Konzepte)
Minuten	**07:30 – 13:02**	**Szene 3 – Übungsaufgaben** (Bearbeitung von Aufgaben zu den neu gelernten Inhalten)
Minuten	**13:02 – 14:10**	**Szene 4 – Zusammenfassung und Abschluss**

Szenen 1, 2 und 3 bestehen aus Unterrichtsgesprächen, in welchen durch Fragen der virtuellen Lehrkraft und Antworten der virtuellen Mitschüler/innen die neuen Inhalte vorgestellt und erarbeitet werden. In Szene 4 spricht nur die virtuelle Lehrkraft und fasst die zentralen Punkte zusammen.

Detailliertes Skript der Unterrichtssituation im IVR Klassenzimmer

L = virtuelle Lehrkraft; **S** = virtuelle/r Mitschüler/in; im Skript erwähnte Folien sind in der Übersicht am Ende abgebildet.

<u>Szene 1 – Einführung</u> (Dauer ca. 3 Minuten)

[Die Lehrkraft kommt ins Klassenzimmer gelaufen, steht dann mittig vor der Klasse und sieht in Richtung der Schüler/innen. Auf der Tafel steht das Thema der Stunde: „Verstehen wie Computer denken". Die Lehrkraft begrüßt die Klasse.]

L: Hallo, willkommen im Kurs „Verstehen wie Computer denken"! Ich muss noch einmal ganz kurz etwas aus dem Lehrerzimmer holen. Ihr könnt euch in der Zwischenzeit einfach schon einmal das Klassenzimmer genauer ansehen. Ich bin dann gleich wieder da.

[Lehrkraft verlässt das Klassenzimmer für 20 Sekunden: Gewöhnungsphase an die VR Umgebung; alle virtuellen Schüler/innen sitzen auf ihrem Platz und sehen sich um, der/die Versuchsteilnehmer/in hat auch die Gelegenheit das zu tun. Nach 20 Sek. kommt die Lehrkraft wieder ins Klassenzimmer und bleibt mittig vor der Klasse stehen.]

L: So, jetzt können wir richtig anfangen. Wir sind hier im Kurs „Verstehen wie Computer denken". Was soll das denn eigentlich heißen, "Verstehen wie Computer denken"? Warum müssen wir uns vorstellen können, wie Computer denken? Wofür ist das gut? Was meint ihr? Hat jemand eine Idee?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S1 auf; sie deutet mit der Hand auf S1 und nickt S1 zu.]*

L: Ja?

S1: Damit wir ihnen sagen können, was sie tun sollen.

L: Ganz genau! Wir wollen, dass Computer Aufgaben für uns übernehmen, damit wir die nicht übernehmen müssen. Wer macht denn sowas normalerweise? Also Computern zu sagen, was sie tun sollen? Oder was tut man dafür, damit man denen das sagen kann?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S2 auf; sie deutet mit der Hand auf S2 und nickt S2 zu.]*

L: Ja?

S2: Programmierer zum Beispiel, die programmieren.

L: Ja, ganz genau. Programmierer sind die Leute, die Computern professionell sagen, was sie tun sollen. Es ist also deren Beruf Computern zu sagen was sie tun sollen. Wofür benutzt man denn Programmierungen. Hat jemand von euch da eine Idee?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S3 auf; sie deutet mit der Hand auf S3 und nickt S3 zu.]*

S3: Man kann Computerspiele machen.

L: Genau, man kann Computerspiele programmieren. Kann man außer Computerspielen noch mehr programmieren? Oder ist Programmierung nur für Computerspiele da? Hat da jemand Ideen?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S4 auf; sie deutet mit der Hand auf S4 und nickt S4 zu.]*

S4: Ich glaub man braucht es auch für Internetseiten, um die zu programmieren zum Beispiel...

L: Ja genau, das sind doch gute Beispiele. Gibt es noch eine Idee?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S3 auf; sie deutet mit der Hand auf S3 und nickt S3 zu.]*

S3: Es gibt Programme, die selbst rechnen können. Oder die Fehler finden.

L: Ja genau, ihr habt bestimmt schon mal auf dem Computer oder dem Smartphone den Taschenrechner gesehen, den man da benutzen kann. Das sind gute Beispiele. Also wenn man programmiert, dann kann man damit Computerspiele machen, Webseiten, Taschenrechner… Aber man kann mit Programmierungen auch alle möglichen Probleme lösen, die auf den ersten Blick vielleicht gar nichts mit Computern zu tun haben. Dafür muss man aber eben verstehen, wie Computer so denken. Dann können wir ihnen sagen was sie tun sollen und sie wissen was sie zu tun haben. Jetzt denken Computer aber natürlich nicht so wie wir Menschen, sonst könnten wir uns einfach vor den Computer setzen und sagen „so, jetzt mach‘ mal das!“ und der Computer würde das dann tun. So funktioniert das aber nicht. Man muss verstehen, wie der Computer denkt, damit man ihm dann quasi in seiner Sprache sagen kann, was er eigentlich machen soll.

<u>Szene 2 – Input</u> (Dauer ca. 4,5 Minuten)

L: Wir wollen uns heute das mal ein bisschen genauer ansehen, wie das funktioniert, dass man Computern in ihrer Sprache sagen kann, was sie tun sollen. Dazu sehen wir uns heute zwei Begriffe etwas näher an. Ihr seht sie hier stehen.

[Die Lehrkraft deutet auf den Videoscreen. Auf dem Videoscreen ist Folie 1 zu sehen, auf der links der Begriff „Sequenz“ und rechts der Begriff „Schleife“ zu lesen ist.]

L: Der erste Begriff ist „Sequenz“. Hat den von euch schon mal jemand gehört?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet.]*

L: Ok. Wer von euch glaubt denn, dass er den Begriff „Sequenz“ auch schon erklären kann, was das ist eine Sequenz?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet.]*

L: Ja das ist gar nicht so einfach. Ich habe euch mal eine Definition mitgebracht, die erklärt was eine Sequenz ist.

[Auf dem Videoscreen erscheint Folie 2 mit der Definition des Begriffs „Sequenz“: „Eine Sequenz ist eine Liste von Befehlen, die in einer bestimmten Reihenfolge ausgeführt werden“.]

L: Wer von euch kann das denn einmal schnell vorlesen?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S5 auf; sie deutet mit der Hand auf S5 und nickt S5 zu.]*

S5: Sequenz. Eine Sequenz ist eine Liste von Befehlen, die in einer bestimmten Reihenfolge ausgeführt werden.

L: Dankeschön. Eine Sequenz ist eine Liste von Befehlen, also das heißt die Befehle sind nacheinander. Und die Befehle werden immer genau in dieser bestimmten Reihenfolge ausgeführt, in der sie auch aufgeschrieben sind. Das heißt, da wird nicht hin- und hergesprungen, sondern der erste Befehl wird als erstes ausgeführt, der zweite Befehl als zweites und dann der dritte Befehl als drittes und immer so weiter. Immer genau in der gleichen Reihenfolge.

[Auf dem Videoscreen erscheint wieder Folie 1 mit den beiden Begriffen „Sequenz“ und „Schleife“. Die Lehrkraft deutet auf den zweiten Begriff „Schleife“ auf dem Videoscreen.]

L: Das zweite Wort „Schleife“ wirkt erstmal ein bisschen leichter, oder? Wer von euch, hat das Wort „Schleife“ schon einmal gehört?

[Alle Schüler/innen melden sich.]

L: Ja, alle. Das hab‘ ich mir gedacht. Wenn wir es hier im Kurs verwenden, dann benutzen wir das Wort Schleife ein bisschen anders. Nicht in der Bedeutung von der Schleife beim Schuhe binden, oder wenn man irgendwie eine schöne Frisur macht und da eine Schleife reinsetzt, sondern wir benutzen das Wort Schleife so.

[Auf dem Videoscreen erscheint Folie 3 mit der Definition des Begriffs „Schleife“: „Eine Schleife ist eine Liste von Befehlen, die mehrmals hintereinander wiederholt ausgeführt wird“. Die Lehrkraft zeigt auf die Definition.]

L: Wer von euch würde das vorlesen?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S6 auf; sie deutet mit der Hand auf S6 und nickt S6 zu.]*

S6: Schleife. Eine Schleife ist eine Liste von Befehlen, die mehrmals hintereinander wiederholt ausgeführt wird.

L: Danke. Liste von Befehlen, das kommt uns vielleicht noch aus der ersten Definition bekannt vor, oder? Die Sequenz war die Liste von Befehlen, die nacheinander in einer bestimmten Reihenfolge ausgeführt werden. So ist das bei der Schleife jetzt auch, aber bei der Schleife, fängt das ganze wieder von vorne an, wenn man am Ende der Sequenz angekommen ist. Wenn wir zum Beispiel vier Befehle haben: Erster Befehl, zweiter Befehl, dritter Befehl, vierter Befehl, bei einer Sequenz ist danach Schluss. Bei der Schleife springen wir dann wieder an den Anfang und machen das Ganze nochmal. Also nach dem vierten Befehl kommt wieder der erste Befehl. Das heißt bei der Schleife wird sozusagen eine Sequenz wiederholt und nicht nur einmal ausgeführt.

[Auf dem Videoscreen ist wieder Folie 1 mit den beiden Begriffen „Sequenz“ und „Schleife“ zu sehen.]

L: Ihr habt vielleicht Ideen oder Beispiele aus eurem Alltag; gibt es irgendwas, das euch jetzt einfällt, was ihr in eurem Alltag beobachten könnt und wobei ihr denkt „ah, dass hier funktioniert wie eine Sequenz oder das hier funktioniert wie eine Schleife“?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S1 auf; sie deutet mit der Hand auf S1 und nickt S1 zu.]*

L: Ja?

S1: Zum Beispiel beim Busfahren. Der Bus fährt los an einer Haltestelle und dann von Haltestelle zu Haltestelle. Vom Anfang bis zum Ende.

L: Genau, 1. Haltestelle, 2. Haltestelle, 3. Haltestelle, die fährt der Bus alle nacheinander ab, also immer in einer bestimmten Reihenfolge. Der fährt nicht mal hier hin und mal da hin, sondern er fährt immer in der gleichen Reihenfolge zu den Haltestellen. Ist das jetzt eine Sequenz oder eine Schleife?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S2 auf; sie deutet mit der Hand auf S2 und nickt S2 zu.]*

L: Ja?

S2: Ich glaube beides, oder? Also zumindest, wenn der Bus eine Runde fährt, ist es eine Sequenz, wenn er immer wieder die gleiche Runde fährt, dann ist es eine Schleife.

L: Genau, einmal den Fahrplan abfahren ist eine Sequenz. Und wenn der Busfahrer zum Beispiel einen langen Arbeitstag hat und immer wieder die gleiche Runde fährt, dann fährt er diese Sequenz als Schleife, weil er wiederholt sie. Das ist ein sehr gutes Beispiel! Hat noch jemand eine Idee, was im Alltag wie eine Schleife oder eine Sequenz funktioniert?

*[**Meldesituation mit Variation in 4 Bedingungen;** Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet. Die Lehrkraft ruft S1 auf; sie deutet mit der Hand auf S1 und nickt S1 zu.]*

S1: Ja also vielleicht bei einer Schleife eine Uhr; die Zeiger machen ja auch immer das gleiche, die drehen sich immer weiter, also es geht eigentlich immer weiter im Kreis.

[Die Lehrkraft zeichnet mit den Fingern einen Kreis in die Luft.]

L: Ganz genau, die Zeiger fangen, wenn man die Uhr zum Beispiel frisch eingestellt hat, ganz oben an und drehen sich dann im Kreis immer an den Zahlen entlang. Die meisten Uhren haben zwei oder drei Zeiger für Stunden und Minuten oder für Stunden, Minuten und Sekunden und jeder von diesen Zeigern hat sozusagen seine eigene Schleife, denn die Zeiger laufen ja unterschiedlich schnell. Ganz genau, das ist ein sehr gutes Beispiel!

Szene 3 – Übungsaufgaben (Dauer ca. 5,5 Minuten)

L: Um jetzt noch ein bisschen tiefer einzusteigen, und nicht nur drüber zu reden, wie Sequenzen und Schleifen funktionieren, habe ich euch zwei Aufgaben mitgebracht, da können wir jetzt mal testen, ob ihr schon verstanden habt, wie Computer in Sequenzen und Schleifen denken.

[Die Lehrkraft zeigt auf den Videoscreen. Dort erscheint Folie 4 mit der ersten Übungsaufgabe „Dein Roboter".]

L: So, das ist die erste Aufgabe. Stellt euch vor, ihr habt einen kleinen Roboter, den könnt ihr programmieren. Dem könnt ihr jetzt aber nicht wie wir Menschen das tun würden einfach irgendwas sagen, was ihr von ihm wollt und dann tut er das. Ihr könnt ihm nur mit zwei Knöpfen sagen was er zu tun hat. Also er hat zwei Knöpfe: Der eine ist der grüne Knopf. Wenn ihr den drückt, dann fährt der Roboter geradeaus. Das heißt, er fährt ein kleines Stück vorwärts. Wenn ihr den lila Knopf drückt, dann dreht sich der Roboter auf der Stelle nach rechts. Ich stell mich jetzt mal so hin, dass ihr es besser verstehen könnt. Ich schaue in die gleiche Richtung wie ihr. Das heißt wenn ihr jetzt den lila Knopf drückt, macht der Roboter so.

[Die Lehrkraft macht die Bewegung des Roboters vor und dreht sich auf der Stelle nach rechts.]

L: Für die Aufgabe müsst ihr jetzt überlegen, welche Sequenz von Knöpfen ihr drücken müsst, in einer bestimmten Reihenfolge, damit der Roboter am Ende nach links gedreht ist.

[Die Lehrkraft steht in Blickrichtung der Schüler/innen und dreht sich dann nach links, um die Zielposition des Roboters vorzumachen.]

L: Das heißt am Anfang steht der Roboter so da. Am Ende soll er so stehen. Ihr habt aber nur die beiden Knöpfe, bei denen er entweder so machen kann oder so.

[Die Lehrkraft demonstriert die beiden Bewegungen des Roboters, macht einen Schritt nach vorne und dreht sich auf der Stelle nach rechts. Dann dreht sich die Lehrkraft zurück zur

Klasse. Die vier Antwortmöglichkeiten, welche zuvor ausgeblendet waren, erscheinen auf der Folie 4.]

L: Ok. Ich gebe euch jetzt ein bisschen Zeit. Schaut euch die vier Antwortmöglichkeiten mal an und überlegt euch dann, welche den Roboter sich so bewegen lässt, dass er am Ende nach links gedreht ist.

[5 Sekunden Pause; die Lehrkraft steht vorne und bewegt sich kaum, Schüler/innen schauen nach vorne auf den Videoscreen.]

L: Habt ihr euch alle Antworten überlegt? Ja? Ok. Dann gehe ich jetzt nacheinander die vier Antwortmöglichkeiten durch. Und ihr meldet euch einfach immer dann, wenn ihr glaubt: Das ist jetzt die richtige Antwort!

[Schüler/innen nicken.]

L: Wer von euch denkt, dass A die richtige Antwort ist?

[Antwort A ist falsch; bei der richtigen Antwort C gelten die üblichen Meldebedingungen, zwischen den restlichen falschen Antworten A, B und D werden die Meldungen der restlichen Schüler/innen (abzüglich der Verteilung je nach Bedingung bei Antwort C) zufällig verteilt.]

L: Wer glaubt, B ist richtig?

[Antwort B ist falsch; bei der richtigen Antwort C gelten die üblichen Meldebedingungen, zwischen den restlichen falschen Antworten A, B und D werden die Meldungen der restlichen Schüler/innen (abzüglich der Verteilung je nach Bedingung bei Antwort C) zufällig verteilt.]

L: Und wer hält Antwort D für richtig?

[Antwort D ist falsch; bei der richtigen Antwort C gelten die üblichen Meldebedingungen, zwischen den restlichen falschen Antworten A, B und D werden die Meldungen der restlichen Schüler/innen (abzüglich der Verteilung je nach Bedingung bei Antwort C) zufällig verteilt.]

L: Und wer glaubt, Antwort C ist richtig?

[Antwort C ist richtig, daher gelten hier die üblichen Meldebedingungen. ***Meldesituation mit Variation in 4 Bedingungen;*** *Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die virtuellen Mitschüler/innen melden, schauen sich einzelne um, um zu sehen, wer sich (noch) meldet.]*

L: Ja, jetzt verrat ich's euch, C ist richtig. Die Frage ist natürlich, warum ist C richtig und nicht die anderen. Ich kann euch mal zeigen, was passiert, wenn man C ausführt mit dem Roboter. Ich stell mich wieder so hin, dass ich für euch in die richtige Richtung schaue. Wenn wir C ausführen, dann macht der Roboter das hier.

[Die Lehrkraft imitiert die Bewegungen des Roboters, welche sie erklärt.]

L: Er dreht sich einmal nach rechts, er dreht sich nach rechts, er dreht sich einmal nach rechts. Und dann ist es vorbei, weil die Sequenz zu Ende ist. Das heißt, jetzt stehe ich so da, dass der Roboter von seiner Ausgangsposition aus nach links gedreht ist. Wir haben also das richtige Ziel erreicht. Könnt ihr das alle nachvollziehen? Ok, also das war jetzt eine Sequenz, weil der Roboter das ganze nur einmal ausgeführt hat und dann ist es vorbei.

[Die Lehrkraft deutet wieder auf den Videoscreen. Dort erscheint Folie 5 mit der zweiten Aufgabe „Der Schildkrötenroboter".]

L: Ok. Jetzt schauen wir uns mit einer zweiten Aufgabe mal an, wie eine Schleife funktionieren könnte. Für diese Aufgabe stellt ihr euch wieder vor, dass ihr einen kleinen Roboter habt, aber diesmal ist es ein Schildkrötenroboter. Der Roboter kann jetzt schon drei Befehle, der versteht also schon ein bisschen mehr als der erste Roboter. Jetzt ist dieser Roboter aber so eingestellt, dass er das Ganze, wenn ihr es ihm als Sequenz gebt, nicht einmal ausführt und dann aufhört, sondern die Sequenz, die ihr ihm gebt, so lange ausführt, wie er an ist. Das heißt ihr müsst den Roboter ausschalten, damit er damit aufhört, die Befehle immer wieder auszuführen, die ihr ihm gegeben habt. Das heißt, die Sequenz, die ihr dem Roboter gebt, die führt er als Schleife aus, solange wie er angeschaltet ist. Die Aufgabe ist jetzt, dass ihr überlegen sollt, wie ihr dem Roboter sagen könnt, dass er ein Quadrat fahren soll.

[Die Lehrkraft dreht sich zurück zur Klasse. Die vier Antwortmöglichkeiten, welche zuvor ausgeblendet waren, erscheinen auf der Folie 5.]

L: Ich gebe euch jetzt noch einmal ein bisschen Zeit, denkt einmal drüber nach, was ihr denkt welche Antwortmöglichkeit die richtige ist und dann besprechen wir es zusammen.

[5 Sekunden Pause; die Lehrkraft steht vorne und bewegt sich kaum, Schüler/innen schauen nach vorne auf den Videoscreen.]

L: So, seid ihr soweit? Haben sich alle eine Antwort überlegt? Prima, dann machen wir es genau wie vorher. Ich sage immer die Antwortmöglichkeit und ihr meldet euch, wenn ihr glaubt, dass die Antwortmöglichkeit die richtige ist. Wer hält Antwort A für richtig?

[Antwort A ist richtig, daher gelten hier die üblichen Meldebedingungen. ***Meldesituation mit Variation in 4 Bedingungen;*** *Bedingung 1: 20% melden sich; Bedingung 2: 35% melden sich; Bedingung 3: 65% melden sich; Bedingung 4: 80% melden sich. Während sich die Schüler melden, schauen sich einzelne um, um zu sehen, wer sich noch meldet.]*

L: Ok. Wer hält Antwort B für die richtige Antwort?

[Antwort B ist falsch; bei der richtigen Antwort A gelten die üblichen Meldebedingungen, zwischen den restlichen falschen Antworten B, C und D werden die Meldungen der restlichen Schüler/innen (abzüglich der Verteilung je nach Bedingung bei Antwort A) zufällig verteilt.]

L: Und wer glaubt, dass Antwort C richtig ist?

[Antwort C ist falsch; bei der richtigen Antwort A gelten die üblichen Meldebedingungen, zwischen den restlichen falschen Antworten B, C und D werden die Meldungen der restlichen Schüler/innen (abzüglich der Verteilung je nach Bedingung bei Antwort A) zufällig verteilt.]

L: Ok. Wer hält die Antwort D für richtig?

[Antwort D ist falsch; bei der richtigen Antwort A gelten die üblichen Meldebedingungen, zwischen den restlichen falschen Antworten B, C und D werden die Meldungen der restlichen Schüler/innen (abzüglich der Verteilung je nach Bedingung bei Antwort A) zufällig verteilt.]

L: Bei dieser Aufgabe ist Antwort A richtig. Ich kann euch einmal zeigen warum. Ich bin jetzt wieder der Schildkrötenroboter und ich führe einfach mal das aus, was da gegeben ist.

[Die Lehrkraft imitiert die Bewegungen des Roboters, welche sie erklärt.]

L: 30 cm vorwärts, nach rechts drehen, 30 cm vorwärts, nach rechts drehen. Jetzt bin ich aber noch kein Quadrat gefahren. Aber dieses Mal ist es ja nicht nur so eine Sequenz, sondern der Roboter führt es als Schleife aus, das heißt er wiederholt es immer wieder. Ich mach also weiter, dann kann ich wieder von vorne anfangen. 30 cm vorwärts, nach rechts, 30 cm vorwärts, nach rechts. Jetzt hätte der Roboter ein Quadrat fertig gefahren, aber wenn ihr ihn nicht ausschaltet, macht er einfach weiter und fährt die ganze Zeit weiter das gleiche Quadrat, weil er immer wieder diese Sequenz ausführt, das heißt er führt sie als Schleife aus. Habt ihr das soweit alle verstanden?

[Schüler/innen nicken.]

<u>Szene 4 – Zusammenfassung und Abschluss</u> (Dauer ca. 1 Minute)

L: Das ist das, was wir heute neu gelernt haben, darüber wie Computer denken, also wie sie verstehen, was wir ihnen sagen.

[Die Lehrkraft deutet nochmals auf den Videoscreen, wo noch die Folie 5 mit der letzten Aufgabe zu sehen ist.]

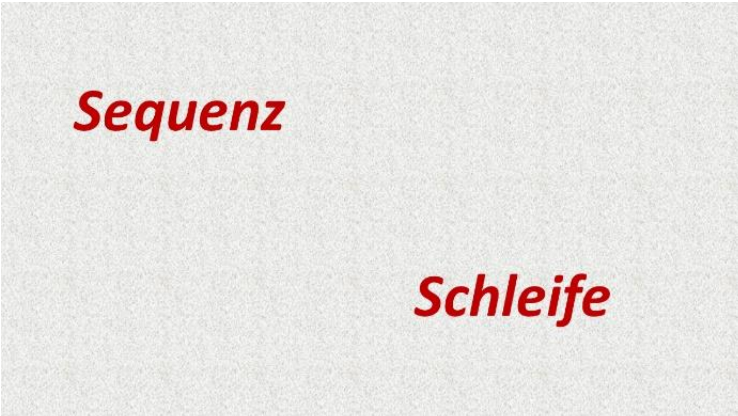
L: In dem Fall hier: Wenn ich dem Schildkrötenroboter sagen würde „fahr‘ in einem Quadrat oder fahr‘ in einem Viereck“, dann versteht er mich nicht. Denn der Roboter kennt nur drei Befehle. Der kann nur verstehen, wenn er sich nach rechts drehen soll, wenn er sich nach links drehen soll oder, wenn er dreißig Zentimeter vorwärtsfahren soll. Wenn ich dem jetzt irgendwas sage, was er gar nicht kennt, z.B. „fahre in einem Quadrat“, dann kann er das nicht machen. Deswegen ist es unsere Aufgabe als Programmierer, dass wir das, was wir dem Computer oder in dem Fall dem Roboter sagen wollen, so zerlegen, dass er es verstehen kann. Wir zerlegen das also in einzelne Befehle, und die ordnen wir dann als Sequenz an, das heißt der Roboter macht die Befehle nacheinander. Oder wir ordnen die Sequenz als Schleife an. Das heißt dann, dass diese Reihe von Befehlen, dann immer wiederholt ausgeführt wird. Versteht ihr was ich meine?

[Schüler/innen nicken.]

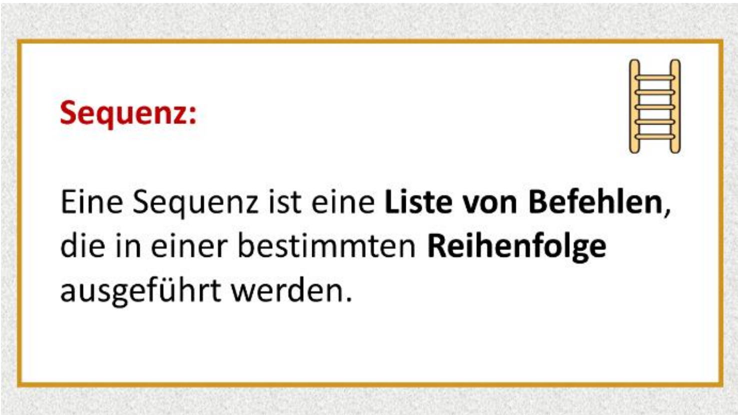
L: Ok, das freut mich. Dann hoffe ich, dass ihr heute schon ein bisschen was darüber lernen konntet, wie Computer denken, und ich freue mich euch das nächste Mal wiederzusehen!

Übersicht Präsentationsfolien

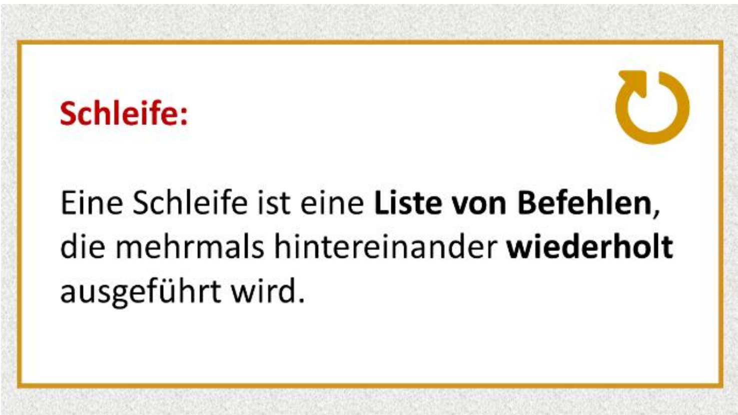
Folie 1



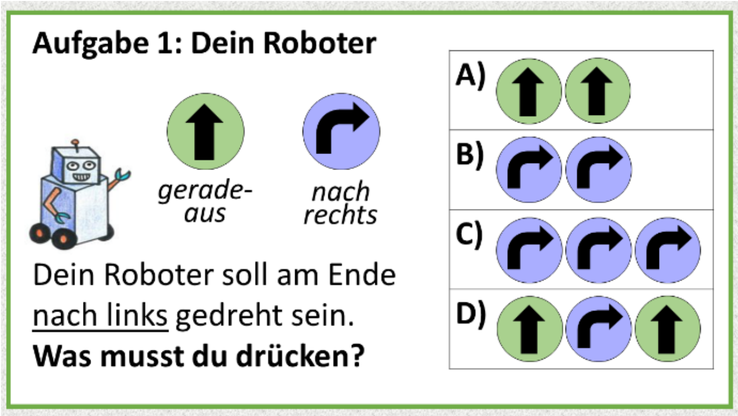
Folie 2



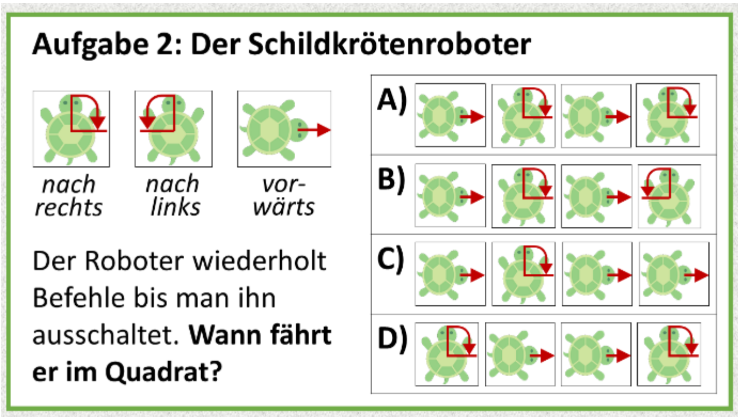
Folie 3



Folie 4



Folie 5



Script of the IVR Classroom Scenario

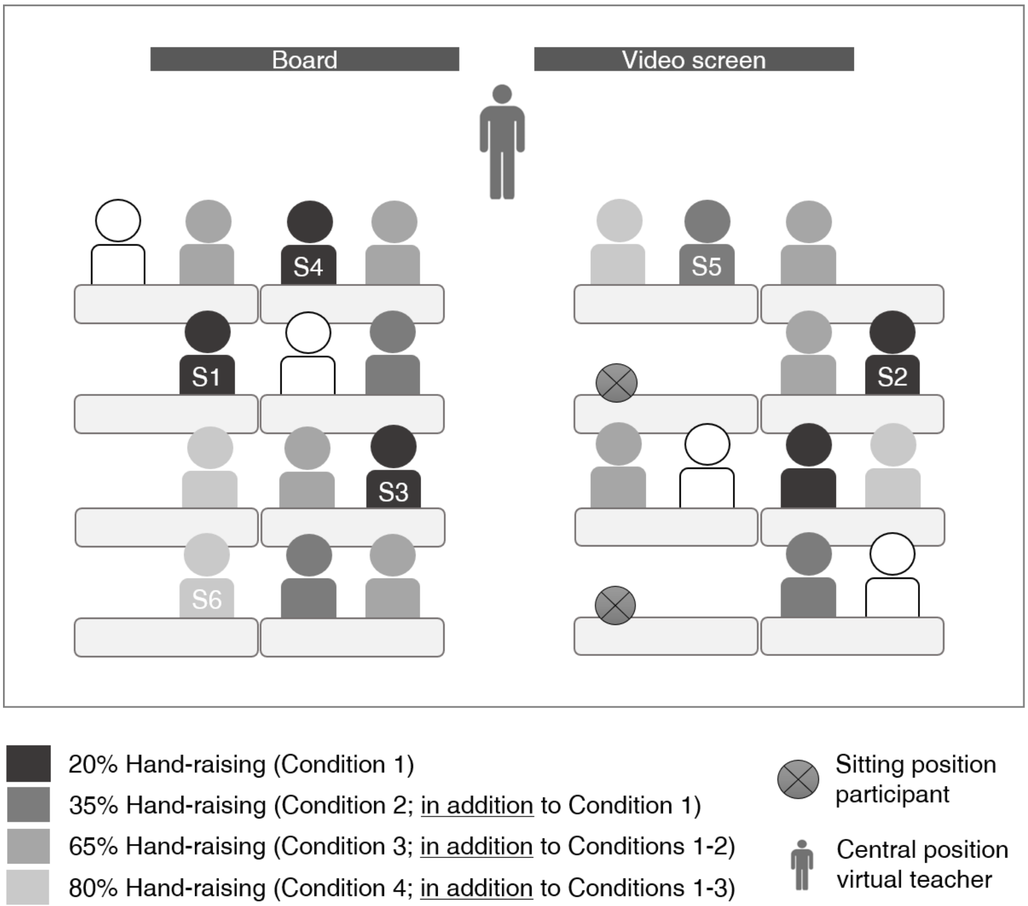
(English Translation)

General Description of the Classroom Scenario in the IVR Classroom

The virtual teacher stands in a central position in the front of the classroom. She mostly stands in one spot and only occasionally shifts her weight from one foot to the other while speaking. When she shows something on the video screen, she moves toward the video screen and stays to the left of it.

The virtual peer learners sit facing forward. Their gaze always follows the teacher (varying naturally between learners). When a peer learner is speaking, some of the other learners look at the person who is speaking. Rows 1-2 and Rows 3-4 behave almost identically. This ensures that the participating students have the same comparison group regardless of the position of their seat in the front/back of the classroom (see the figure below), especially when seated in the front (a) when they around and look at the whole class as well as (b) when they do not turn around and barely look at the last two rows.

The classroom scenario is available in four different conditions regarding the performance level of the virtual classmates. The performance level is manipulated via the hand-raising behavior of the virtual classmates (i.e., the proportion of students raising their hands in response to the teacher's questions or to indicate that they know the correct solution to a task). Situations with hand-raising in which the four conditions are implemented are highlighted in bold in the script: ***Hand-raising situation that varies across four conditions;*** *Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands.*



Overall, the classroom scenario can be divided into four phases (scenes):

Minutes	**00:00 – 03:03**	**Scene 1 – Introduction** (including becoming familiar with the IVR classroom)
Minutes	**03:03 – 07:30**	**Scene 2 – Input** (learning new terms and concepts)
Minutes	**07:30 – 13:02**	**Scene 3 – Exercises** (working on tasks related to the new material that is learned)
Minutes	**13:02 – 14:10**	**Scene 4 – Summary and conclusion**

Scenes 1, 2, and 3 consist of a dialogue between the virtual teacher and classmates in which the new material is introduced and explained through questions from the virtual teacher and answers from the classmates. In Scene 4, only the virtual teacher speaks and summarizes the key points.

Detailed Script of the Classroom Scenario in the IVR Classroom

T = virtual teacher; **S** = virtual student/peer learner; slides mentioned in the script are shown in the overview at the end.

Scene 1 – Introduction (duration approximately 3 min)

[The teacher comes into the classroom, stands in the middle in front of the class, and looks toward the students. The topic of the lesson is written on the blackboard: "Understanding how computers think." The teacher greets the class.]

T: Hello, welcome to the course "Understanding how computers think"! I need to quickly get something from my office. In the meantime, you can take a closer look at the classroom. I'll be right back.

[The teacher leaves the classroom for 20 s: familiarization phase with VR environment; all peer learners sit in their seats and look around; the participant also has the opportunity to do so. After 20 s, the teacher comes back into the classroom and stays in the middle in front of the class.]

T: So, now we can really start. This course is about "Understanding how computers think." What does that actually mean, "understanding how computers think"? Why do we need to be able to imagine how computers think? What is it good for? What do you think? Does anyone have any ideas?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S1; she points to S1 with her hand and nods.]*

T: Yes?

S1: So we can tell them what to do.

T: That's right! We want computers to do tasks for us so we don't have to do them. Who normally does that, telling computers what to do? Or how do you tell them to do certain things?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S2; she points to S2 with her hand and nods.]*

T: Yes?

S2: Programmers, for example, they code.

T: Yes, exactly. Programmers are the people who professionally tell computers what to do. So, it's their job to tell computers what to do. What do you use programming for? Do you have any ideas?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S3; she points to S3 with her hand and nods.]*

S3: You can make computer games.

T: That's right, you can program computer games. Can you program more than computer games? Or is programming only for computer games? Does anyone have any ideas about this?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S4; she points to S4 with her hand and nods.]*

S4: I think you also need it for websites, for example, to program them...

T: Yes exactly, these are good examples. Any other ideas?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S3; she points to S3 with her hand and nods.]*

S3: Or there are programs that can calculate themselves. Or find errors automatically.

T: Yeah right, you've probably seen a calculator on a computer or smartphone that you can use. These are good examples. So, if you code, you can use it to make computer games, websites, calculators... But you can also use programming to solve all kinds of problems that at first glance might have nothing to do with computers. But to do this, you have to understand how computers think. Then you can tell them what to do in a way that they will understand. Of course, computers don't think the way we humans do. Otherwise, we could just sit down in front of the computer and say, “Do this!” and the computer would do it. But that's not how it works. You have to be able to understand how the computer thinks, so that you can then tell it in its language, so to speak, what it should actually do.

<u>Scene 2 – Input</u> (duration approximately 4.5 min)

T: Today we want to take a closer look at how to tell computers what to do in their own language. To do this, we're going to take a look at two terms today. You can see them here.

[The teacher points to the video screen. The video screen shows Slide 1, which shows the term “sequence” on the left and the term “loop” on the right.]

T: The first term is "sequence." Have any of you heard this one before?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand.]*

T: Okay. Who among you thinks they can already explain the term “sequence”? What is a sequence?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand.]*

T: Yes, it's not that simple. I brought you a definition that explains what a sequence is.

[On the video screen, Slide 2 appears with the definition of the term “sequence”: “A sequence is a list of commands executed in a specific order.”]

T: Who among you would like to read it out loud?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual*

classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S5; she points to S5 with her hand and nods.]

S5: Sequence. A sequence is a list of commands executed in a specific order.

T: Thank you. A sequence is a list of commands, so this means the commands follow one after the other. And the commands are always executed exactly in the particular order in which they are written down. That is, there is no jumping back and forth, but the first command is executed first, the second command is executed second, and then the third command is executed third, and so on and on. Always in exactly the same order.

[On the video screen, Slide 1 appears again with the two terms "sequence" and "loop." The teacher points to the second term "loop" on the video screen].

T: The second word "loop" seems a bit easier at first, doesn't it? How many of you have heard the word loop before?

[All the students raise their hands.]

T: Yes, all of you. That's what I thought. When we use it here in class, we use the word loop a little bit differently. Not in the meaning of tying a loop when you tie your shoes, or when you make a nice hairstyle with a loop in your hair. We use the word loop as follows.

[On the video screen, Slide 3 appears with the definition of the term "loop": "A loop is a list of commands that is executed repeatedly several times in a row." The teacher points to the definition.]

T: Who among you would like to read it out loud?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S6; she points to S6 with her hand and nods.]*

S6: Loop. A loop is a list of commands that is repeatedly executed several times in a row.

T: Thank you. List of commands—that might look familiar from the first definition, right? The sequence was the list of commands executed one after another in a certain order. That's the same for a loop, but with the loop, it starts all over again when you get to the end of the sequence. For example, if we have four commands: first command, second command, third command, fourth command; with a sequence, it ends after that. With a loop, we then jump back to the beginning and do the whole thing again. So, after the fourth command, we go back to the first command again. This means that with the loop, a sequence is repeated, so to speak, and not executed only once.

[The video screen again shows Slide 1 with the two terms "sequence" and "loop."]

T: When you think of your everyday life, is there anything that comes to your mind now that you can observe in your everyday life and think “ah, this works like a sequence or this works like a loop”?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S1; she points to S1 with her hand and nods.]*

T: Yes?

S1: For example, when riding the bus. The bus leaves from one stop and then from stop to stop. From the beginning to the end.

T: Exactly, first stop, second stop, third stop, the bus goes to all of them one after the other, so always in a certain order. It doesn't go here and there, but it always goes to the stops in the same order. Is that a sequence or a loop now?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S2; she points to S2 with her hand and nods.]*

T: Yes?

S2: I think both, right? At least if the bus does one lap, it's a sequence, if it does the same lap over and over, it's a loop.

T: Exactly, running the schedule once is a sequence. And if the bus driver, for example, has a long day at work and makes the same round over and over again, then he's making that sequence as a loop, because he's repeating it! This is a very good example! Does anyone else have an example of how a loop or a sequence works in everyday life?

*[**Hand-raising situation that varies across four conditions;** Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand. The teacher calls on S1; she points to S1 with her hand and nods.]*

S1: Yes, so maybe with a clock, the hands of the clock always do the same thing, they turn around and around in the circle.

[The teacher draws a circle in the air with her fingers.]

T: That's right, the hands start at the top when the clock is freshly set, for example, and then turn in a circle along the numbers. Most watches have two or three hands for hours and minutes, or for hours, minutes, and seconds, and each of these hands has its own loop that it runs, so to speak, because the loops run at different speeds. That's right, that's a very good example!

<u>Scene 3 – Exercises</u> (duration approximately 5.5 min)

T: In order to dive in a little deeper and not just talk about how sequences and loops work, I've brought along two exercises for you to test whether you now understand how computers think in sequences and loops.

[The teacher points to the video screen. Slide 4 appears with the first exercise called "Your robot."]

T: So, this is the first task. Imagine you have a little robot, you can program it, but you can't just tell it what you want it to do like we humans would, and then it does it. You can only tell it what to do with two buttons. The robot has two buttons. One is the green button. When you press it, the robot goes straight ahead. That is, it moves a little bit forward. If you press the purple button, the robot turns to the right on the spot. I'm going to stand in such a way that you can understand it better. I am standing facing the same direction as you. This means that if you press the purple button, the robot will do this.

[The teacher demonstrates the movement of the robot and turns to the right on the spot.]

T: For the task, you now have to think about which sequence of buttons you have to press, in a certain order, so that the robot is turned to the left at the end.

[The teacher stands facing forward and then turns to the left to demonstrate the final position of the robot.]

T: This means that, in the beginning, the robot is standing like this. At the end, it should stand like this. But you only have the two buttons where he can either go like this or go like that.

[The teacher demonstrates the two movements of the robot, taking a step forward and turning to the right on the spot. Then the teacher turns back to the class. The four answer choices, which were previously hidden, appear on Slide 4.]

T: Ok. I'm going to give you some time now. Take a look at the four answer choices and think about which one will make the robot move so that it will end up turned to the left.

[There is a 5 s pause; the teacher stands at the front and hardly moves; the students look forward at the video screen.]

T: Have you thought about the answers? Yes? Okay. Then I'll go through the four possible answers one after the other. And you just raise your hands whenever you think it's the right answer!

[Students nodding.]

T: Which one of you thinks A is the right answer?

[Answer A is incorrect; for the correct answer C, the usual hand-raising conditions apply; between the remaining incorrect answers A, B, and D, the numbers of the remaining students who raise their hands (minus the distribution depending on the condition for Answer C) are randomly distributed.]

T: Who thinks B is correct?

[Answer B is incorrect; for the correct answer C, the usual hand-raising conditions apply; between the remaining incorrect answers A, B, and D, the numbers of the remaining students who raise their hands (minus the distribution depending on the condition for Answer C) are randomly distributed.]

T: And who thinks Answer D is correct?

[Answer D is incorrect. For the correct answer C, the usual hand-raising conditions apply; between the remaining incorrect answers A, B, and D, the numbers of the remaining students who raise their hands (minus the distribution depending on the condition for Answer C) are randomly distributed.]

T: And who thinks Answer C is correct?

[Answer C is correct, so the usual hand-raising conditions apply here. ***Hand-raising situation that varies across four conditions;*** *Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand.]*

T: Yes, C is correct. The question, of course, is why is C correct and not the others? I can show you what happens when you execute C with the robot. I'll stand facing forward again so that I'm looking in the same direction as you. When we execute Option C, the robot does this.

[The teacher imitates the robot's movements while explaining.]

T: It turns to the right, it turns to the right again, and it turns to the right again. And then it's over because the sequence is over. That is, now the robot has turned to the left from its starting position by turning to the right three times. So, we have achieved the right goal. Can you all understand this? Okay, so that was a sequence because the robot did the whole thing only once and then it was over.

[The teacher points to the video screen again. Slide 5 appears there with the second task, "The Turtle Robot."]

T: Ok. Now let's take a look at how a loop might work with a second task. For this task, imagine again that you have a small robot, but this time it's a turtle robot. The robot can now do three commands, so it already understands a bit more than the first robot. But this robot is set up so that when you give it a sequence of commands, it doesn't do the whole thing once and then stops, but it keeps doing the sequence you give it for as long as it's on. That is, you have to turn it off so that it stops executing the commands you gave it over and over again. That is, the sequence that you give the robot is executed as a loop as long as the robot is turned on. The task now is for you to figure out how to tell the robot to move in a square.

[The teacher turns to the class. The four answer choices, which were previously hidden, appear on Slide 5.]

T: I'm going to give you some time again to think about what you think is the right answer, and then we'll discuss it together.

[There is a 5 s pause; the teacher stands at the front and hardly moves; the students look forward at the video screen.]

T: So, are you guys ready? Has everyone thought of an answer? Great, then we'll do it just like before. I'll always say the answer choice, and you'll raise your hands if you think the answer is the right one. Who thinks Answer A is correct?

[Answer A is correct, so the usual reporting conditions apply here. ***Hand-raising situation that varies across four conditions;*** *Condition 1: 20% of the students raise their hands; Condition 2: 35% of the students raise their hands; Condition 3: 65% of the students raise their hands; Condition 4: 80% of the students raise their hands. As virtual classmates are raising their hands, some of the virtual learners look around to see who (else) is raising their hand.]*

T: Okay. Who thinks Answer B is the correct answer?

[Answer B is incorrect; for the correct answer A, the usual hand-raising conditions apply; between the remaining incorrect answers B, C, and D, the numbers of the remaining students who raise their hands (minus the distribution depending on the condition for Answer A) are randomly distributed.]

T: And who thinks Answer C is correct?

[Answer C is incorrect; for the correct answer A, the usual hand-raising conditions apply; between the remaining incorrect answers B, C, and D, the numbers of the remaining students who raise their hands (minus the distribution depending on the condition for Answer A) are randomly distributed.]

T: Okay. Who thinks Answer D is correct?

[Answer D is incorrect; for the correct answer A, the usual hand-raising conditions apply; between the remaining wrong answers B, C and D, the numbers of the remaining students who raise their hands (minus the distribution depending on the condition for Answer A) are randomly distributed.]

T: For this task, Answer A is correct. Let me show you why. Now I am the turtle robot again, and I just execute what is given there.

[The teacher imitates the robot's movements while explaining.]

T: 30 cm forward, turn to the right, 30 cm forward, turn to the right. Now I haven't done a square yet, but this time it's not just a sequence like that, but the robot executes it as a loop, which means it repeats it over and over again. So, I keep going, then I can start again from the beginning. 30 cm forward, to the right, 30 cm forward, to the right. Now the robot would have finished a square, but if you don't turn it off, it just keeps going and keeps doing the same square all the time because it keeps doing this sequence, that is, it keeps doing it as a loop. Do you all understand this?

[Students nodding.]

Scene 4 – Summary and conclusion (duration approximately 1 min)

T: That's what we've learned today about how computers think, which is important for us to know so that they can understand what we tell them.

[The teacher points again to the video screen, where Slide 5 with the last task is still visible.]

T: In the case here: If I were to tell the turtle robot, "Go in a rectangle or go in a square," it wouldn't understand me. Because the robot only knows three commands. It can only understand when it should turn to the right, when it should turn to the left, or when it should drive 30 cm forward. If I tell it something it doesn't know, for example, drive in a square, then it can't do that. That's why it's our job as programmers to break down what we want to tell the computer, or in this case the robot, in such a way that it can understand our commands. So we break it down into individual commands, and then we arrange them as a sequence, which means the robot does the commands one after the other, or we arrange the sequence as a loop. This means that this series of commands is then executed repeatedly. Do you understand what I mean?

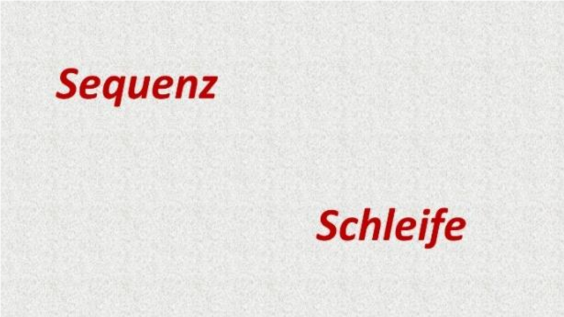
[Students nodding.]

T: Okay, I'm glad to hear that. Then I hope you've been able to learn a little bit about how computers think today, and I look forward to seeing you next time!

Overview of Presentation Slides

Original | **English Translation**

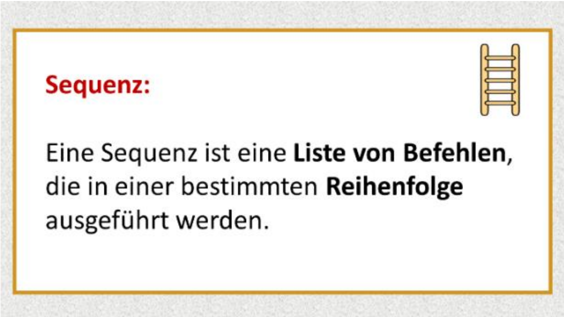
Slide 1



Sequence

Loop

Slide 2



Sequence:

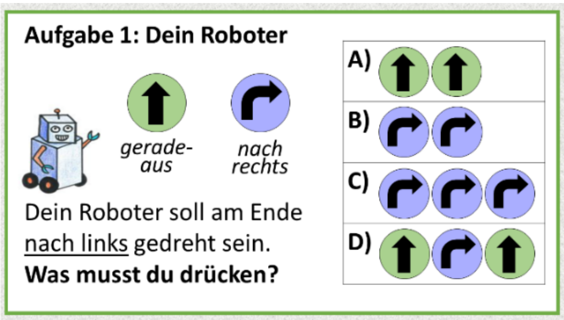
A sequence is a list of commands that are to be executed in a certain order.

Slide 3

Loop:

A loop is a list of commands that are executed repeatedly after each other.

Slide 4

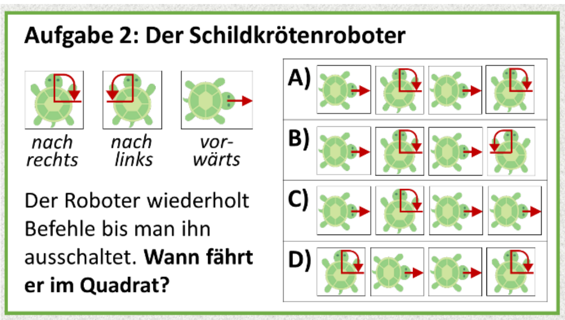


Exercise 1: Your robot

[keys:] *straight / to the right*

Your robot should be turned <u>to the left</u> in the end. **Which keys do you need to press?**

Slide 5



Exercise 2: The turtle robot

[keys:] *to the right / to the left / straight*

The robot repeats commands until you turn it off. **When does it move in a square?**