Universität Duisburg-Essen
Fakultät für Bildungswissenschaften
Lehrstuhl für Lehr-Lernpsychologie

Die Rolle und das Potenzial räumlicher Fähigkeiten für den Studienerfolg in MINT-Fächern

Dissertation zur Erlangung des Grades Dr. phil.
vorgelegt von **Nils Nolte**

Erstgutachter: Prof. Dr. Dr. hc. Detlev Leutner, Universität Duisburg-Essen
Zweitgutachter: Prof. Dr. Jörn Sparfeldt, Universität des Saarlandes

Tag der mündlichen Prüfung: 11. Juli 2023

DOI: 10.17185/duepublico/81963
URN: urn:nbn:de:hbz:465-20240517-064428-0

Inhalt

Danksagung

An dieser Stelle möchte ich mich bei allen Personen herzlichst bedanken, die maßgeblich zum Gelingen dieser Arbeit beigetragen haben.

Zuallererst gebührt Dir, Detlev Leutner, mein ganz besonderer Dank. Du hast mir nicht nur diese Möglichkeit geboten, meine Dissertation bei Dir anzufertigen, sondern mir auch viele Freiheiten bei der inhaltlichen Ausgestaltung innerhalb des Projekts und darüber hinaus gegeben. Danke für Deine kontinuierliche breite und wertvolle Unterstützung fachlicher wie organisatorischer Natur, Dein stets offenes Ohr bei jeglichen Problemen und die Fülle an bei Weitem nicht selbstverständlichen Möglichkeiten und Chancen, die Du mir geboten hast. Ich reihe mich in die lange Liste Promovierender ein, die Dir viel verdanken!

Auch Ihnen, Herr Sparfeldt, möchte ich für Ihr Interesse an meiner Arbeit und die Bereitschaft, diese Arbeit als Zweitgutachter zu betreuen, danken. Die unkomplizierte Zusammenarbeit mit Ihnen hat mir Freude bereitet und den Prozess deutlich vereinfacht!

Danke auch Dir, Jens Fleischer – nicht nur in der Projektphase hast Du mich mit deiner umfangreichen Erfahrung unterstützt, sondern mir auch darüber hinaus bei all meinen Fragen zur Seite gestanden. Es freut mich, dass das auch über Projekt und Uniwechsel hinaus so geblieben ist!

Weiterhin gilt mein Dank der Vielzahl an Kolleginnen und Kollegen, die mich auf dem Weg hierher begleitet haben. Hier möchte ich zentral meine Projektkollegen, die ALSTERaner Salome, Seb, Vero, Flo und Christian (in Projektreihenfolge) und Vanessa als Koordinatorin, nennen, mit denen ich (mindestens!) drei schöne Jahre in unserem gemeinsamen Bürokonglomerat verbringen durfte. Ich bin immer gern ins Büro gekommen und habe durch euch auch interdisziplinäre Perspektiven gewonnen, die meinen Horizont unglaublich erweitert haben. Sowohl diese Arbeit als auch ich wären ärmer ohne unsere gemeinsame Zeit. Außerdem gebührt mein Dank auch dir, Sabrina Hilz, die du immer ein offenes Ohr für meine Sorgen und Nöte hattest und mir dieselben organisatorischen und bürokratischen Prozesse auch ein siebtes Mal mit einem Lächeln erklärt hast.

Meine unendliche Dankbarkeit gilt abschließend meiner Familie: Meinen Eltern, Andrea und Hartmut – ihr habt mich Zeit meines Lebens gefördert und bei allem unterstützt, was ich mir in den Kopf gesetzt habe, und wart in allen Höhen und Tiefen für mich da. Und meiner Ehefrau und Lebenspartnerin Katrin – deine Liebe und geduldige Unterstützung waren mir zentrale Stütze während der gesamten Promotion und darüber hinaus, aber gerade in der sehr arbeitsreichen und beruflich wie privat turbulenten letzten Phase. Ohne euch hätte ich es niemals so weit geschafft. Danke, dass ihr Teil meines Lebens seid.

Zusammenfassung

Räumliche Fähigkeiten sind eine zentrale Facette in vielen Intelligenzmodellen, die auch in vielen Anwendungsbereichen von Bedeutung ist. Beispielsweise zeigten zahlreiche Studien, dass räumliche Fähigkeiten einen starken Prädiktor für sowohl Interesse an als auch Leistung in naturwissenschaftlich-technischen Fächern darstellen (bspw. Wai et al., 2009). Aufgrund der häufig gefundenen Überlappung von räumlichen Fähigkeiten mit der generellen Intelligenz *g* (Lohman, 1996) und fluider Intelligenz, ihrem stärksten Indikator (Carroll, 1993) war es aber bisher in wenigen Studien möglich, die Bedeutung räumlicher Fähigkeiten für Leistungen in naturwissenschaftlich-technischen Fächern von Einflüssen anderer Intelligenzfacetten wie dem fluiden Denken zu trennen, da nur wenige Studien zur Kontrolle auch Tests anderer Intelligenzfacetten mit erhoben. Ein zusätzliches Problem stellt hier dar, dass auch Tests räumlicher Fähigkeiten (wie Tests aller Intelligenzfacetten) immer auch eine *g*-Komponente mit messen (bspw. Carroll, 1993; Spearman, 1904), also auch innerhalb eines Tests nicht zweifelsfrei gesagt werden kann, ob die Testleistung die abhängige Variable vorhersagt, weil der Test räumliche Fähigkeiten misst oder weil er *g* misst und die räumliche Komponente für die Vorhersage kaum von Bedeutung ist. Die vorliegende Arbeit befasst sich mittels zwei unterschiedlicher Herangehensweisen mit diesem Problem: In der ersten vorgestellten Studie wurde der mentale Rotationstest (Vandenberg & Kuse, 1978) psychometrisch dahingehend untersucht, welche kognitiven Prozesse bei seiner Bearbeitung eine Rolle spielen – konkret in Bezug auf die beschriebene Problematik, inwieweit räumliche Fähigkeiten über *g* hinaus die Testleistung vorhersagen. In der zweiten und dritten Studie hingegen werden studienbezogene Leistungsmaße (Fachwissenserwerb und Klausurnoten) mittels räumlicher Fähigkeiten (gemessen anhand des in der ersten Studie untersuchten mentalen Rotationstests) vorhergesagt und dabei für fluide Intelligenz kontrolliert wurde. Die erste Studie konnte hierbei zeigen, dass die Leistung im mentalen Rotationstest in der Tat sowohl von *g*, aber eben auch darüber hinaus von räumlichen Fähigkeiten vorhergesagt wurde, er also beide als voneinander abgrenzbare Komponenten misst. Die zweite und dritte Studie belegten hingegen, dass räumliche Fähigkeiten substanziell die untersuchten Studienleistungen voraussagten – und zwar durch sowohl ihre *g*-Komponente (was sich darin äußerte, dass die durch einen Test fluiden Denkens vorhergesagte Varianz vollständig auch durch den mentalen Rotationstest erklärt wurde) als auch darüber hinaus durch ihre räumliche Komponente (da der mentale Rotationstest auch über die mit fluidem Denken überlappende erklärte Varianz noch signifikant mehr Varianz erklärte).

Neben der beschrieben psychometrischen Problematik wurde der Zusammenhang räumlicher Fähigkeiten mit Leistungen in naturwissenschaftlich-technischen Fächern bisher nur entweder für jeweils ein einzelnes Fach gezeigt (typischerweise mittels schwer vergleichbarer Leistungsmaße; bspw. Sorby et al., 2018) oder nur deskriptiv durchschnittliche räumliche Fähigkeiten

zwischen Fächern verglichen, ohne konkrete Leistungsmaße zu berücksichtigen (bspw. Shea et al., 2001). Entsprechend war es bisher nicht möglich, die tatsächlichen Zusammenhänge zwischen räumlichen Fähigkeiten und fachspezifischen Leistungen zwischen unterschiedlichen Fächern zu vergleichen, um so belastbare Aussagen darüber treffen zu können, in welchen Fächern räumliche Fähigkeiten für die tatsächliche Leistung (statt nur der bloßen Fachwahl oder binärer Erlangung eines Abschlusses) besonders relevant sind. Um diese Frage zu beantworten, untersuchte die zweite Studie der vorliegenden Arbeit drei naturwissenschaftlich-technische Fächer und ein gesellschaftswissenschaftliches Fach, um die Bedeutung räumlicher Fähigkeiten für den Fachwissenserwerb im ersten Semester dieser Studiengänge zu vergleichen. Insgesamt zeigte sich das erwartete Muster: In den naturwissenschaftlich-technischen Fächern war der Zusammenhang deutlich stärker (mit weiteren Abstufungen innerhalb der Gruppe) als in der Gesellschaftswissenschaft. Allerdings stellte sich hierbei heraus, dass sich die Bedeutung für den *universitären* Fachwissenserwerb (unter Kontrolle von Vorwissen und Abiturnote) zwischen den vier Fächern tatsächlich nicht signifikant unterschied. Fachunterschiede zeigten sich hingegen deutlich auf den indirekten Pfaden des berechneten Mediationsmodells, gerade hinsichtlich der Bedeutung räumlicher Fähigkeiten für den *vor-universitären* (typischerweise also schulischen) Fachwissenserwerb. Insgesamt konnte diese zweite Studie somit einerseits endlich zeigen, inwieweit sich unterschiedliche Fächer hinsichtlich des Zusammenhangs räumlicher Fähigkeiten mit fachspezifischen Leistungsmaßen unterscheiden, und eröffnete andererseits eine neue Fragestellung hinsichtlich des Unterschieds zwischen universitärem und schulischem Lernen in Bezug auf die Nutzung und Notwendigkeit räumlicher Fähigkeiten.

Nachdem der korrelative Zusammenhang zwischen räumlichen Fähigkeiten und Fachwissenserwerb in spezifischen Fächern sowohl in der hier vorgestellten zweiten Studie als auch in früheren Studien bereits demonstriert wurde, sollte die dritte Studie der vorliegenden Arbeit diesen nun zur kausalen Verbesserung des Fachwissenserwerbs einer Stichprobe von Ingenieursstudierenden nutzen. Hierbei sollte ein Videospieltraining als in der Literatur bisher vernachlässigte dritte Kategorie räumlicher Fähigkeitstrainings nach Uttal et al. (2013) verwendet werden, um auch die Effektivität dieser Trainingskategorie für den Transfer auf fachliche Leistungsmaße zu evaluieren. Es zeigte sich, dass das Videospieltraining die räumlichen Fähigkeiten der Studierenden verbesserte, wobei sich kein signifikanter Unterschied im Trainingseffekt zwischen einem räumlich sehr anspruchsvollen und einem räumlich weniger anspruchsvollen Spiel ergab. Die trainierten räumlichen Fähigkeiten sagten anschließend sowohl den Fachwissenserwerb als auch (mediiert über den verbesserten Fachwissenserwerb) die Klausurnoten der Studierenden vorher – erneut unter Kontrolle von fluider Intelligenz und fachlichem Vorwissen.

Abstract

Spatial ability is a central facet in many models of intelligence, and it is also important in many applicational contexts. For example, numerous studies showed that spatial ability is a strong predictor of both interest and achievement in science and technology disciplines (e.g., Wai et al., 2009). However, an overlap of spatial ability with general intelligence *g* (Lohman, 1996) and fluid intelligence, its strongest indicator (Carroll, 1993), is frequently found. Few studies have been able to separate the importance of spatial ability for performance in science and technology disciplines from effects of other facets of intelligence, such as fluid intelligence, because few studies have included corresponding tests as control variables. An additional problem is that tests of spatial ability (like tests of all facets of intelligence) always measure a *g*-component as well (e.g., Carroll, 1993; Spearman, 1904). Thus, even within a spatial test, it cannot be said beyond doubt whether test performance predicts the dependent variable because the test measures spatial ability or because it measures *g* and the spatial component is of little importance for prediction. The present work addresses this problem using two different approaches. On the one hand, in the first study presented the mental rotation test (Vandenberg & Kuse, 1978) was psychometrically investigated to determine which cognitive processes play a role in completing it – specifically, to which extent spatial ability beyond *g* predict test performance. In the second and third studies, on the other hand, study-related performance measures (content knowledge acquisition and exam grades) were predicted using spatial ability (measured using the mental rotation test investigated in the first study), controlling for fluid intelligence. The first study showed that performance on the mental rotation test was indeed predicted by both *g* and spatial ability, which means the test measured both as distinct components. The second and third studies demonstrated that spatial ability substantially predicted study performance through both their *g*-component (as evidenced by the fact that the variance explained by a test of fluid intelligence was fully explained by the mental rotation test as well) and their spatial component (as the mental rotation test explained significant variance beyond fluid intelligence).

In addition to the psychometric issues described above, the relationship of spatial ability with performance in science and technology disciplines has only been shown either for a single discipline at a time (typically using performance measures that are difficult to compare; e.g., Sorby et al., 2018) or only descriptively compared average spatial ability between disciplines without considering specific performance measures (e.g., Shea et al., 2001). Accordingly, it has not been possible to compare the actual relationships between spatial ability and discipline-specific performance between different disciplines to make robust statements about which disciplines spatial ability are particularly relevant for actual performance (rather than mere choice of study discipline or binary attainment of a degree). To answer this question, the second study in this paper examined three science and technology disciplines and one social science discipline to compare the importance of spatial skills for content knowledge acquisition in the first semester of these courses. Overall, the expected pattern emerged: the correlation was significantly stronger in the science and technology disciplines (with further variance within that group) than in the social science discipline. However, it turned out that the importance for *university* content knowledge acquisition (controlling for prior knowledge and GPA) actually did not differ significantly among the four disciplines. Instead, discipline differences were clearly evident on the indirect paths of the calculated mediation model, especially with respect to the importance of spatial skills for *pre-university* (typically in school) content knowledge acquisition. Overall, this second study was thus able to finally show the extent to which different disciplines differ with respect to the association of spatial skills with discipline-specific performance measures, and further opened up a new line of inquiry regarding the difference between university and school learning with respect to the use and necessity of spatial skills.

Having already demonstrated the correlative relationship between spatial skills and content knowledge acquisition in specific disciplines both in the second study presented here and in previous studies, the third study of the present work was now intended to use this to causally improve content knowledge acquisition in a sample of engineering students. Here, video game training was used as the third type of spatial ability training according to Uttal et al. (2013) that has been neglected in the literature so far. The goal was to evaluate the effectiveness of this type of spatial ability training for achieving transfer effects on discipline-specific performance measures. Video game training was found to improve students' spatial ability, with no significant difference in training effect between a highly spatially challenging game and a less spatially challenging game. The trained spatial skills subsequently predicted both content knowledge acquisition and (mediated by improved content knowledge acquisition) students' exam grades – again controlling for fluid intelligence and prior knowledge.

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Räumliche Fähigkeiten sind eine zentrale Komponente in vielen Intelligenzmodellen. Bereits 1938 nahm Thurstone sie als Primärfähigkeit in sein Intelligenzmodell auf, auch in späteren Modellen wie dem von Carroll (1993) oder der in jüngerer Zeit erstellten zusammenfassenden Taxonomie kognitiver Fähigkeiten, dem Cattell-Horn-Carroll-Modell (CHC-Modell; Flanagan & Dixon, 2013), finden sich räumliche Fähigkeiten als eigenständiger Faktor. Räumliche Fähigkeiten lassen sich in weitere Unterfacetten aufgliedern, unter anderem auch die Fähigkeit, die umgangssprachlich als *räumliches und bildliches Vorstellungsvermögen* bezeichnet wird. Im wissenschaftlichen Kontext wird hier eher von der Facette *räumlicher Visualisierungsfähigkeit* (*spatial visualization*) gesprochen, die "Prozesse des Erfassens, Kodierens und Manipulierens räumlicher Formen" umfasst (Carroll, 1993, S. 309). Hierzu kann in Anwendungskontexten auch die Erschaffung und Manipulation mentaler Modelle im visuell-räumlichen Teil des Arbeitsgedächtnisses gezählt werden (Lohman, 1996).

Räumliche Visualisierungsfähigkeit hat offensichtliche Anwendungsmöglichkeiten in Beruf und Alltag. Gerade in naturwissenschaftlich-technischen Berufen nimmt die Arbeit mit Modellen und Visualisierungen eine zentrale Rolle ein. Zahlreiche Befunde belegen, dass räumliche Fähigkeiten im Allgemeinen und *spatial visualization* im Besonderen mit sowohl Leistung als auch Interesse im MINT-Bereich (MINT: Mathematik, Informatik, Naturwissenschaft und Technik; im Englischen STEM: Science, Technology, Engineering, and Mathematics) zusammenhängen (Khine, 2017). Beispielsweise zeigten Wai et al. (2009) in längsschnittlichen Studien, dass die räumlichen Fähigkeiten von Jugendlichen in enger Beziehung zu ihren Fachpräferenzen in der Schule, der Wahl eines Studiengangs fünf Jahre später sowie dem weitere zehn Jahre später ausgeübten Beruf stehen: Personen mit stärkeren räumlichen Fähigkeiten gaben häufiger MINT-Fächer als Lieblingsschulfach an, begannen häufiger MINT-Studiengänge und arbeiteten später häufiger in Berufen mit MINT-Bezug. Entsprechend wird deutlich, dass räumliche Fähigkeiten für viele Abschnitte des (Aus-)Bildungswesens von Bedeutung sind.

Obwohl eine wachsende Zahl Studien einen Zusammenhang zwischen räumlichen Fähigkeiten und Fachpräferenzen sowie unterschiedlichen Leistungsmaßen gerade im MINT-Kontext zeigt, lässt die bisherige Forschung dennoch einige Lücken, die in den folgenden Kapiteln herausgearbeitet werden. So untersuchen viele der bisherigen Studien die räumlichen Fähigkeiten als einzigen Prädiktor, ohne weitere kognitive Variablen zu berücksichtigen. Dies ist dahingehend besonders kritisch, dass Tests räumlicher Fähigkeiten auch substanziell generelle Intelligenz mit messen – bei alleiniger Betrachtung eines räumlichen Fähigkeitstests als Prädiktor lässt sich also nicht zweifelsfrei sagen, ob dessen Vorhersagekraft tatsächlich auf die räumliche Komponente des Tests zurückzuführen ist oder Resultat der mit gemessenen generellen Intelligenz, was die Interpretierbarkeit der Befunde einschränkt. Weiterhin ist es anhand bisheriger Studien nicht möglich, differentiell zu beurteilen, inwieweit sich unterschiedliche MINT-Fächer in Hinblick auf die Stärke des Zusammenhangs von räumlichen Fähigkeiten mit Leistungen in den jeweiligen Fächern unterscheiden. Dadurch ist eine Beurteilung, in welchen Fächern räumliche Fähigkeiten von besonders hoher Bedeutung sind und sich Interventionen entsprechend besonders lohnen würden, nur eingeschränkt möglich. Weiterhin sind unterstützende Interventionen für räumliche Fähigkeiten bisher nicht über das gesamte Spektrum sich als effektiv erwiesen habender Arten auch im MINT-Kontext erprobt worden: Der Ansatz der videospielbasierten Interventionen wurde in diesem Kontext bisher vernachlässigt.

Die in der vorliegenden Arbeit vorgestellten Studien leisten einen Beitrag, um diese Lücken zu schließen.

1.1 Räumliche Fähigkeiten

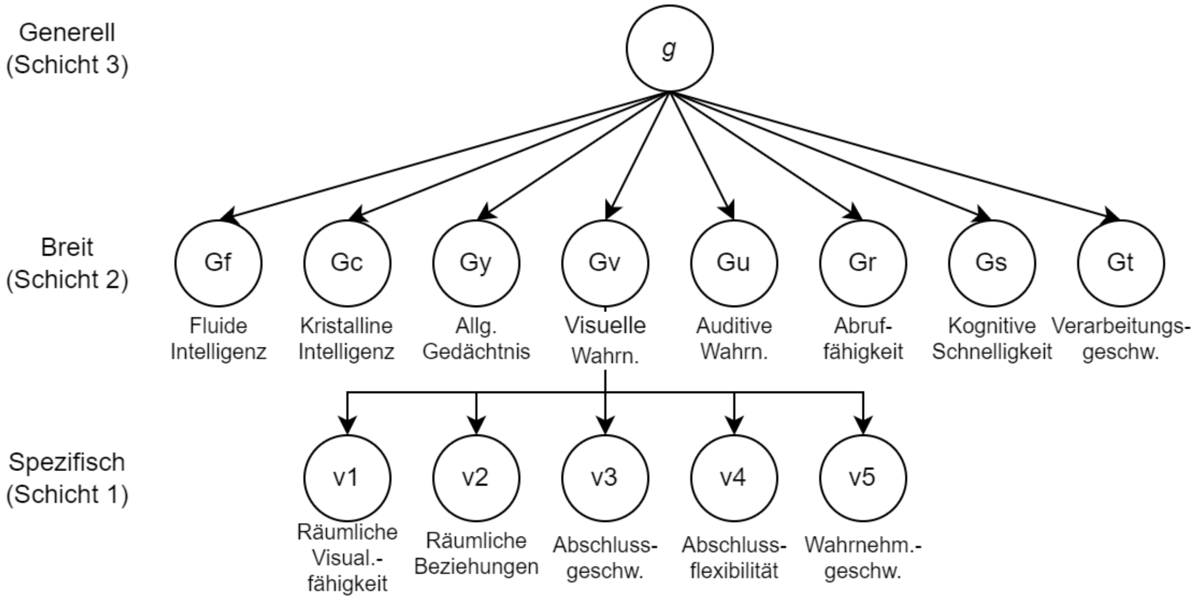
Die in dieser Dissertation dargestellten Studien folgen der faktoranalytischen Klassifikation kognitiver Fähigkeiten von Carroll (1993), die in Abbildung 1.1 dargestellt ist. In dieser Klassifikation werden die räumlichen Fähigkeiten in fünf spezifische Facetten aufteilt:

1. Räumliche Visualisierungsfähigkeit (im Englischen *spatial visualization ability*): “Prozesse des Erfassens, Kodierens und Manipulierens räumlicher Formen” (Carroll, 1993, S. 309).
2. Räumliche Beziehungen (*spatial relations*): eng verwandt mit räumlicher Visualisierungsfähigkeit, deckt aber eher einfachere Aufgaben unter Zeitdruck ab, wohingegen räumliche Visualisierungsfähigkeit komplexere Aufgaben ohne Zeitdruck beinhaltet.
3. Abschlussgeschwindigkeit (*closure speed*): Erkennen unbekannter, verschleierter Formen.
4. Abschlussflexibilität (*closure flexibility*): visuelle Suche nach einer vorgegebenen Form in ablenkendem oder verschleierndem Kontext.
5. Wahrnehmungsgeschwindigkeit (*perceptual speed*): typischerweise entweder Finden einer vorgegebenen Form ohne ablenkenden Kontext oder Vergleich mehrerer Abbildungen auf Identität.

Die erste Facette der räumlichen Visualisierungsfähigkeit steht im Fokus der folgenden Arbeit, da sie von besonderer Bedeutung für MINT-Fächer ist (für ein Review siehe Uttal & Cohen, 2012). Viele MINT-Fächer erfordern die Anwendung räumlicher Visualisierungsfähigkeit in typischen Aufgaben, wie in Kapitel 1.2 ausführlicher dargestellt wird.

Abbildung 1.1

Carrolls Drei-Schichten-Modell der Intelligenz



Notiz. Adaptiert nach Carroll (1993).

Räumliche Fähigkeiten hängen zudem eng mit fluider Intelligenz und logischem Schlussfolgern zusammen (im Englischen sowie im Folgenden dieser Arbeit als *Reasoning* zusammengefasst und abgekürzt), die als stärkste Indikatoren für den generellen Intelligenzfaktor *g* gelten (Carroll, 1993), sowie mit *g* selbst (Lohman, 1996). Aufgrund dessen ging Spearman (Spearman & Wynn Jones, 1950) sogar so weit, die Trennbarkeit räumlicher Fähigkeiten von *g* in Frage zu stellen und Tests räumlicher Fähigkeiten lediglich als ungenaue Messungen von *g* zu bezeichnen. Jüngere Modelle wie beispielsweise das CHC-Modell (Flanagan & Dixon, 2013) widersprechen dem hingegen entschieden und enthalten einen von *g* getrennten Faktor räumlicher Fähigkeiten. Lohman (1996) argumentierte, dass die besonders starke Überlappung von räumlichen Fähigkeiten mit *g* daher komme, dass Tests beider Fähigkeiten die Generierung, Manipulation und Aufrechterhaltung mentaler Modelle und damit sehr ähnliche kognitive Prozesse erfordern. Diese Prozesse im Umgang mit mentalen Modellen seien sowohl für Reasoning-Aufgaben als auch für räumliche Aufgaben gleichermaßen zentral – und spielten sich vorwiegend im visuell-räumlichen Teil des Arbeitsgedächtnisses (visuell-räumlicher Notizblock; Baddeley, 2003) ab. Diese Visualisierungsfähigkeit stelle dementsprechend die Grundlage für erfolgreiches logisches Schlussfolgern, ja das Denken an sich dar: “Thought without imagery would be like prose without metaphor” (Lohman, 1996, S. 12). Das Vorstellen von Bildern jedweder Art fällt typischerweise unter die eingangs bereits beschriebene räumliche Visualisierungsfähigkeit, also eine Facette räumlicher Fähigkeiten. Lohman argumentierte, dass man somit Spearmans Aussage auch umdrehen und befinden könnte, “that measures of G are largely unreliable measures of the ability to generate and transform different types of mental models in working memory” (Lohman, 1996, S. 12). Die erste in dieser Dissertation vorgestellte Studie behandelt die Frage näher, welche kognitiven Prozesse beim Bearbeiten räumlicher Tests eine Rolle spielen und von welchen Aufgabeneigenschaften (mit Fokus auf die Rotationsachsen eines mentalen Rotationstests) die bei der Bearbeitung der Aufgaben angewandten Strategien abhängen. Hierbei wird auch das Zusammenspiel von Reasoning und räumlichen Fähigkeiten bei der Bearbeitung solcher Tests näher beleuchtet.

1.1.1 Mentale Rotationstests

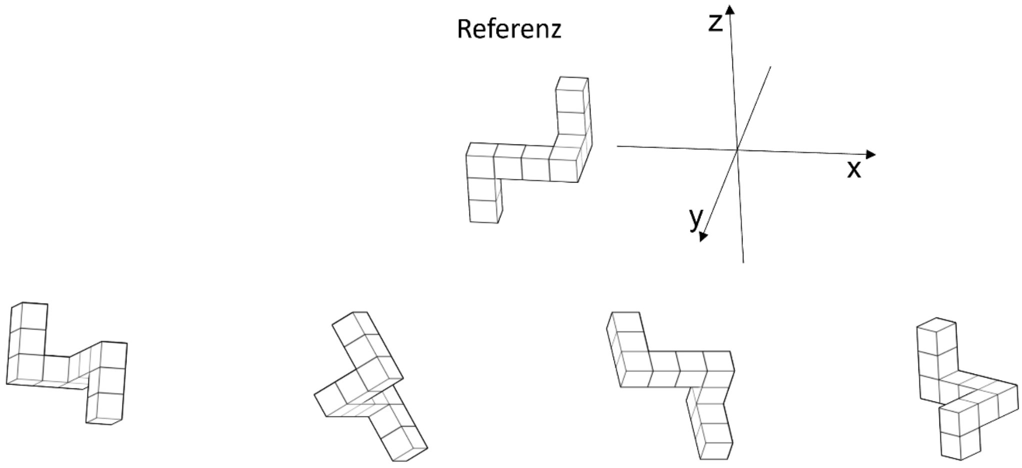
In der Forschung zu räumlichen Fähigkeiten sind mentale Rotationstests weit verbreitet, da sie einerseits leicht erstell- und einsetzbar sind (beispielsweise aufgrund der Fülle möglicher Stimuli, inklusive der frei verfügbaren Datenbank von Peters & Battista, 2008) sowie andererseits typischerweise mit räumlicher Visualisierungsfähigkeit eine in vielen Zusammenhängen relevante Fähigkeit erfassen.

Je nach Zusammensetzung der Items können mentale Rotationstests allerdings sowohl den Faktor räumliche Visualisierungsfähigkeit als auch den Faktor räumliche Beziehungen messen. Die Zuordnung eines Tests zu einem der beiden Faktoren hängt maßgeblich von der Komplexität der Stimuli und Aufgabenstellung sowie dem Vorhandensein eines Zeitlimits ab. Während komplexe Aufgaben ohne Zeitlimit (*Power*-Tests) eher auf räumliche Visualisierungsfähigkeit laden, laden einfachere Aufgaben unter Zeitdruck eher auf räumliche Beziehungen (*Speed*-Tests; Carroll, 1993). Zentrale Itemcharakteristika, die die Komplexität mentaler Rotationstests beeinflussen, werden in Kapitel 1.1.2 näher beschrieben. Da es in Hinblick auf die Aufgabenkomplexität viele Abstufungen gibt, lässt sich bei einzelnen Tests nicht immer eindeutig feststellen, welcher der beiden Facetten sie zuzuordnen sind. Umgekehrt führt diese Problematik auch dazu, dass sich diese beiden Facetten in vielen Studien nicht klar trennen ließen (Carroll, 1993).

Mentale Rotationstests werden mit einer großen Bandbreite an Stimuli durchgeführt. Diese reichen von einfachen zweidimensionalen Formen wie beispielsweise im Card Rotation Test (Ekstrom et al., 1976) bis hin zu Alltagsgegenständen (bspw. Rahe et al., 2021) oder menschlichen Körpern (Doyle & Voyer, 2018). Klassische Stimuli sind die Würfelfiguren von R. N. Shepard und Metzler (1971), die in jüngerer Zeit auch in einer digitalen Stimulusdatenbank (Peters & Battista, 2008) verfügbar sind. Diese dreidimensionalen Figuren bestehen typischerweise aus neun bis zehn aneinander gereihten Würfeln mit drei 90°-Winkeln in ihrer Kette. Einzelne Segmente bestehen jeweils aus zwei bis fünf Würfeln, wobei die beiden mittleren Segmente immer aus mindestens dreien bestehen. Eine Beispielaufgabe, die diese Würfelfiguren nutzt, ist in Abbildung 1.2 dargestellt und erläutert. Typischerweise bestehen Rotationsaufgaben aus einem Referenzstimulus und einem oder mehreren Vergleichsstimuli, unter denen die gedrehten Versionen des Referenzstimulus gefunden werden sollen. Aber auch für die Darbietung der Stimuli gibt es unterschiedliche Varianten. Die beiden verbreitetsten sind das Shepard und Metzler-Paradigma sowie das Vandenberg und Kuse-Paradigma. Im Shepard und Metzler-Paradigma wird nur ein Vergleichsstimulus neben dem Referenzstimulus gezeigt und es soll eine ja/nein-Entscheidung getroffen werden, ob die beiden Stimuli rotierte Versionen voneinander sind. Dieses Paradigma wird üblicherweise unter Zeitdruck bearbeitet und die Reaktionszeit in unterschiedlichen Versuchsbedingungen (bspw. Vergleichsstimulus rotiertes Objekt versus nicht rotiert, sondern verschieden) als abhängige Variable betrachtet. Die Betrachtung der Genauigkeit ist aufgrund der typischen sehr hohen Lösungsquoten von etwa 95% (z. B. bei R. N. Shepard & Metzler, 1971) nicht sinnvoll. Im Vandenberg und Kuse-Paradigma hingegen werden typischerweise vier Vergleichsstimuli gezeigt, von denen zwei die zu identifizierenden rotierten Versionen des Referenzstimulus sind, während die anderen beiden Distraktoren darstellen. Distraktoren in beiden Paradigmen sind meist gespiegelte (und zusätzlich rotierte) Versionen des Referenzstimulus, manchmal auch strukturell leicht unterschiedliche Objekte. Das Vandenberg und Kuse-Paradigma kann mit und ohne Zeitdruck durchgeführt werden und resultiert in der Regel in deutlich niedrigeren Lösungsquoten von durchschnittlich 40-50% (bspw. Krüger & Suchan, 2016; Peters, 2005; Vandenberg & Kuse, 1978. Entsprechend kann in diesem Paradigma auch die Lösungsquote als abhängige Variable betrachtet werden, ohne dass die Messung unter einem starken Deckeneffekt leidet. Auch sind die Ergebnisse weniger durch Ratewahrscheinlichkeit verzerrt, da hier die Aufgaben nur als richtig gewertet werden, wenn beide Zielstimuli gefunden wurden (was bei zwei Zielen und zwei Distraktoren einer 1/6 Chance entspricht). Im Anwendungskontext wird darum besonders das Vandenberg und Kuse-Paradigma verwendet, um individuelle Unterschiede in der Testleistung als Prädiktor für unterschiedliche Praxisaufgaben im MINT-Bereich in Bildung und Beruf nutzen zu können. Diese reichen von Schul- und Studienleistungen (bspw. Ganley et al., 2014; Sorby & Baartmans, 2000) wie dem Erkennen identischer Molekülstrukturen in der Chemie (Stieff et al., 2014) über die Zuordnung von Schnittbildern zu 3D-Modellen in der Zahnmedizin (Hegarty et al., 2009) bis hin zur Präzision bei der Durchführung einer Endoskopie (Rogister et al., 2021).

Abbildung 1.2

Beispielaufgabe des Vandenberg und Kuse Mentalen Rotationstests



Ziel 1
125° z-Achse

Ziel 2
170° z-Achse +
60° x-Achse
= 172° schiefe Achse

Distraktor 1
170° z-Achse +
140° x-Achse
= 176° schiefe Achse
& gespiegelt

Distraktor 2
50° z-Achse
strukturell
unterschiedlich

Notiz. Ebenfalls dargestellt und beschrieben sind die verschiedenen Möglichkeiten für Rotationen um Kardinal- oder schiefe Achsen sowie die beiden Arten von Distraktoren. Zielstimulus 1 ist 125° um die z-Achse (die vertikal in der Bildebene liegende Kardinalachse) gedreht, Zielstimulus 2 erst 170° um die z- und dann 60° um die x-Achse (also die horizontal und vertikal in der Bildebene liegenden Kardinalachsen), in Summe somit 172° um eine schiefe Achse. Distraktor 1 ist eine gespiegelte und ebenfalls schief rotierte Version des Referenzstimulus (resultierend aus erst einer 170° Rotation um die z- und anschließend 140° um die x-Achse), während Distraktor 2 sich strukturell von den anderen Referenzstimulus unterscheidet (die einzelnen Segmente sind unterschiedlich lang), aber für einen direkten Vergleich mit dem Referenzstimulus ebenfalls um 50° um die z-Achse rotiert werden muss.

1.1.2 Charakteristika von mentalen Rotationstests

Zahlreiche Aufgabencharakteristika beeinflussen die Komplexität von mentalen Rotationsaufgaben und dadurch auch die von ihnen gemessene Facette räumlicher Fähigkeiten. Der wohl am konsistentesten gefundene Faktor ist die Größe des Rotationswinkels: Je weiter ein Objekt mental rotiert werden muss, desto länger dauert der Prozess (R. N. Shepard & Metzler, 1971). Dieser lineare Anstieg der Bearbeitungszeit (was einer konstanten Rotationsgeschwindigkeit entspricht) passt zur in Studien mit bildgebenden Verfahren (bspw. fMRT) gezeigten Ähnlichkeit der neuronalen Aktivitäten bei der Verarbeitung mentaler und manueller physischer Rotation von Objekten, da manuelle Rotation ebenfalls mit einer konstanten Geschwindigkeit geschieht (für ein Review siehe Zacks, 2008). Ähnliche Befunde wie für die Reaktionszeit fanden Boone und Hegarty (2017) auch für die Genauigkeit der Antworten. Weiterhin ist die Rotation von zweidimensionalen Objekten einfacher (also mit höherer Lösungsquote) und schneller (also mit geringerer Reaktionszeit) als die von dreidimensionalen Objekten (S. Shepard & Metzler, 1988). Die Art der benötigten Rotation (um eine der drei Kardinalachsen des Bildes oder um mehrere zugleich, was in entweder einer mehrschrittigen Rotation oder einer direkten Rotation um eine schiefe Achse resultiert) beeinflusst ebenfalls die Testleistung. Hier unterscheiden sich Befunde zur Reaktionszeit und zur Genauigkeit dahingehend, bei Rotation um welche Kardinalachse Rotationsaufgaben mit der kürzesten Reaktionszeit bzw. höchsten Lösungsquote gelöst werden. Gemein ist beiden Maßen allerdings, dass Rotation um Kardinalachsen sowohl schneller als auch akkurater ist als die Rotation um eine schiefe Achse (für Reaktionszeiten siehe Parsons, 1987; für Genauigkeit, siehe Stieff, 2007). Weitere Einflussfaktoren auf die Testleistung sind die Komplexität der zu rotierenden Objekte selbst (bspw. die Anzahl an Merkmalen wie Segmente und Knicke, die ein zu rotierendes Objekt aufweist), wobei Aufgaben mit komplexeren zu rotierenden Objekten seltener gelöst werden als solche mit simpleren Stimulusobjekten (bspw. Cooper, 1975; Heil & Jansen-Osmann, 2008; Stieff, 2007; Stieff et al., 2018), sowie die Vertrautheit der Probanden mit den Stimuli, wobei Aufgaben mit vertrauteren Stimuli wie Alltagsgegenständen häufiger korrekt gelöst werden als Aufgaben, bei denen unbekannte Stimulusobjekte rotiert werden müssen (bspw. Doyle & Voyer, 2018; Muto & Nagai, 2020; Stieff, 2007).

Zusammengefasst: Tests mit simplen, oft zweidimensionalen Stimuli und knappem Zeitlimit wie der Card Rotation Test fallen typischerweise unter räumliche Beziehungen, besonders wenn die Stimuli im Shepard und Metzler-Paradigma dargeboten werden. Werden hingegen unvertraute (und meist dreidimensionale) Stimuli mit komplexeren Merkmalen und variierenden Rotationsachsen verwendet und ohne Zeitlimit im Vandenberg und Kuse-Paradigma dargeboten (beispielsweise im klassischen Vandenberg und Kuse mentalen Rotationstest), lädt ein solcher Test typischerweise auf räumliche Visualisierungsfähigkeit.

Aufgrund seines Zusammenhangs mit Leistungen im MINT-Kontext fokussiert sich die vorliegende Arbeit auf den Faktor räumliche Visualisierungsfähigkeit. Um diesen Faktor zuverlässig zu messen, wurden in den Studien, die in der vorliegenden Arbeit dargestellt werden, die klassischen dreidimensionalen Würfelfiguren von Shepard und Metzler verwendet. Diese wurden im komplexeren Paradigma nach Vandenberg und Kuse ohne Zeitlimit dargeboten und zudem zur Lösung der Aufgaben Rotationen um wechselnde Kardinal- und schiefe Rotationsachsen gefordert. Hierdurch sollte die zur Messung räumlicher Visualisierungsfähigkeit benötigte Komplexität der Aufgaben gesichert werden. Die Nutzung wechselnder Rotationsachsen, inklusive schiefer Achsen, die sich nicht an der Form des zu rotierenden Objekts orientieren, stellen auch eine höhere ökologische Validität des Tests her: In der realen Welt sind Objekte, die es zu erfassen und zu vergleichen gilt, selten perfekt gerade orientiert, sondern liegen eher schief; sie werden von unterschiedlichen Perspektiven wahrgenommen, haben vielleicht sogar verdeckte Elemente.

Die Darbietung ohne Zeitlimit führt zudem auch dazu, dass der üblicherweise gefundene Geschlechtsunterschied zugunsten von Männern (für eine Metaanalyse siehe Voyer et al., 1995) deutlich reduziert wird (für eine Metaanalyse siehe Voyer, 2011). Dies ist wünschenswert, um mögliche Geschlechtseffekte als Störvariablen zu minimieren – auch wenn Geschlechtsunterschiede in MINT-Stichproben bereits schwächer ausgeprägt sein sollten als in der Gesamtbevölkerung, da die aus Shea et al.'s (2001) Ergebnissen hervorgehenden Selbstselektionseffekte bei der Studienfachwahl bereits dazu führen sollten, dass diejenigen Frauen, die ein MINT-Studium beginnen, überdurchschnittliche räumliche Fähigkeiten aufweisen.

1.2 Studienerfolg in naturwissenschaftlich-technischen Fächern

MINT-Studiengängen leiden seit Jahren unter enorm hohen Studienabbruchsquoten (bspw. Chen, 2009; Heublein et al., 2020; Price, 2010). Dem steht der zunehmende Fachkräftemangel als vielfach beklagtes Problem für Wirtschaftswachstum und Innovation gegenüber, besonders im MINT-Bereich (CBI, 2011, 2019; National Academies Committee on Science, Engineering, and Public Policy [COSEPUP], 2007). Dieser Diskrepanz zu begegnen ist eine maßgebliche Herausforderung für die Bildungssysteme (Uttal & Cohen, 2012) – besonders, aber nicht ausschließlich für die Universitäten.

Der von MINT-Studienabbrechenden am häufigsten genannte Grund sind Probleme mit der Bewältigung der Leistungsanforderungen ihres Fachs (Heublein & Schmelzer, 2018). Gezielte Unterstützung der Studierenden bei der Bewältigung der Leistungsanforderungen ist somit ein vielversprechender Ansatz, die Studienabbruchquote zu reduzieren. Hierfür werden Einblicke in Einflussfaktoren auf die Leistung der Studierenden aus bildungswissenschaftlicher Forschung benötigt. Darauf aufbauend können einerseits besonders talentierte Studienanwärter und Studierende ausgewählt beziehungsweise gefördert werden (im Sinne einer Exzellenzförderung) und andererseits Studierende mit schlechteren Startbedingungen gezielt beraten und unterstützt werden.

Ein zentrales Ziel eines Studiums ist der Erwerb von Kompetenzen und Wissen (York et al., 2015). Darauf aufbauend stellt das Fachwissen in den im Rahmen dieser Dissertation durchgeführten Studien die zentrale abhängige Variable dar.

1.2.1 Prädiktoren des Studienerfolgs

Klassische Prädiktoren für leistungsbezogene Maße des Studienerfolgs wie Fachwissenserwerb und Klausurnoten (die das erworbene Fachwissen formal messen sollen) sind Intelligenz und Abiturnote (bspw. Burton & Ramist, 2001; Neisser et al., 1996) sowie fachspezifisches Vorwissen (Formazin et al., 2011; Freyer et al., 2014; Hailikari, 2009). Während die beiden ersteren leicht erfassbar und fachübergreifend verwendbar sind, birgt die Erfassung fachspezifischen Vorwissens im Gegenzug mehr Vorhersagepotenzial für das spezifische Fach (für eine Metaanalyse siehe Kuncel et al., 2001). Beispielsweise fanden Formazin et al. (2011), dass fachlich relevantes Vorwissen aus den Bereichen Psychologie, Mathematik, Biologie und Englisch mehr Varianz in Psychologie-Klausurnoten erklärte als es Intelligenz tat (35% zu 10%), während Hailikari (2009) einen vergleichbaren Effekt von mathematischem Vorwissen und Abiturnote auf Klausurnoten in Mathematikkursen fand. Weiterhin zeigte Averbeck (2021), dass es für Studienanfänger in der Chemie ohne durch einen vorherigen Chemie-Leistungskurs erworbenes Vorwissen kaum möglich war, im ersten Semester das für das Bestehen der Klausur im typischen Einführungskurs zur Allgemeinen Chemie benötigte Fachwissen zu erwerben. Entsprechend ist der Einbezug fachspezifischen Vorwissens in die Vorhersage von mindestens leistungsbezogenen Studienerfolgsmaßen zu empfehlen, insofern eine Messung des fachbezogenen Vorwissens vorliegt bzw. möglich ist. Weitere psychologische Variablen wie die Persönlichkeit, Motivation und Interesse oder das akademische Selbstkonzept spielen ebenfalls eine Rolle für den Studienerfolg (bspw. Fleischer et al., 2019; Kryshko et al., 2020; Schnettler et al., 2020). Auf diese wird in der vorliegenden Arbeit allerdings nicht näher eingegangen.

Bezüglich der kognitiven Einflussfaktoren auf Studienerfolg sind räumliche Fähigkeiten in den vergangenen Jahren als weiterer Faktor in den Blick der Forschung gerückt. Besonders in MINT-Fächern spielen räumliche Fähigkeiten eine nicht zu vernachlässigende Rolle, wie beispielsweise Shea et al. (2001) und Wai et al. (2009) in Übersichtsarbeiten für eine große Bandbreite an Fächern und Zeitpunkten der Bildungslaufbahn im Längsschnitt zeigten. So stellten die beiden aufeinander aufbauenden Arbeiten dar, dass Jugendliche mit höheren räumlichen Fähigkeiten häufiger MINT-Fächer als Lieblingsschulfach nannten (und umgekehrt Jugendliche mit niedrigen räumlichen Fähigkeiten diese häufiger als am wenigsten liebstes Fach nannten), fünf Jahre später häufiger MINT-Studiengänge absolvierten und zehn Jahre später häufiger in Berufen im MINT-Bereich arbeiteten.

Auch für spezifische Fächer konnte der Zusammenhang zwischen räumlichen Fähigkeiten und höherem fachlichen Erfolg gezeigt werden, beispielweise im Ingenieurwesen (bspw. Duesbury & O'Neil, 1996; Sorby et al., 2018; für ein Review siehe Ha & Fang, 2016), in der Chemie (bspw. Talley, 1973; Wu & Shah, 2004), der Mathematik (bspw. Hegarty & Kozhevnikov, 1999; Mix et al., 2021), der Physik (bspw. Kozhevnikov et al., 2007; Pallrand & Seeber, 1984) oder der Biologie und Medizin (bspw. Hegarty et al., 2009; Rochford, 1985). Bei Betrachtung typischer Aufgaben in MINT-Berufen wird schnell offensichtlich, woher die vielfach gefundenen Verbindungen zu räumlichen Fähigkeiten rühren: Beispielsweise müssen Ingenieure anhand von Visualisierungen verstehen können, wie unterschiedliche Bauteile zusammenpassen und sich in Abhängigkeit voneinander bewegen, sich vorstellen, wie Komponenten auf begrenztem Raum am effizientesten verbaut werden können, oder komplexe topografische Karten visualisieren, erstellen und lesen und daraus die tatsächlichen Geländeverhältnisse und ihre Bedeutung für ein Bauprojekt ableiten (Sorby et al., 2013). Ebenso ist es für Chemiker wichtig, sich die Struktur von Molekülen anhand von Darstellungen vorstellen zu können (und umgekehrt solche Darstellungen produzieren zu können), um beispielsweise chemische Reaktionen wie Syntheseprozesse zu verstehen (Wu & Shah, 2004). Biologen und Ärzte müssen Querschnitte von Organen lesen, verstehen und anfertigen (Hegarty et al., 2009), oder Geologen anhand der durch Erosion veränderten Landschaft mental rekonstruieren können, wie diese Landschaft früher einmal ausgesehen haben mag (Hambrick et al., 2012). Die von diesen Aufgaben gestellten Anforderungen haben offensichtliche Überlappungen mit räumlichen Fähigkeiten im Allgemeinen und der Facette räumliche Visualisierungsfähigkeit im Besonderen. Insofern ist naheliegend, dass sich diese Facette als die für viele MINT-Disziplinen relevanteste herausgestellt hat (siehe bspw. Hsi et al., 1997; Lowrie et al., 2017; Miller & Halpern, 2013; Shea et al., 2001; Sorby, 2009; Wai et al., 2009; Wu & Shah, 2004; für ein Review siehe Uttal & Cohen, 2012).

Eine Reihe experimenteller Studien nutzte den Zusammenhang zwischen räumlichen Fähigkeiten und MINT-Erfolg, um mit Hilfe von räumlichen Fähigkeitstrainings Leistungen in MINT-Fächern zu verbessern (meistens in Form von Klausurnoten oder eigens entwickelten Fachwissenstests). Die meisten Befunde gibt es hierbei im Bereich der Mathematik in unterschiedlichen Bildungsabschnitten (bspw. Cheng & Mix, 2014; Judd & Klingberg, 2021; Lowrie et al., 2017; Lowrie et al., 2019; Mix et al., 2021) sowie im Bereich des Ingenieurwesens, hier vorwiegend im universitären Kontext (hauptsächlich von Sorby et al., bspw. 2013; 2018; 2010; sowie darauf aufbauenden Projekten, bspw. Benedičič et al., 2022; Grzybowski et al., 2014). In diesen Studien zeigten sich Transfereffekte des räumlichen Fähigkeitstrainings auch auf die jeweiligen fachbezogenen Maße. In Kapitel 1.3 wird näher auf verschiedene Arten räumlicher Fähigkeitstrainings sowie deren Nutzung im MINT-Kontext eingegangen.

1.2.2 Räumliche Fähigkeiten und Studienerfolg im Fächervergleich

Zusammenfassend kann festgehalten werden, dass es eine umfangreiche Befundlage für die Existenz eines robusten Zusammenhangs zwischen räumlichen Fähigkeiten und MINT-Erfolgen gibt, der anhand einiger Interventionsstudien auch als kausal angenommen werden kann. Allerdings gibt es bisher keine Studien, die die Stärke dieses Effekts zwischen unterschiedlichen Fächern verglichen haben, sodass keine klare differentielle Aussage darüber getroffen werden kann, in welchen MINT-Fächern sie bedeutsamer für Studienerfolgsmaße sind und in welchen weniger bedeutsam. Die von Shea et al. (2001) gefundenen Muster kognitiver Fähigkeiten (in Form durchschnittlicher Ausprägungen mathematischer, verbaler und räumlicher Fähigkeiten) bei Bachelor-Absolvierende unterschiedlicher Fächer lassen darauf schließen, dass sich die Bedeutung gerade der räumlichen

Fähigkeiten auch innerhalb der MINT-Fächer noch weiter differenzieren lässt. So wiesen Studienteilnehmende mit Ingenieur-Bachelorabschluss die mit Abstand höchsten durchschnittlichen räumlichen Fähigkeiten auf, gefolgt von den „physical sciences“ (Physik, Chemie, Geowissenschaften und Astronomie) sowie Mathematik und Informatik mit ebenfalls überdurchschnittlichen räumlichen Fähigkeiten. Dem gegenüber steht die Biologie, deren Absolventinnen und Absolventen sogar unterdurchschnittliche räumliche Fähigkeiten aufwiesen (ungefähr mittig zwischen Mathematik/Informatik und den Gruppen der nicht-MINT-Fächer), obwohl die Biologie klassischerweise auch zu den Naturwissenschaften gezählt wird. Ähnliche, wenn auch weniger granulare Muster zeigten sich auch bei Schulfachpräferenzen sowie späteren Berufsgruppen. Auch wenn diese Befunde starke Selbstselektionseffekte beinhalten und keine Leistungsunterschiede innerhalb der Fächer (wie Abschlussnoten) betrachten, legen diese großen querschnittlichen Unterschiede zwischen den Fächern auch innerhalb des MINT-Bereichs nahe, dass sich Aussagen zur Stärke des Zusammenhangs von räumlichen Fähigkeiten und Erfolgsmaßen nicht über alle MINT-Fächer hinweg verallgemeinern lassen.

Bisherige Studien (wie auch die zuvor beschriebene von Shea et al., 2001) ermöglichen es allerdings nicht, Vergleiche zwischen den Effekten räumlicher Fähigkeiten auf Erfolg in MINT-Kontexten (akademisch wie beruflich) unterschiedlicher Fächer anzustellen, da diesen Studien einige Probleme gemein sind, die einen direkten Vergleich erschweren. Viele dieser Studien betrachteten nur ein einzelnes Fach (bspw. Ingenieurwesen; Sorby et al., 2000; 2010; 2013; 2018) und verwendeten eine Vielzahl unterschiedlicher und schwer vergleichbarer abhängiger Variablen, die oft spezifisch für das jeweilige Fach (oder sogar den konkreten Studiengang) waren. Die wenigen Studien hingegen, die mehrere Fächer betrachteten, verwendeten dabei entweder sehr allgemeine, oft dichotome Ergebnisse (beispielsweise spätere Berufswahl in einem MINT-Bereich oder nicht in einem MINT-Bereich, ohne detailliertere Fachunterscheidung), oder verglichen rein querschnittlich die durchschnittlichen räumlichen Fähigkeiten zwischen unterschiedlichen Fächern (bspw. Humphreys et al., 1993; Shea et al., 2001; Wai et al., 2009). Beide Varianten beinhalten keine Varianz in den Erfolgsmaßen, die sich durch räumliche Fähigkeiten erklären ließen. Insgesamt ist somit ein Vergleich der tatsächlichen Bedeutung räumlicher Fähigkeiten für den Erfolg in unterschiedlichen Fächern anhand der bisherigen Studien kaum möglich – die ein Fach betrachtenden Studien erlauben nur die Untersuchung des Zusammenhangs in einem Fach, während die fächerübergreifenden Studien keine Erfolgsmaße mit durch räumliche Fähigkeiten erklärbarer Varianz verwendeten und somit keine näheren Zusammenhänge innerhalb der einzelnen Fächer untersuchen konnten.

Diese Lücke sollte mit Studie 2 anhand vier exemplarischer Fächer geschlossen werden, indem anhand der gleichen Richtlinien parallel entwickelte Fachwissenstests als fachspezifische und leistungsbezogene Erfolgsmaße verwendet wurden. Durch die parallele Entwicklung der

Fachwissenstests blieben sie dennoch zwischen den Fächern vergleichbar – im Gegensatz zu beispielsweise (unabhängig voneinander erstellten) Klausuren unterschiedlicher Fächer. Eine weitere Besonderheit der hier dargestellten Studie ist, dass der Effekt räumlicher Fähigkeiten auf Fachwissen als ausgewähltes Studienerfolgsmaß nicht nur bivariat oder unter Kontrolle einzelner Kovariaten untersucht wird, sondern multivariat unter Kontrolle der zuvor genannten kognitiven und wissensbezogenen Einflussfaktoren Intelligenz, Abiturnote und fachliches Vorwissen. Das stärkte einerseits die ziehbaren Schlussfolgerungen und ermöglichte andererseits den Vergleich des Effekts räumlicher Fähigkeiten mit den Effekten der anderen Variablen.

1.3 Training räumlicher Fähigkeiten

1.3.1 Training räumlicher Fähigkeiten und Studienerfolg

Die Metaanalyse von Uttal et al. (2013) zeigte, dass räumliche Fähigkeiten auf unterschiedlichen Arten von Trainings verbesserbar sind. Für den praktischen Nutzen des Wissens um den Zusammenhang zwischen räumlichen Fähigkeiten und MINT-Erfolg ist dies von großer Bedeutung: Räumliche Fähigkeiten können dadurch nicht nur dazu dienen, die von vorneherein bestqualifizierten Personen für MINT-Studiengänge oder -Berufe auszuwählen, sondern ihre Verbesserung durch eine Intervention auch als ein Weg dienen, Personen mit schlechteren Startbedingungen zu unterstützen. Sie können also nicht nur zur Selektion, sondern auch zur Unterstützung durch Intervention genutzt werden. In der Metaanalyse beschrieben Uttal et al. (2013) das Potenzial solcher Trainings in einem Gedankenexperiment: Ein flächendeckend auf Populationsebene angewandtes Training mit einer auf Basis ihrer Daten konservativ geschätzten Effektstärke von Hedges $g = 0.4$ würde die Anzahl der Personen *verdoppeln*, die mindestens genauso gute räumliche Fähigkeiten haben wie durchschnittliche Ingenieure derzeit. Auch wenn eine so breite Intervention als Teil eines Forschungsprojekts kaum möglich ist, zeigt es dennoch das Potenzial solcher Interventionen auf – besonders wenn es sich beispielsweise in Schulcurricula verankern ließe, was einer populationsweiten Anwendung am nächsten käme.

Präferenz und Kompetenz hängen hierbei eng zusammen – Webb et al. (2007, S. 409) sprechen von einer „Fähigkeits-Präferenz-Passung“ (*ability-preference fit*) beim Verfolgen von MINT-Laufbahnen bei räumlich starken Personen. Daraus lässt sich die Hoffnung ableiten, dass eine Verbesserung der räumlichen Fähigkeiten bei beispielsweise Schülerinnen und Schülern auch zu einem gesteigerten Interesse an MINT-Fächern und -Karrieren führt. Dies könnte eine Option darstellen, um mehr Personen für MINT-Fächer zu begeistern und dort auch erfolgreich werden zu lassen, was wiederum den in Kapitel 1.2 genannten hohen Studienabbruchquoten sowie dem Fachkräftemangel entgegenwirken kann.

Kognitive Fähigkeiten im Allgemeinen und räumliche Fähigkeiten im Besonderen sind für die Vorhersage von MINT-Leistungen am bedeutsamsten, wenn der bisherige Wissensstand der untersuchten Personen niedrig ist (Hambrick & Meinz, 2011; Uttal & Cohen, 2012), und verliert mit zunehmendem Wissenserwerb an Bedeutung. Diese Interaktion zeigte sich beispielsweise bei Hambrick et al. (2012): Bei Personen mit geringem geowissenschaftlichem Wissen wurde die Leistung in einer Gesteinskartierungsaufgabe durch räumliche Fähigkeiten vorhergesagt, wohingegen die Leistung bei erfahrenen Geologen unabhängig von deren räumlichen Fähigkeiten war. Uttal und Cohen (2012) führen diese Interaktion von Fachwissen und räumlichen Fähigkeiten darauf zurück, dass Experten passende fachspezifische Repräsentationen verinnerlicht haben, die ihnen die Lösung fachspezifischer Probleme erlauben, ohne auf fachunabhängige räumliche Fähigkeiten zurückgreifen zu müssen. Umgekehrt kann ein Mangel an räumlichen Fähigkeiten eine Einstiegshürde beim Erwerb des benötigten Grundlagenwissens darstellen (Uttal et al., 2013), ohne das wiederum der Aufbau der mentalen Repräsentationen nicht möglich ist, die die Bedeutsamkeit räumlicher Fähigkeit später reduzieren. Daraus folgt, dass Trainings räumlicher Fähigkeiten vor allem bei noch geringem Fachwissensniveau hilfreich sind und sich in besseren fachlichen Leistungen niederschlagen, da sie betroffenen Personen helfen könnten, diese Einstiegshürde zu überwinden. Entsprechend wäre der beste Zeitpunkt für eine solche Intervention „so früh wie möglich“. Hier birgt die Schule großes Potenzial, aber auch die Studieneingangsphase stellt ein vielversprechendes Zeitfenster dar – gerade bei Studiengängen, die kein Schulfach als direktes Pendant haben und darum bei Studienbeginn noch geringes fachliches Vorwissen vorliegt (bspw. im Ingenieurwesen). In der Studieneingangsphase ist es bereits zu spät, um durch Interventionen bei einer größeren Gruppe von Schülerinnen und Schülern Interesse am jeweiligen Fach im Sinne der Fähigkeits-Präferenz-Passung zu wecken. Allerdings bieten in spezifischen Studiengängen verankerte Trainings den Vorteil, dass das Training für das jeweilige Fach auch tatsächlich von Bedeutung ist und gegebenenfalls mit fachbezogenen Inhalten und Repräsentationen angereichert werden kann, wie es beispielsweise Sorby und Baartmans (2000) getan haben.

1.3.2 Kategorien räumlicher Fähigkeitstrainings

Uttal et al. (2013) gruppierten in ihrer Metaanalyse die Interventionen in drei Kategorien:

1. Training mittels wiederholter Bearbeitung räumlicher Aufgaben, oft Varianten klassischer Tests räumlicher Fähigkeiten.
2. Besuch von räumlich anspruchsvollen Kursen, oft als semesterbegleitende Veranstaltungen im Studienkontext.
3. Videospielbasierte Trainings, in denen unterschiedliche (teils kommerzielle, teils eigens entwickelte) Videospiele zum Training räumlicher Fähigkeiten genutzt wurden.

Die erste Kategorie entspricht einem direkten beziehungsweise spezifischen Training, während die Kategorien 2 und 3 indirekte Trainings sind (Baenninger & Newcombe, 1989). Alle drei Trainingskategorien erwiesen sich in der Metaanalyse als effektiv, wobei die Videospieltrainings den größten durchschnittlichen Effekt aufwiesen (Hedges $g = 0.54$), Kurstrainings den niedrigsten (Hedges $g = 0.41$) und der Effekt von Trainings mittels räumlicher Aufgaben mittig dazwischen lag (Hedges $g = 0.48$). Die Trainingseffekte stellten sich in der Metaanalyse auch als stabil heraus – zumindest in den untersuchten Zeiträumen verzögerter Posttests, die bei maximal einem Monat lagen. Auch lagen starke Transfereffekte auf andere, nicht direkt trainierte räumliche Tests vor (Hedges $g \geq 0.51$), sodass grundsätzlich davon ausgegangen werden kann, dass es sich bei den gefundenen Trainingseffekten nicht nur um testspezifische Übungseffekte handelte. Zudem profitierten Personen mit geringen räumlichen Fähigkeiten stärker von den Trainings – in den Studien, in denen gezielt räumlich schwache Personen für die Teilnahme am Training ausgewählt wurden, war der gefundene Trainingseffekt signifikant größer als bei Studien mit reinen Zufallsstichproben (Hedges $g = 0.68$ zu Hedges $g = 0.44$). Dies belegt den Nutzen solcher Interventionen zur Unterstützung von Personen mit geringen räumlichen Fähigkeiten. Beispielsweise könnten so Studierende in der Studieneingangsphase unterstützt werden, in der sie noch mehr auf ihre räumlichen Fähigkeiten angewiesen sind, bevor die als Teil des (darauffolgend erworbenen) Fachwissens angeeigneten fach*spezifischen* mentalen Repräsentationen die fach*unabhängigen* räumlichen Fähigkeiten weniger wichtig werden lassen können.

In Kapitel 1.2 wurde bereits angeschnitten, dass einige Studien bereits räumliche Trainings in MINT-Fächern durchgeführt haben. Die bisherigen Studien nutzten hierfür allerdings vorwiegend die Trainingskategorien 1 und 2 nach Uttal et al. (2013): Entweder wurden räumliche Testbatterien zur Übung verwendet (bspw. Lowrie et al., 2019; Mix et al., 2021), oder räumlich anspruchsvolle Kurse mit Fachbezug stellten die Intervention dar (bspw. Molina-Carmona et al., 2019; Sorby et al., 2018). Videospiele hingegen wurden bisher nicht zu Trainingszwecken im MINT-Kontext genutzt – obwohl sie in der Metaanalyse von Uttal et al. (2013) den stärksten Trainingseffekt erzielten. Die dritte in der vorliegenden Arbeit präsentierte Studie sollte diese Lücke schließen, indem sie ein Videospieltraining bei Bauingenieur-Erstsemesterstudierenden vorstellt. Bauingenieur-Erstsemesterstudierende eignen sich für ein solches Training besonders gut: Die Studierenden starten mangels eines direkten Vorläuferfachs in der Schule hier mit geringem fachlichen Vorwissen und sind somit gerade zu Beginn noch stärker auf ihre räumlichen Fähigkeiten zur Aneignung räumlich anspruchsvoller Studieninhalte angewiesen. Der grundsätzliche Nutzen eines räumlichen Trainings in dieser Zielgruppe wurde zudem von Sorby et al. (bspw. 2000; 2013; 2018) mittels eines Kurstrainings wiederholt belegt.

Videospiele können ein breites Spektrum an kognitiven Fähigkeiten trainieren, wie die Metaanalyse von Bediou et al. (2023) für das Genre der Actionvideospiele zeigt. Dieses Genre stand zu Beginn der Videospielforschung im Fokus, da das Spielen darunter fallender Spiele zu sehr auffälligen Trainingseffekten gerade im Bereich der visuellen Aufmerksamkeit führt (bspw. Green & Bavelier, 2003). Auch räumliche Fähigkeiten wie mentale Rotation können durch Actionvideospiele trainiert werden (Spence & Feng, 2010). Aber auch wenn die Fülle durch Actionvideospiele verbesserbarer kognitiver Fähigkeiten an sich positiv ist, stellt sie für die gezielte Untersuchung von Transfereffekten eines räumlichen Trainings ein Problem dar. Denn wenn mit einer Intervention mehrere Fähigkeiten zugleich trainiert werden, ist eine Zurückführung des Transfereffekts auf genau eine der trainierten Fähigkeiten schwierig. Jüngere Studien untersuchten allerdings auch andere Arten von Spielen – beispielsweise fanden Shute et al. (2015), dass auch das 3D-Puzzlespiel „Portal 2" räumliche Fähigkeiten sowie Problemlösefähigkeiten trainiert (wobei hier auch die Anwendung von Regeln anhand von Ravens Matrizen [Raven, 2000] im Fokus der Messung stand, die zentral als Maß für Reasoning gesehen werden, wie die Autoren der Studie ebenfalls darstellten). Portal 2 ist in Hinblick auf die vom Spiel gestellten Anforderung insofern besonders, dass, im Gegensatz zu Actionvideospielen, Aufmerksamkeitsfunktionen und Reaktionsgeschwindigkeit praktisch keine Rolle spielen. Stattdessen werden in erster Linie räumliche Fähigkeiten benötigt, um im Spiel weiterzukommen. Aufgrund dessen eignet es sich hervorragend als Trainingsspiel zur Untersuchung des Transfereffekts eines solchen Trainings auf MINT-Leistungen. In Studie 3 wird daher Portal 2 (im Vergleich zu einem weniger räumlich anspruchsvollen Vergleichsspiel als aktiver Kontrollbedingung) genutzt, um direkt die räumlichen Fähigkeiten und indirekt als Transfer den Fachwissenserwerb der Bauingenieur-Erstsemesterstudierenden zu verbessern. In der Studie werden das Spiel und seine besonderen räumlichen Anforderungen auch genauer beschrieben.

1.4 Fragestellungen der Arbeit

Vor diesem theoretischen Hintergrund eröffnen sich mehrere aufeinander aufbauende Fragestellungen zur Bedeutung und zum Potenzial räumlicher Fähigkeiten im Studienkontext (mit Fokus auf MINT-Fächer), die, wie an entsprechender Stelle bereits angeschnitten, in den in dieser Arbeit vorgestellten Studien bearbeitet wurden.

1.4.1 Studie 1

Die erste Studie nutzte einen psychometrischen Ansatz, um die kognitiven Prozesse näher zu untersuchen, die beim Lösen räumlicher Aufgaben angewandt werden – insbesondere der Beteiligung räumlicher Prozesse wie mentaler Manipulation (z. B. Rotation) der Stimuli gegenüber Reasoning-basierten Strategien wie dem Vergleich struktureller Elemente (*feature comparison*, Burin et al., 2000). Der Fokus lag hierbei auf dem klassischen mentalen Rotationstest nach Vandenberg und Kuse (1978), da dieser Test auch in den folgenden Studien als Maß räumlicher Fähigkeiten verwendet wurde. Das Ziel der Studie war die Untersuchung, ob die zur Lösung eines Items benötigten Rotationsachsen einen Einfluss auf die Dimensionalität des gemessenen Konstrukts haben und ob möglicherweise vorliegende trennbare Faktoren unterschiedlich von *g* und räumlicher Visualisierungsfähigkeit vorhergesagt werden. Konkret wurde untersucht, ob Items, die die Rotation um eine schiefe Achse erfordern, noch einen über die gewöhnliche Rotation um eine Kardinalachse hinausgehenden Faktor messen und ob diese beiden Faktoren (schiefe und kardinale Rotation) in unterschiedlichem Maße von *g* und einem davon separierten Faktor räumlicher Visualisierungsfähigkeit vorhergesagt werden. Dies ließe Rückschlüsse auf die beteiligten kognitiven Prozesse und angewandten (möglicherweise auch nicht räumlichen) Strategien zu: Wenn beide Rotationsfaktoren im selben Maße von *g* und räumlicher Visualisierungsfähigkeit vorhergesagt würden, dann ließe das darauf schließen, dass auch dieselben kognitiven Prozesse bei der Lösung der Aufgaben aktiv waren, entsprechend auch dieselbe Art von Lösungsstrategien angewandt wurde. Wenn hingegen räumliche Visualisierungsfähigkeit beispielsweise nur den generellen Rotationsfaktor vorhersagen würde, deutete dies darauf hin, dass bei der Lösung der schiefen Rotationsaufgaben (möglicherweise aufgrund ihrer höheren Komplexität) nicht auf räumliche Lösungsstrategien zurückgegriffen wurde. Auf diese Weise sollte einerseits ein besserer Einblick in die in diese beiden Aufgabentypen einfließenden kognitiven Prozesse gewonnen und andererseits die Konstruktvalidität des Tests zur Messung räumlicher Fähigkeiten gegenüber Reasoning bzw. *g* abgesichert werden, die – wie in Kapitel 1.1 erläutert – bekannterweise starke Überlappungen aufweisen. Entsprechend diente diese Studie im Rahmen der Arbeit als eine Absicherung für die Verwendung des Tests in den darauffolgenden Studien.

1.4.2 Studie 2

Die zweite Studie begegnete dem Problem, dass die besondere Bedeutung räumlicher Fähigkeiten zwar für zahlreiche einzelne Fächer sowie für MINT-Studienerfolg im Allgemeinen bekannt ist, aber die bisherigen Studien keine direkten differentiellen Vergleiche zwischen unterschiedlichen Fächern erlaubten. Um diese Lücke zu schließen, untersuchte Studie 2 semesterbegleitend die Entwicklung des Fachwissens Erstsemesterstudierender in Abhängigkeit von ihren räumlichen Fähigkeiten in vier exemplarisch ausgewählten Studiengängen - drei MINT-Fächer und ein nicht-MINT-Fach. Das nicht-MINT-Fach wurde hierbei mit in die Studie aufgenommen, um die besondere Bedeutung räumlicher Fähigkeiten für MINT-Studiengänge gegenüber nicht-MINT-Studiengängen zu überprüfen, die zum Beispiel aus den durchschnittlichen räumlichen Fähigkeiten der Absolventinnen und Absolventen unterschiedlicher Studiengänge in Shea et al. (2001) hervorgehen. Denn auf Basis dieser bisherigen Befunde ließ sich nicht sagen, ob räumliche Fähigkeiten wirklich für MINT-Fächer wichtiger sind als für beispielsweise geistes- oder gesellschaftswissenschaftliche Fächer. Es ließ sich nur sagen, dass Personen mit hohen räumlichen Fähigkeiten häufiger MINT-Studiengänge (und später Berufe) wählen, nicht aber, ob die differentiellen Unterschiede in den räumlichen Fähigkeiten von Studierenden *innerhalb* dieser Studiengänge tatsächlich besser zwischen leistungsstarken und leistungsschwachen Studierenden differenzieren können als es in einem gesellschaftswissenschaftlichen Studiengang (mit niedrigeren *durchschnittlichen* räumlichen Fähigkeiten, aber nach wie vor differentiellen Unterschieden zwischen den Studierenden des Faches) der Fall wäre. Die Aufnahme eines nicht-MINT-Fachs in die vorliegende Studie ermöglichte eine Untersuchung auch dieser Frage.

In dieser Studie wurde das Fachwissen als leistungsbezogenes Studienerfolgsmaß verwendet, da Fachwissenstests sich zwischen verschiedenen Fächern konzeptionell vergleichbar erstellen und innerhalb eines Fachs auch an unterschiedlichen Universitäten identisch durchführen lassen. Klausurergebnisse unterliegen hingegen deutlich stärker weiteren fach- und universitätsspezifischen Einflussfaktoren, die einen angemessenen Vergleich erschweren. Eine weitere Besonderheit der hier vorgestellten Studie ist, dass der Zusammenhang von räumlichen Fähigkeiten und (hier durch das Fachwissen operationalisiertem) Studienerfolg unter Kontrolle weiterer bekannter leistungsbezogener Prädiktoren (Intelligenz, Abiturnote und fachliches Vorwissen) geprüft wurde. Dies stärkt einerseits die Aussagekraft der gefundenen Ergebnisse und ermöglicht andererseits einen Vergleich der Bedeutsamkeit dieser Prädiktoren miteinander. Dies ist insbesondere für den Vergleich mit dem Effekt von *g* interessant, um die Überlappung und Trennbarkeit von räumlichen Fähigkeiten und *g* auch in Bezug auf die Vorhersage externer Variablen näher untersuchen zu können.

1.4.3 Studie 3

Einzelne Interventionsstudien nutzten bereits räumliche Trainings, um durch Transfereffekte auch MINT-Erfolgsmaße zu verbessern und so den kausalen Effekt räumlicher Fähigkeiten auf diese Erfolgsmaße zu belegen. Allerdings wurden bisher mit der Übung räumlicher Aufgaben sowie meist fachbezogenen räumlich anspruchsvollen Kursen nur zwei der drei von Uttal et al. (2013) beschriebenen Interventionsarten für diesen Zweck genutzt, nicht aber Videospielinterventionen als diejenige Interventionsart, die die größten durchschnittlichen Trainingseffekte in Uttal et al.'s (2013) Metaanalyse zeigte. Diese Lücke sollte die dritte Studie der vorliegenden Arbeit schließen, indem semesterbegleitend ein Videospieltraining bei Bauingenieur-Erstsemesterstudierenden angewandt wurde, um ihre räumlichen Fähigkeiten zu verbessern und dadurch auch ihren Fachwissenserwerb zu unterstützen. Hierfür wurde ein räumlich anspruchsvolles 3D-Puzzlespiel in der Experimentalgruppe und ein weniger räumlich anspruchsvolles Krimispiel in der aktiven Kontrollgruppe verwendet. Es wurde untersucht, ob das räumlich anspruchsvolle Spiel einen stärkeren Trainingseffekt auf die räumlichen Fähigkeiten der Studierenden hatte, und ob die so trainierten räumlichen Fähigkeiten studienbezogene Leistungsmaße vorhersagen. Neben dem erworbenen Fachwissen wurden hier auch Klausurnoten als Leistungsmaß untersucht, da diese bei nur einer untersuchten Kohorte (im Vergleich zur zweiten Studie) sinnvoll nutzbar sind und letztlich das formal ausschlaggebende (weil das sich auf dem Zeugnis widerspiegelnde) Maß für den Studienerfolg darstellen. Hierbei wurde ebenfalls geprüft, ob räumliche Fähigkeiten auch direkt auf den Klausurerfolg wirken oder allein durch die Unterstützung des Fachwissenserwerbs (also ob der Effekt räumlicher Fähigkeiten auf Klausurerfolg vollständig oder partiell über Fachwissen mediiert wird).

Insgesamt sollte die vorliegende Arbeit also zentral drei Fragen untersuchen:

1. Bieten räumliche Fähigkeiten über *g* hinaus einen Mehrwert für einerseits die Vorhersage der Leistung in einem typischen räumlichen Fähigkeitstest (Studie 1) und andererseits bei der Vorhersage externer studienbezogener Leistungsvariablen (Studie 2 & 3)?
2. Wie unterscheiden sich unterschiedliche Studiengänge (besonders im Vergleich MINT- vs. nicht-MINT-Fächer) hinsichtlich der Bedeutung räumlicher Fähigkeiten (Studie 2)?
3. Kann ein Videospieltraining zum Training räumlicher Fähigkeiten genutzt werden und als Transfereffekt auch studienbezogene Leistungen verbessern (Studie 3)?

1.5 Literaturverzeichnis

Averbeck, D. (2021). *Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums: Der Einfluss kognitiver und affektiv-motivationaler Variablen*. Dissertation. *Studien zum Physik- und Chemielernen: Bd. 308*. Logos.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews. Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baenninger, M. & Newcombe, N. S. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*(5-6), 327–344. https://doi.org/10.1007/BF00287729

Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S. & Bavelier, D. (2023). Effects of action video game play on cognitive skills: A meta-analysis. *Technology, Mind, and Behavior*, *4*(1). https://doi.org/10.1037/tmb0000102

Benedičič, U., Duffy, G. & Sorby, S. A. (2022). An ongoing spatial intervention project in irish secondary schools for improved engineering education. In *Towards a new future in engineering education, new scenarios that european alliances of tech universities open up* (S. 943–951). Universitat Politècnica de Catalunya. https://doi.org/10.5821/conference-9788412322262.1321

Boone, A. P. & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(7), 1005–1019. https://doi.org/10.1037/xlm0000370

Burin, D. I., Delgado, A. R. & Prieto, G. (2000). Solution strategies and gender differences in spatial visualization tasks. *Psicológica*, *21*(2), 275–286. https://www.redalyc.org/articulo.oa?id=16921204

Burton, N. W. & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980*. Research Report No. 2001-2. College Entrance Examination Board.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

CBI. (2011). *Building for growth: Business priorities for education and skills*. https://help.open.ac.uk/students/_data/documents/careers/building-for-growth-cbi-education-and-skills-survey-2011.pdf

CBI. (2019). *Getting young people 'work ready'*. https://www.cbi.org.uk/media/2960/cbi_work-readiness.pdf

Chen, X. (2009). Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education. Stats in Brief. NCES 2009-161. *National Center for Education Statistics*.

Cheng, Y.-L. & Mix, K. S. (2014). Spatial training improves children‘s mathematics ability. *Journal of Cognition and Development*, *15*, 2–11. https://doi.org/10.1080/15248372.2012.725186

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20–43. https://doi.org/10.1016/0010-0285(75)90003-1

Doyle, R. A. & Voyer, D. (2018). Photographs of real human figures: Item types and persistent sex differences in mental rotation. *Quarterly Journal of Experimental Psychology (2006)*, *71*(11), 2411–2420. https://doi.org/10.1177/1747021817742079

Duesbury, R. T. & O‘Neil, H. F. (1996). Effect of type of practice in a computer-aided design environment in visualizing three-dimensional objects from two-dimensional orthographic projections. *Journal of Applied Psychology*, *81*, 249–260. https://doi.org/10.1037/0021-9010.81.3.249

Ekstrom, R. B., French, J. W., Harman, H. H. & Dermen, D. (1976). *Manual for the kit of factor-referenced cognitive tests* (Bd. 102). Educational Testing Service. http://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf

Flanagan, D. P. & Dixon, S. G. (2013). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest & E. Fletcher-Janzen (Hrsg.), *Wiley online library. Encyclopedia of special education: A reference for the education of children adolescents and adults with disabilities and other exceptional individuals* (4. Aufl.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118660584.ese0431

Fleischer, J., Leutner, D., Brand, M., Fischer, H. E., Lang, M., Schmiemann, P. & Sumfleth, E. (2019). Vorhersage des Studienabbruchs in naturwissenschaftlich-technischen Studiengängen. *Zeitschrift für Erziehungswissenschaft*, *22*(5), 1077–1097. https://doi.org/10.1007/s11618-019-00909-w

Formazin, M., Schroeders, U., Köller, O., Wilhelm, O. & Westmeyer, H. (2011). *Studierendenauswahl im Fach Psychologie* (Bd. 62). https://doi.org/10.1026/0033-3042/a000093

Freyer, K., Epple, M., Brand, M., Schiebener, J. & Sumfleth, E. (2014). Studienerfolgsprognose bei Erstsemesterstudierenden in Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, *20*(1), 129–142. https://doi.org/10.1007/s40573-014-0015-3

Ganley, C. M., Vasilyeva, M. & Dulaney, A. (2014). Spatial ability mediates the gender difference in middle school students‘ science performance. *Child Development*, *85*(4), 1419–1432. https://doi.org/10.1111/cdev.12230

Green, C. S. & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. https://doi.org/10.1038/nature01647

Grzybowski, D. M., Stavridis, O., Barclay, L. A., Abrams, L., Sorby, S. A., Thomas, J. & Merril, J. A. (2014). Impact of optional supplemental course to enhance spatial visualization skills in first-year engineering students. In American Society for Engineering Education (Vorsitz), *2014 ASEE Annual Conference*.

Ha, O. & Fang, N. (2016). Spatial ability in learning engineering mechanics: critical review. *Journal of Professional Issues in Engineering Education and Practice*, *142*(2), 4015014. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000266

Hailikari, T. (2009). *Assessing university students' prior knowledge: Implications for theory and practice*. Helsinki University Print. https://helda.helsinki.fi/handle/10138/19841

Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., Turner, S. P., Rench, T. A. & LaDue, N. D. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General*, *141*, 397–403.

Hambrick, D. Z. & Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science*, *20*(5), 275–279. https://doi.org/10.1177/0963721411422061

Hegarty, M., Keehner, M., Khooshabeh, P. & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, *19*, 61–70. https://doi.org/10.1016/j.lindif.2008.04.006

Hegarty, M. & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689. https://doi.org/10.1037/0022-0663.91.4.684

Heil, M. & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology (2006)*, *61*(5), 683–689. https://doi.org/10.1080/17470210701822967

Heublein, U., Richter, J. & Schmelzer, R. (2020). *Die Entwicklung der Studienabbruchquoten in Deutschland.* https://doi.org/10.34878/2020.03.DZHW_BRIEF

Heublein, U. & Schmelzer, R. (2018). Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen: Berechnungen auf Basis des Absolventenjahrgangs 2016. *DZHW-Projektbericht*.

Hsi, S., Linn, M. C. & Bell, J. E. (1997). The role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education*, *86*, 151–158.

Humphreys, L. G., Lubinski, D. & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology*, *78*, 250–261. https://doi.org/10.1037//0021-9010.78.2.250

Judd, N. & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour.* Vorab-Onlinepublikation. https://doi.org/10.1038/s41562-021-01118-4

Khine, M. S. (Hrsg.). (2017). *Visual-spatial Ability in STEM Education*. Springer. https://doi.org/10.1007/978-3-319-44385-0

Kozhevnikov, M., Motes, M. A. & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31*, 549–579. https://doi.org/10.1080/15326900701399897

Krüger, J. K. & Suchan, B. (2016). You should be the specialist! Weak mental rotation performance in aviation security screeners – reduced performance level in aviation security with no gender effect. *Frontiers in Psychology*, *7*, 333. https://doi.org/10.3389/fpsyg.2016.00333

Kryshko, O., Fleischer, J., Waldeyer, J., Wirth, J. & Leutner, D. (2020). Do motivational regulation strategies contribute to university students‘ academic success? *Learning and Individual Differences*, *82*, 101912. https://doi.org/10.1016/j.lindif.2020.101912

Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological Bulletin*, *127*(1), 162–181. https://doi.org/10.1037/0033-2909.127.1.162

Lohman, D. F. (1996). Spatial ability and G. In I. Dennis & P. Tapsfield (Hrsg.), *Human abilities: Their nature and measurement* (S. 97–116). Lawrence Erlbaum Associates, Inc.

Lowrie, T., Logan, T. & Hegarty, M. (2019). The influence of spatial visualization training on students’ spatial reasoning and mathematics performance. *Journal of Cognition and Development*, *20*(5), 729–751. https://doi.org/10.1080/15248372.2019.1653298

Lowrie, T., Logan, T. & Ramful, A. (2017). Visuospatial training improves elementary students‘ mathematics performance. *The British Journal of Educational Psychology*, *87*, 170–186. https://doi.org/10.1111/bjep.12142

Miller, D. I. & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences*, *26*, 141–152. https://doi.org/10.1016/j.lindif.2012.03.012

Mix, K. S., Levine, S. C., Cheng, Y.-L., Stockton, J. D. & Bower, C. (2021). Effects of spatial training on mathematics in first and sixth grade children. *Journal of Educational Psychology*, *113*(2), 304–314. https://doi.org/10.1037/edu0000494

Molina-Carmona, R., Pertegal-Felices, M. L., Jimeno-Morenilla, A. & Mora-Mora, H. (2019). Assessing the Impact of Virtual Reality on Engineering Students’ Spatial Ability. In A. Visvizi, M. D. Lytras & L. Daniela (Hrsg.), *Emerald studies in higher education innovation and technology. The future of innovation and technology in education: Policies and practices for teaching and* *learning excellence* (S. 171–185). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78756-555-520181013

Muto, H. & Nagai, M. (2020). Mental rotation of cubes with a snake face: The role of the human-body analogy revisited. *Visual Cognition*, *28*(2), 106–111. https://doi.org/10.1080/13506285.2020.1727598

National Academies Committee on Science, Engineering, and Public Policy. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. National Academies Press.

Neisser, U., Bodoo, G., Bouchard, T. J., JR., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., Urbina, S. & Boodoo, G. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Pallrand, G. J. & Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching*, *21*, 507–516. https://doi.org/10.1002/tea.3660210508

Parsons, L. M. (1987). Visual discrimination of abstract mirror-reflected three-dimensional objects at many orientations. *Perception & Psychophysics*, *42*(1), 49–59. https://doi.org/10.3758/BF03211513

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, *57*(2), 176–184. https://doi.org/10.1016/j.bandc.2004.08.052

Peters, M. & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260–264. https://doi.org/10.1016/j.bandc.2007.09.003

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*(6), 901–910. https://doi.org/10.1016/j.econedurev.2010.07.009

Rahe, M., Ruthsatz, V. & Quaiser-Pohl, C. (2021). Influence of the stimulus material on gender differences in a mental-rotation test. *Psychological Research*, *85*(8), 2892–2899. https://doi.org/10.1007/s00426-020-01450-w

Raven, J. (2000). The Raven‘s progressive matrices: change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Rochford, K. (1985). Spatial learning disabilities and underachievement among university anatomy students. *Medical Education*, *19*(1), 13–26. https://doi.org/10.1111/j.1365-2923.1985.tb01134.x

Rogister, F., Pottier, L., El Haddadi, I., Monseur, J., Donneau, A.-F., Diep, A. N., Camby, S., Defaweux, V., Bonnet, P., Tombu, S., Lefebvre, P. & Poirrier, A.-L. (2021). Use of Vandenberg and Kuse mental rotation test to predict practical performance of sinus endoscopy. *Ear, Nose & Throat Journal*, 1455613211000599. https://doi.org/10.1177/01455613211000599

Schnettler, T., Bobe, J., Scheunemann, A., Fries, S. & Grunschel, C. (2020). Is it still worth it? Applying expectancy-value theory to investigate the intraindividual motivational process of forming intentions to drop out from university. *Motivation and Emotion*, *44*(4), 491–507. https://doi.org/10.1007/s11031-020-09822-w

Shea, D. L., Lubinski, D. & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604–614. https://doi.org/10.1037//0022-0663.93.3.604

Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Shepard, S. & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 3–11. https://doi.org/10.1037/0096-1523.14.1.3

Shute, V. J., Ventura, M. & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, *80*, 58–67. https://doi.org/10.1016/j.compedu.2014.08.013

Sorby, S. A. (2009). Developing spatial cognitive skills among middle school students. *Cognitive Processing*, *10*, 312-315. https://doi.org/10.1007/s10339-009-0310-y

Sorby, S. A. & Baartmans, B. J. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year eengineering students. *Journal of Engineering Education*, *3*(89), 301–307.

Sorby, S. A., Casey, M. B., Veurink, N. & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, *26*, 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sorby, S. A. & Veurink, N. (2010). Long-term results from spatial skills intervention among first-year engineering students. In American Society for Engineering Education (Vorsitz), *Proceedings of the 65th midyear meeting of the engineering design graphics division of ASEE*.

Sorby, S. A., Veurink, N. & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is ‘yes’. *Learning and Individual Differences*, *67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Spearman, C. E. (1904). „General intelligence“ - objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spearman, C. E. & Wynn Jones, L. (1950). *Human ability: A continuation of „The abilities of man“*. Macmillan.

Spence, I. & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92–104. https://doi.org/10.1037/a0019491

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*(2), 219–234. https://doi.org/10.1016/j.learninstruc.2007.01.012

Stieff, M., Dixon, B. L., Ryu, M., Kumi, B. C. & Hegarty, M. (2014). Strategy training eliminates sex differences in spatial problem solving in a stem domain. *Journal of Educational Psychology*, *106*(2), 390–402. https://doi.org/10.1037/a0034823

Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D. & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology*, *110*(8), 1160–1174. https://doi.org/10.1037/edu0000258

Talley, L. H. (1973). The use of three-dimensional visualization as a moderator in the higher cognitive learning of concepts in college level chemistry. *Journal of Research in Science Teaching*, *10*(3), 263–269. https://doi.org/10.1002/tea.3660100311

Thurstone, L. L. (1938). *Primary mental abilities*. The University of Chicago Press.

Uttal, D. H. & Cohen, C. A. (2012). Spatial thinking and STEM education. In B. H. Ross (Hrsg.), *The psychology of learning and motivation* (Bd. 57, S. 147–181). Elsevier. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C. & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin & Review*, *18*(2), 267–277. https://doi.org/10.3758/s13423-010-0042-0

Voyer, D., Voyer, S. & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Wai, J., Lubinski, D. & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835. https://doi.org/10.1037/a0016127

Webb, R. M., Lubinski, D. & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420. https://doi.org/10.1037/0022-0663.99.2.397

Wu, H.-K. & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492. https://doi.org/10.1002/sce.10126

York, T. T., Gibson, C. & Rankin, S. (2015). *Defining and measuring academic success.* https://doi.org/10.7275/HZ5X-TX03

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19. https://doi.org/10.1162/jocn.2008.20013

2 Studie 1: Rotational Complexity in Mental Rotation Tests: Cognitive Processes in Tasks Requiring Mental Rotation Around Cardinal and Skewed Rotation Axes

2.1 Abstract

Mental rotation tests have been extensively studied regarding item characteristics that affect difficulty, e.g., angular disparity, item dimensionality, and object complexity. In the present study, we applied a psychometric approach to examine whether complex skewed-axis rotation requires incremental processes that can be distinguished from simple cardinal-axis rotation. Participants (*N* = 372) completed a battery of cognitive tests, including a mental-rotation test requiring mental rotation of Shepard and Metzler type figures around a single cardinal axis or around two cardinal axes (resulting in a skewed-axis rotation). When comparing a nested-factor measurement model to a one-factor model, results showed that complex skewed-axis rotation is not identifiable as a nested specific factor. This suggests that the processes resulting in individual differences in mental rotation are either the same in both item types, or at least substantially correlated. Including spatial visualization tests and reasoning tests in a prediction model suggested that participants used spatial strategies over and above reasoning to solve the mental rotation items. These results generalize the findings of Just and Carpenter (1985) on simple cardinal-axis and complex skewed-axis rotation of cubes to more complex objects that allow more flexible mental rotations. It can be concluded that mental rotation represents a unitary ability. From an individual-differences perspective, this ability can be assessed equally with simple cardinal-axis and complex skewed-axis rotation items.

This manuscript was published in:

Nolte, N., Schmitz, F., Fleischer, J., Bungart, M., & Leutner, D. (2022). Rotational complexity in mental rotation tests: Cognitive processes in tasks requiring mental rotation around cardinal and skewed rotation axes. *Intelligence*, 91, 101626. https://doi.org/10.1016/j.intell.2022.101626

2.2 Introduction

Mental rotation tasks have received considerable attention in ability research. As early as 1938, Thurstone conceived of spatial ability as a primary mental ability, and a corresponding factor is also included in Carrol's model (Carroll, 1993) as well as in the most recent taxonomy of human abilities, the CHC model (Flanagan & Dixon, 2013). Further, mental rotation tests have been used as ability indicators because they are closely related with reasoning, fluid intelligence (Carroll, 1993), and the general factor g of intelligence (Lohman, 1996), although they appear to be dissociable from the latter (Carroll, 1993; Flanagan & Dixon, 2013; Thurstone, 1938). According to Carroll (1993), mental rotation is part of the spatial visualization facet of spatial ability, which he defined as "processes of apprehending, encoding, and mentally manipulating spatial forms".

Experimental cognitive psychology has shed light on low-level processes that are assumed to affect performance in mental rotation tests. These processes relate, among others, to task characteristics such as angular disparity (e.g., R. N. Shepard & Metzler, 1971, for 3D tasks, and Cooper, 1975, for 2D tasks), stimulus dimensionality (S. Shepard & Metzler, 1988), or axis of rotation (Parsons, 1987). It is less known, however, to what extent these task characteristics are relevant to the assessment of individual differences in mental rotation ability, since they may call for different cognitive functions and abilities. If this were the case, it could affect how mental rotation should be assessed for diagnostic purposes.

The aim of the present contribution is to investigate mental rotation ability as assessed with the Vandenberg and Kuse paradigm (S. G. Vandenberg & Kuse, 1978) which is one of the most frequently used rotation tests in applied assessment. In this introduction we start by reviewing the relevance of the paradigm in research and its validity for ability assessment. Next, adopting a cognitive processing perspective we discuss task requirements and processing strategies; a particular focus is on the complexity of mental rotation as determined by the axes of rotation. We then summarize the evidence available today concerning individual differences. Finally, we elaborate on the aim of this contribution, namely: to explore (1) the dimensionality of mental rotation ability from an assessment perspective and (2) its incremental validity over and above reasoning ability.

2.1.1 The Vandenberg and Kuse Paradigm

The Vandenberg and Kuse paradigm (S. G. Vandenberg & Kuse, 1978) is a highly popular measure of mental rotation ability. In this task, the classic Shepard and Metzler figures (R. N. Shepard & Metzler, 1971) are used. These are 3-dimensional figures composed of conjoined regular cubes (see Figure 2.1). Participants are instructed to inspect an anchor stimulus and to select two identical but rotated target stimuli from a row of stimuli including two distractors (which are either mirrored or differently composed). Performance is measured as the percentage of correct answers (“found both target stimuli in the item”) across a series of items, and usually scores around 40-50% accuracy (e.g., Krüger & Suchan, 2016; Peters, 2005; S. G. Vandenberg & Kuse, 1978). This is considerably less than the 95% accuracy level frequently obtained in the Shepard and Metzler paradigm (R. N. Shepard & Metzler, 1971). Consequently, the Vandenberg and Kuse paradigm does not suffer from ceiling effects; this is beneficial for the assessment of individual differences.

In fact, the Vandenberg and Kuse paradigm has been frequently used in studies on spatial ability and its relevance in applied contexts: For instance, it has been shown to be a predictor of academic and professional success in science, technology, engineering, mathematics (i.e., the STEM fields) and medicine, ranging from study achievement (e.g., Sorby et al., 2018) to dentistry (Hegarty et al., 2009), endoscopy (Rogister et al., 2021), or aviation security (Krüger & Suchan, 2016). Furthermore, the paradigm has been used in studies on sex differences in rotation ability, which revealed that male participants typically outperform female participants (for a review see Voyer et al., 1995). Hormones have been discussed as a possible cause of the observed group differences (Hausmann et al., 2000).

The composition of the test sets varies between studies, especially since Peters and Battista's (2008) stimulus library has enabled researchers to easily create their own test sets.

2.1.2 Task Characteristics

Performance in ability tests generally depends on task characteristics. For mental rotation tests, some of the most important of these are, among others, angular disparity, stimulus dimensionality and axis of rotation.

Probably the best-documented factor is angular disparity. R. N. Shepard and Metzler (1971) first described a consistent linear relationship between angular disparity and reaction time, which other researchers confirmed under different circumstances (e.g., Cooper, 1975, for 2D tasks). The common explanation for this is the overlap between mental rotation and manual rotation found in fMRI studies (for a review, see Zacks, 2008). Boone and Hegarty (2017) described a trend similar to that for reaction time for accuracy up to 90° of angular disparity.

In contrast, the rotated stimulus’ dimensionality appears to mainly affect the intercept of the reaction-time-by-angular-disparity function (S. Shepard & Metzler, 1988), indicating that dimensionality predominantly affects the encoding of the stimulus and, to a lesser extent, the actual rotation process.

The required axis of rotation affects the slope of the reaction-time-by-angular-disparity function, with the horizontal axis displaying the shallowest slope and the line-of-sight axis, perpendicular to the picture plane, displaying the steepest slope (Parsons, 1987). However, Stieff et al. (2018) showed that this order changes when accuracy is considered, instead of reaction time: Rotation in the picture plane – that is, a rotation around an axis perpendicular to the picture plane – was more accurate than any rotation in depth, that is, a rotation around an axis not perpendicular to the picture plane.

Other task characteristics of interest include stimulus complexity (e.g., Cooper, 1975; Heil & Jansen-Osmann, 2008; Stieff, 2007; Stieff et al., 2018) and the participant's familiarity with the type of stimulus (e.g., Doyle & Voyer, 2018; Muto & Nagai, 2020; Stieff, 2007).

Another characteristic affecting performance that is not intrinsic to the stimuli, is the application of a time limit, as is commonly applied in the Vandenberg and Kuse paradigm: Shorter time limits lead to weaker performance (Peters, 2005). Time limits also exacerbate gender differences: In a meta-analysis, Voyer (2011) found a linear relationship between the length of the time limit and the magnitude of sex differences, and the smallest sex difference in studies with no time limit. The effects of time limits may be mediated by strategy choice, with participants adopting more time-efficient strategies being less impacted by a time limit.

2.1.3 Processing Strategies

From a cognitive processing perspective, a core question pertains to what strategies are applied when participants complete a test. For mental rotation, two strategies have been proposed as most prevalent, namely holistic (spatial) and analytical (reasoning-based; Burin et al., 2000).

The holistic spatial strategy is expected for mental rotation tasks as it necessitates the imagination of rotating the shown stimulus (i.e., mentally rotating it; Burin et al., 2000). When this strategy is used, the test is expected to measure spatial visualization defined as "processes of apprehending, encoding, and mentally manipulating spatial forms" (Carroll, 1993).

Opposed to that, analytical strategies in spatial tasks are mainly centered around feature matching or feature comparison (e.g., Glück et al., 2002). This is an analytical approach that requires reasoning instead of spatial ability (in contrast to the holistic spatial approach that requires mental rotation). Participants using an analytical strategy focus on salient features to aid their same-different decision – most frequently the shape of either the whole stimulus or separate parts of the stimulus is used (Burin et al., 2000), but other features such as color patterns are also possible (Khooshabeh & Hegarty, 2010).

It has been argued that in the Vandenberg and Kuse paradigm participants often use both types of strategies combined (Hegarty, 2018). This may be prompted by items that have two types of distractors: on the one hand, mirrored versions of the target stimulus (mirror foils), and on the other hand stimuli differing slightly in shape (structure foils). Structure foils can be eliminated using feature matching, while mirror foils cannot. Therefore, participants predominantly using analytic strategies such as feature matching perform much worse on items having only mirror foils than on items having only structure foils (Geiser et al., 2006).

Time limits interact with strategy choice, as holistic strategies are usually more time efficient and therefore lead to increased performance under restricted time, compared to analytical strategies. This is another possible reason for the male advantage in the Vandenberg and Kuse paradigm, which is usually applied with a time limit: Compared to women, men are more likely to use efficient strategies (Hirnstein et al., 2009) and are less likely to use more analytical strategies (such as feature matching; Geiser et al., 2006).

Stimulus complexity and axes of rotation also affect strategy choice. While a 90° rotation around one edge of a simple cube may be represented as turning the cube over one step, rotations of more complex stimuli and around less pre-defined rotational axes can lead to different types of rotational strategy (Just & Carpenter, 1985; Parsons, 1987) or, due to increased difficulty, to switching from rotation to feature matching, as Boone and Hegarty's (2017) results on changes in response strategy in items with high angular disparity indicate.

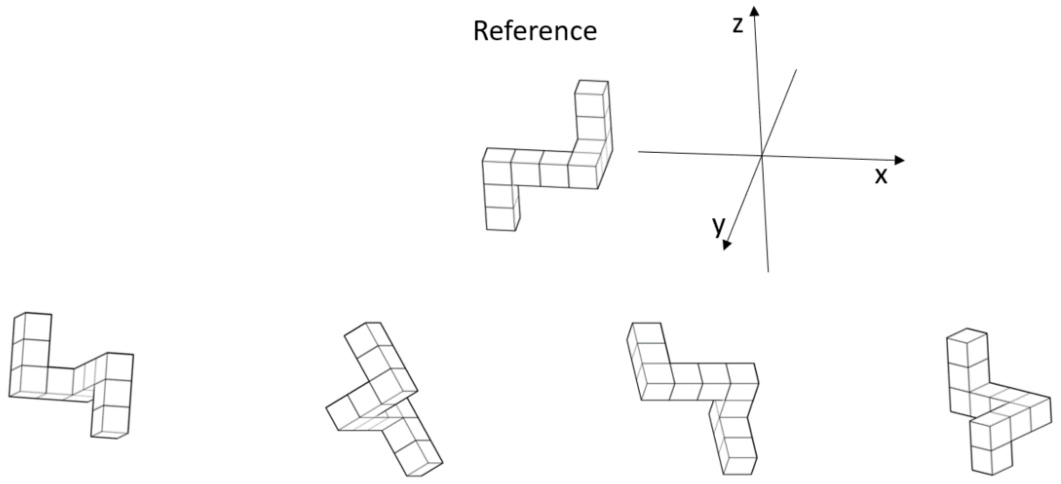
2.1.4 Cardinal Axes

With the use of stimuli composed of regular cubes, cardinal axes corresponding to the edges of the cubes pop out relatively easily. Similarly, in the Shepard and Metzler figures the cardinal axes are given by the spatial orientation of segments of the figures, as the segments are orthogonal to each other and thereby span a 3D reference frame for rotation (see Figure 2.1). In the original R. N. Shepard and Metzler (1971) studies and in most similar studies, figures were rotated around one of these cardinal axes.

However, mental rotation in the real world is often considerably more complex, and requires rotation around a skewed axis that does not correspond to any of the cardinal axes defined by the edges of the object. Just and Carpenter (1985) and Parsons (1987) both describe two types of rotational strategy that may be pursued to this end: On the one hand, participants could use a multiple-step rotation (“rotation by dimension”) around the cardinal axes. On the other hand, participants may rotate a stimulus in a single-step (or complex skewed) rotation, which is referred to by Parsons as “shortest path rotation”. In multiple-step rotation the rotation axes are given by the cardinal axes of the reference frame, whereas in single-step rotation, the skewed rotation axis is not given by and does not coincide with the cardinal axes of the reference frame. In this latter case, the skewed rotation axis has to be identified by the participant, allowing for a direct, single-step rotation.

Figure 2.1

Example Item of the Vandenberg & Kuse Mental Rotation Test



Target 1
125° z-axis

Target 2
170° z-axis +
60° x-axis
= 172° skewed

Distractor 1
170° z-axis +
140° x-axis
= 176° skewed
mirrored

Distractor 2
50° z-axis
structurally
different

Note. Target 1 represents a simple cardinal-axis rotation in relation to the reference figure while Target 2 represents a complex skewed-axis rotation in relation to the reference figure. Distractor 1 is a mirror foil while Distractor 2 is a structure foil.

As could be expected, items requiring rotation around a skewed axis showed longer reaction times compared to items including the same degree of rotation but around a single cardinal axis (Just & Carpenter, 1985; Parsons, 1987). This can be explained in relation to both rotation strategies: Multiple-step “rotation by dimensions” takes longer as it entails a longer rotation path as well as a switch between rotation axes. The direct “shortest path” rotation also takes longer than rotation around a single cardinal axis as it requires identification of the item-specific skewed rotation axis before performing the single-step rotation. The need to first identify a rotation axis is an additional requirement in the mental rotation process that makes such tasks more complex. In line with this, Ziemek et al. (2012) showed that participants performed with higher accuracy on items with mechanical stimuli whose features were aligned to the axes of rotation than items with anatomical stimuli that lacked this property.

Just and Carpenter (1985) used eye-tracking measures to show that, in cubes, only high-spatial participants were able to efficiently utilize the most efficient shortest-path-rotation strategy (Parsons, 1987), while low-spatial participants used the less efficient rotation-by-dimension strategy. This indicates that individual differences in spatial visualization ability influence what type of rotation strategy participants use. Likewise, Arendasy and Sommer (2013) showed that if a task becomes too difficult relative to a participant’s ability, participants are more likely to use feature matching strategies instead. Just and Carpenter (1985) described a similar notion.

From an assessment perspective, differences in strategies for solving different types of mental rotation items turn out to be a challenge. It seems that items requiring more complex (skewed or multiple-step) rotations require different cognitive processes – i.e., either direct rotation, including identification of the task-specific rotation axis, or stepwise rotation, including a switch of rotation axes – that may be distinguishable from the cognitive processes that are needed to solve classic cardinal rotation items. This raises the question whether solving complex mental rotation items (requiring rotation around axes that are not given by the frame of reference) necessitates spatial abilities that go beyond the spatial abilities that are needed to solve more simple mental rotation items (requiring rotation around cardinal axes that are given by the frame of reference), and to what extent task performance reflects spatial visualization ability at all (rather than a reasoning-related shortcut process such as feature matching; Hegarty, 2018).

2.2 Aims of the Present Study

This study aims to contribute to a better understanding of mental rotation ability, its assessment, and its factorial structure. While classic tests of mental rotation have used simple rotation around cardinal axes (see e.g., S. Shepard & Metzler, 1988), rotation requirements in the real world usually are more complex and may comprise complex skewed-axis rotation (e.g., for body parts, as described in Wohlschläger & Wohlschläger, 1998). In turn, the latter can be achieved by finding the most appropriate direct (skewed) rotation axes or by completing a stepwise sequence of cardinal rotations including the requirement to switch between these. In any case, this provokes the question whether complex skewed-axis rotation necessitates cognitive functions and processes that are psychometrically dissociable from those used for simple rotation around cardinal axes. Put differently, we seek to answer the question whether complex skewed-axis rotation requires incremental abilities that can be distinguished from simple cardinal-axis rotation. Besides its theoretical relevance, this question is also of practical relevance for the specific selection of items in mental rotation ability tests, as Peters and Battista's (2008) stimulus library enables researchers to tailor a mental rotation test to specific needs.

Technically, we seek to address this question by means of fitting a nested-factor CFA to indicators requiring either simple cardinal-axis rotation or complex skewed-axis rotation, with a shared general rotation factor capturing variance in all indicators and a nested specific factor accounting only for the incremental requirements in indicators requiring complex skewed-axis rotation. If the nested factor is identified (substantial and statistically significant), this would be interpreted as evidence that complex skewed-axis rotation captures an ability incremental to that involved in simple cardinal-axis rotation items. In turn, a distinguishable skewed-axis rotation ability could potentially possess differential relations and incremental validity for relevant outcome variables.

Further, we consider previous evidence that mental rotation tasks are generally closely related with reasoning ability, as most ability indicators are heavily intercorrelated (Carroll, 1993), and note that cognitive research has shown that participants may resort to reasoning-related processing strategies (e.g., feature matching) instead of using spatial visualization-related processes, especially with difficult items (Arendasy & Sommer, 2013; Just & Carpenter, 1985). These previous findings raise the question whether the currently available tests used to assess mental rotation, such as tests following the Vandenberg and Kuse paradigm, capture it as an ability that can be distinguished from reasoning ability.

We sought to test this analogously by administering reasoning and visualization tests as relevant predictors. Nested-factor CFA was used to specify a general factor of reasoning ability (where reasoning ability may serve as the closest proxy to general intelligence *g*; Carroll, 1993) accounting for variance in all indicators. This general reasoning factor would likely account for a substantial portion of variance in visualization tests because of the positive manifold inherent in all ability tests (Carroll, 1993; Spearman, 1904). Importantly, a nested visualization-specific factor was tested that accounts for specific variance in visualization tests only. In turn, this nested specific factor captures portions of variance that are unique to visualization, and thus, independent of the general reasoning factor. In turn, the predictions of the nested factor would indicate a contribution from a specific visualization ability.

These factors of reasoning and visualization, respectively, were tested as predictors of the ability factors modeled for the mental rotation tests. A contribution from the general reasoning factor would be predicted (i.e., reflecting positive manifold; Carroll, 1993; Spearman, 1904). More importantly, we sought to test whether there is visualization-specific ability in those factors accounting for the performance in mental rotation tests. This would indicate that participants indeed use holistic spatial strategies and not only reasoning-based analytical strategies in solving mental rotation tasks – both for stimuli requiring simple cardinal-axis rotations and for stimuli requiring complex skewed-axis rotations.

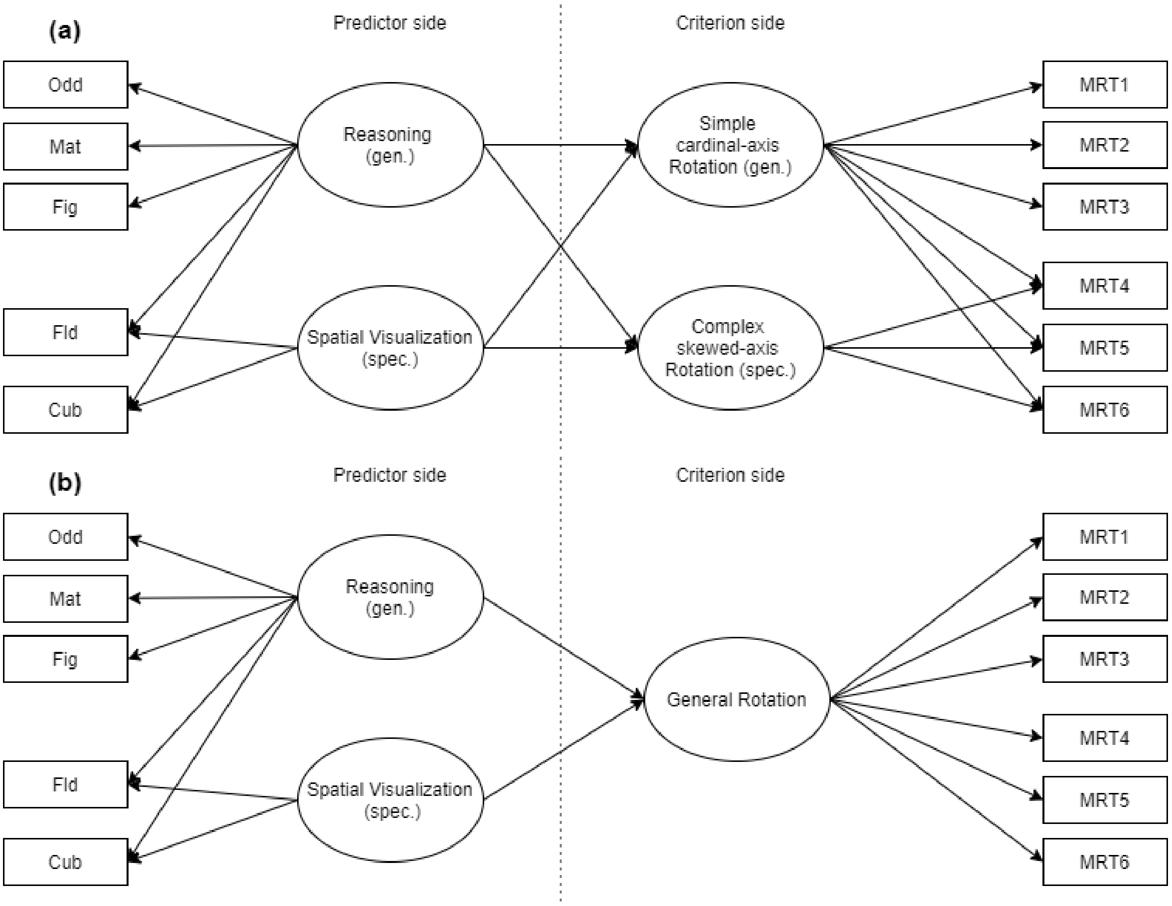
Hypotheses:

1. The nested-factors measurement model on the criterion side (depicted on the right-hand side in Figure 2.2a) fits the data better than the one-factor measurement model (depicted on the right-hand side in Figure 2.2b), indicating that including complex skewed-axis rotation as a separate factor provides incremental value for measuring mental rotation ability.

2. Spatial visualization ability predicts mental rotation ability over and above reasoning ability, indicated by significant path coefficients between the nested specific spatial-visualization factor and the factor(s) accounting for variance in the spatial rotation items, let this either be both the simple cardinal-axis rotation factor and the complex skewed-axis rotation factor (in the case where the model depicted in Figure 2.2a is adopted) or a general rotation factor (in the case where the model depicted in Figure 2.2b is adopted).

Figure 2.2

Prediction Models With Either Two Separate Factors, a General and a Nested Specific Factor (a), or a Single Factor (b) on the Criterion Side



Note. MRT1-MRT6 are parcels of the mental rotation test, with MRT1-MRT3 consisting of items requiring simple cardinal-axis rotation and MRT4-MRT6 consisting of items requiring complex skewed-axis rotation. Odd = Odd-Man-Out Task; Mat = Figural Matrices Task; Fig = Figure Classification Task; Fld = Paper Folding Task; Cub = Cube Construction Task

2.3 Method

2.3.1 Sample

Participants were recruited online via academic teachers and student representatives of psychology and teacher education programs at two German universities, with participants also being encouraged to disseminate the survey in their social environment. Participants needed to be at least 18 years old, not suffering from any psychiatric or neurological disease, and to have normal or corrected eyesight. Of the 557 participants who began the survey, 372 fitted these inclusion criteria and completed at least one of the six tests with a score above guessing probability. Of these, 85% were younger than 30 years (median: 23, range: 18-63) and 58% were female. Most participants reported being university students (74%), employed (25%), or in apprenticeships (4%). Reporting multiple occupations was possible, but only 7% of the participants did report more than one (all double reports were “university student” and “employed”). All participants received a 10 € coupon for completing the survey.

2.2.1 Procedure and Materials

The study protocol was in accordance with the human subjects guidelines of the Declaration of Helsinki. At the time of data collection, institutional review board approvals were neither required nor customary for these types of studies in Germany. Data collection was conducted online using LimeSurvey (LimeSurvey GmbH, 2019) hosted on a university server. For an overview of the tests used, including item numbers and time limits, see Table 2.1.

Table 2.1

Overview of Tests Administered

Construct	Test name	n items	Time limit	Source
Spatial Visualization	Mental Rotation Task	24	None	S. G. Vandenberg & Kuse (1978)
Spatial Visualization	Cube Construction Task	10	1 minute per item	Bungart (2018)
Spatial Visualization	Paper Folding Task	25	9 minutes	Heller & Perleth (2000)
Reasoning	Figure Classification Task	15	8 minutes	Heller & Perleth (2000)
Reasoning	Figural Matrices Task	15	4 minutes	Weiß (2006)
Reasoning	Odd-Man-Out Task	15	5 minutes	Weiß (2006)

2.3.1.1 Spatial Visualization Tests

2.3.1.1.1 Mental Rotation Task

Mental rotation stimuli were presented in the Vandenberg & Kuse (1978) paradigm, in which participants are asked to pick the two rotated versions of a reference stimulus (see Figure 2.1). Stimuli were taken from the stimulus library of Peters and Battista (2008) and were selected following the rules for distinct items described in Krüger and Suchan (2016): Reference stimuli were selected from the stimulus library, going through the different shapes and angular differences included in the library to get distinct items. The first of the two target stimuli was created by rotating the reference stimulus only around the horizontal x axis in half the items and only around the vertical z axis in the other half of the items (simple cardinal-axis rotation; for an example, see Figure 2.1). Mean angular disparity was 118° (range: 110° to 125°) for these stimuli. The second target stimulus was created by rotating the reference stimulus around both the horizontal x axis and the vertical z axis. This resulted in a skewed one-step rotation (complex skewed-axis rotation; see Target 2 in Figure 2.1). Mean angular disparity (for the skewed one-step rotation) in these items was 115° (range: 77° to 175°). The first distractor was a mirrored version of the reference stimulus that, in half of the items, was also rotated around both the x axis and the z axis (resulting in a complex skewed-axis rotation). The second distractor was a structurally different stimulus that was rotated around the z axis in half of the items, and around both the x axis and the z axis in the other items.

For each item, the target and distractor stimuli were displayed in randomized order. No time limit was given for attempting these items.

Responses for both target stimuli were treated as different items in the present study. Thereby, each mental rotation item yielded two bits of information, one on correct choice of the target stimulus at a simple cardinal-axis rotation, the other for correct choice of the target stimulus at a complex skewed-axis rotation of the reference stimulus. This allowed computing two scale scores from the 24 items of the mental rotation task: one by calculating the percentage of correctly chosen target stimuli for simple cardinal-axis (1-axis) rotation, the other by calculating the percentage of correctly chosen target stimuli for complex skewed-axis (2-axis) rotation. Means and standard deviations of the two scales were similar ($M_{\text{simple}} = 0.76$, $SD_{\text{simple}} = 0.16$; $M_{\text{complex}} = 0.75$, $SD_{\text{complex}} = 0.17$) and both scales showed good reliability (omega total $\omega_{\text{simple}} = .88$, $\omega_{\text{complex}} = .89$; after McNeish, 2018).

2.3.1.1.2 Cube Construction Task

For a three-dimensional test of spatial visualization we used the distractor-based Cube Construction Task created by Bungart (2018) on the basis of the distractor-free Cube Construction Task of Thissen et al. (2018). The test consists of 23 items of increasing difficulty, presented successively. In each item, a reference cube is presented from three different perspectives and the respondent needs to identify the correct unfolded cube surface among six options (see Figure 2.3). Difficulty is manipulated by varying the number of given symbols on the unfolded sheets. Empty faces on the unfolded cube surfaces could represent any symbol on the reference cube. Item difficulty is greatest with three empty faces and three given symbols.

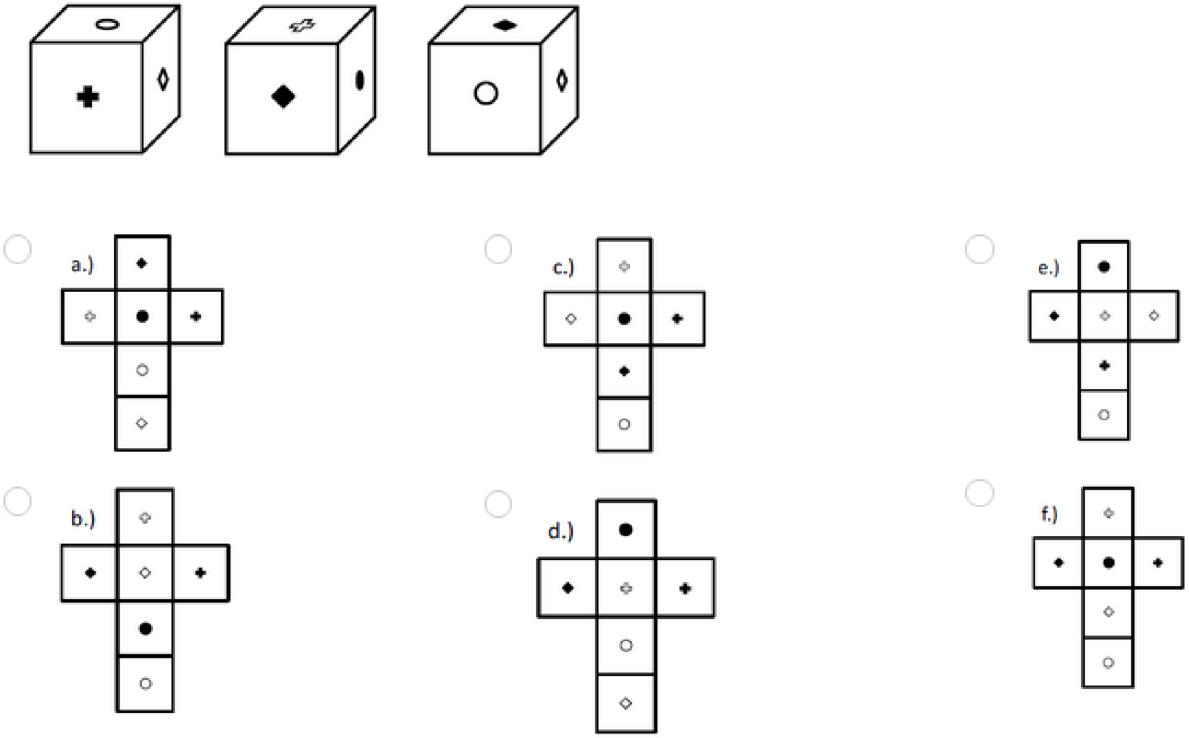
We administered a shortened version of the original test with 10 items and a time limit of 1 minute for each item. Items were selected for high discrimination and short average response times for test efficiency based on data from the original work of Bungart (2018). In our sample, the test had a mean of 0.32 (*SD* = 0.22) and showed good reliability despite the short time limit and high degree of difficulty ($\omega_{\text{cub}} = .79$).

2.3.1.1.3 Paper folding task

We used a Paper Folding Task as a classic measure of spatial visualization (Carroll, 1993). The test was taken from a battery of cognitive tests by Heller and Perleth (2000). The test had 25 items, a mean of 0.55 (*SD* = 0.21), and good reliability ($\omega_{\text{Fld}} = .86$).

Figure 2.3

Example Item of the Cube Construction Task With no Empty Faces on the Answer Options



2.3.1.2 Reasoning Tests

To measure reasoning ability, we administered a group of tests from established test batteries. This included first a Figure Classification Task (Heller & Perleth, 2000; subscale N1) with 15 items, a mean of 0.58 (*SD* = 0.19) and good reliability (ω_{Fig} = .89). Second, we used an “Odd-Man-Out” Task (Weiß, 2006; subscale 2) with 15 items, a mean of 0.75 (*SD* = 0.15) and ω_{Odd} = .68. Third, we used a Figural Matrices Test from the same battery (Weiß, 2006; subscale 3), similar to Raven’s matrices (Raven, 2000). The test had 15 items, a mean of 0.80 (*SD* = 0.16), and good reliability (ω_{Mat} = .83).

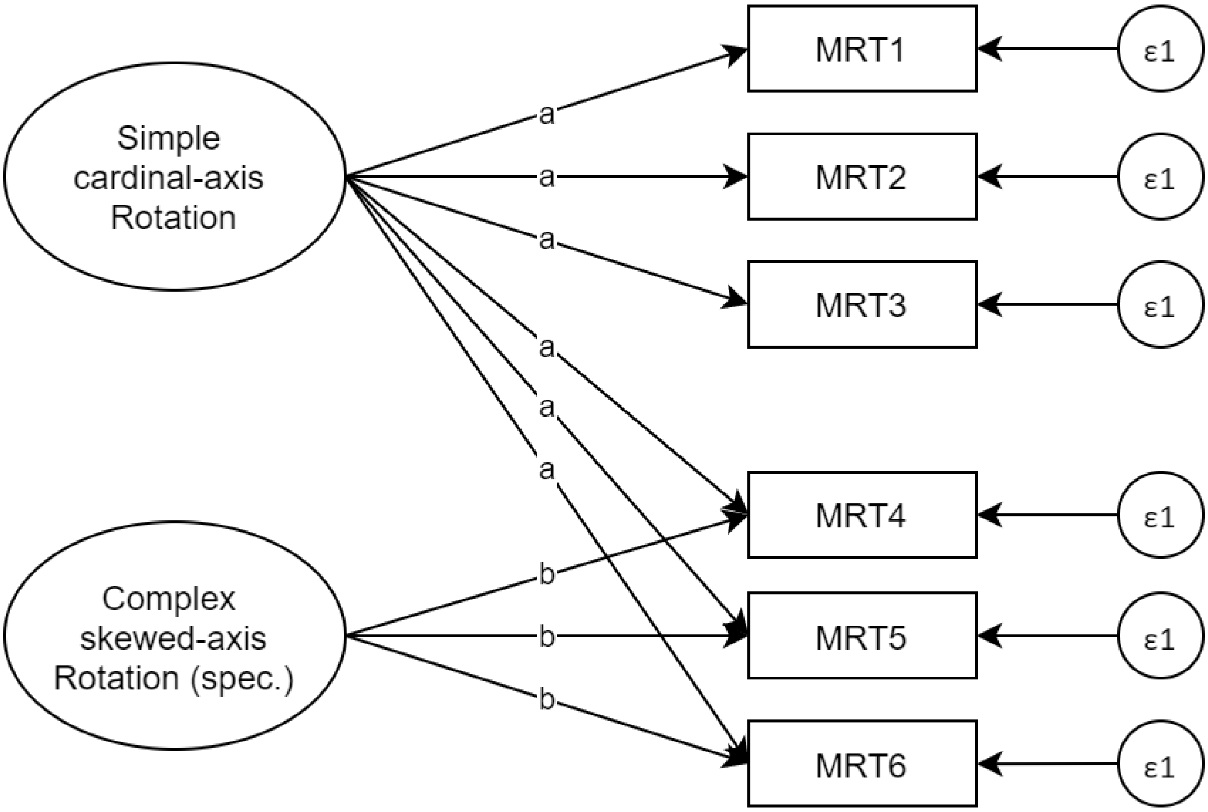
2.3.2 Data Preparation

All test items were scored as correct (1) or incorrect (0). If a participant did not complete items, due to skipping single items or to elapsed time limit, these were scored as incorrect if the participant had not skipped the test entirely. If a participant had skipped a test, their score on this test was scored as missing. As we assumed that skipped tests resulted from survey discontinuation unrelated to relevant variables, we treated these missings as “missing at random” and estimated them using a FIML estimator to use all available information from the data (Enders & Bandalos, 2001).

To test a nested-factor measurement model of the mental rotation test (Figure 2.4), its items were collapsed to form six parcels: three parcels with simple cardinal-axis rotation items (MRT1-MRT3) and three parcels with complex skewed-axis rotation items (MRT4-MRT6). Each parcel included eight items. Parcels were constructed (100,000 combinations tested) by optimizing measurement invariance in terms of equal loadings, equal intercepts, and equal residual variances of the parcels within the nested-factor measurement models. There were no constraints concerning the overall magnitude of the loadings other than that loadings on the same factor within the measurement model should be equal. The optimization criterion was the least difference between corresponding parameters in terms of a minimum p value of the chi-squared difference test. For comparison, all models used in the analyses were also investigated on the basis of parcels optimized only for equality of parameters in their own one-factor models; this yielded highly similar results.

Figure 2.4

Measurement Model for Parcel Creation



Note. MRT1-MRT3 include simple cardinal-axis rotation items, MRT4-MRT6 include complex skewed-axis rotation items.

The proportions of correct responses in each test or test parcel were used as indicators of their respective latent variables. One item had to be excluded for 42 participants, due to programming error; hence, scores for both simple cardinal-axis rotation and complex skewed-axis rotation for these participants were based on the remaining seven items, with a system missing value for the erroneous item.

2.3.3 Analyses

The difference in item difficulty of the complex skewed-axis and the simple cardinal-axis rotation scales was evaluated via a two-sample one-sided t-test (or Welch test, if variance homogeneity was violated).

We used nested-factor measurement models both for predictor variables and criterion variables. Nested-factor models allow for the separation of variance allocated for by a shared general factor and a nested specific factor of interest (see Morin et al., 2020, for details on bifactor confirmatory factor analysis), leaving only unique variance to the specific factor. This way, the reasoning factor was extracted from the two spatial visualization tests on the predictor side, and the simple cardinal-axis rotation factor was extracted from the three complex skewed-axis rotation parcels on the criteria side.

Data preparation and analyses were conducted in R (R Core Team, 2022) using the lavaan package (Rosseel, 2012). Standard errors of all models were obtained using bootstrapping (10,000 resamples). All tests and analyses used $\alpha = .05$.

2.3.3.1 Models of Mental Rotation Ability

In a first step, we compared two competing measurement models to test the incremental validity of a nested specific complex skewed-axis rotation factor as indicator of a change in cognitive processes when moving from simple cardinal-axis rotation to complex skewed-axis rotation tasks:

a) A one-factor model in which all parcels loaded on a single mental rotation factor

b) A nested-factor model consisting of a general factor on which both the simple cardinal-axis rotation parcels and the complex skewed-axis rotation parcels loaded, representing the simple cardinal-axis mental rotation processes required and used in both item types; and a specific factor on which only the complex skewed-axis rotation parcels loaded, representing unique cognitive processes that are required and used in these items, over and above those cognitive processes that are required and used in simple cardinal-axis rotation items.

In both models, loadings on the same factor, intercepts, and residual variances of corresponding parcels within the measurement model were constrained to be equal, to reduce model complexity, in particular since these item parameters were optimized to be comparable in data preparation. The relevance of the specific factor was then tested by evaluating its factor loadings and variance and by comparing the nested-factor model with the one-factor model in terms of model fit. As the models are nested, a comparison via χ^2 difference test was possible. Akaike information criterion (AIC), Bayesian information criterion (BIC), and sample-size adjusted BIC (aBIC) were also considered.

In a second step, mental rotation ability was predicted by the reasoning factor and by the spatial visualization factor. In doing so, we specified a nested-factor measurement model for the two predictors, with reasoning as the general factor and spatial visualization as the nested specific factor. This allowed investigating the specific contribution of spatial visualization in predicting mental rotation ability, over and above the contribution of reasoning.

2.4 Results

2.4.1 *Hypothesis 1: Simple Cardinal-axis Rotation and Complex Skewed-axis Rotation as Distinct Constructs*

Simple and complex rotation items did not differ in difficulty (two-sided paired t-test: $M_{\text{simple}} = 0.761$, $M_{\text{complex}} = 0.751$, $\Delta M = 0.011$, 95% CI = [-0.002;0.023], $t(371) = 1.632$, $p = .103$; d = -0.06).

Fit statistics of both the nested-factor model and the one-factor model were well within accepted fit criteria (Hu & Bentler, 1999; see Table 2.2). Model comparison showed that adding the spatial-visualization factor in the nested-factor model did not significantly improve model fit compared to the more parsimonious one-factor model ($\Delta\chi^2(1) = 3.512$, $p = .061$). Furthermore, the specific factor in the bifactor model was not adequately identified as it had weak standardized loadings ($\lambda = .23$), and its variance was not statistically significant ($s^2 = 0.002$, $SE = 0.001$, $p = .120$). This indicates that distinguishing between simple cardinal-axis rotation and complex skewed-axis rotation adds little value to the measurement model of spatial abilities for individual differences.

2.4.2 *Hypothesis 2: Prediction of Mental Rotation Factor(s) by Reasoning (g) and Spatial Visualization*

To examine whether relations with reasoning and spatial visualization differ between simple cardinal-axis rotation and complex skewed-axis rotation we compared the two prediction models depicted in Figure 2.2a and 2.2b.

Descriptive fit statistics of the prediction model with mental rotation as a single factor showed that it fitted the data well (see Table 2.3; structurally depicted in Figure 2.2b). Although the

χ^2 test was significant (which is frequently observed with large sample sizes; R. J. Vandenberg, 2006), the other fit criteria demonstrated that the model fitted the data well. Regression paths from both the general factor Reasoning and the specific factor Spatial Visualization were significant (Reasoning: $p < .001$; Visualization: $p = .022$; see Figure 2.5).

Descriptive statistics of the prediction model with a nested-factor structure of mental rotation indicated good fit as well (see Table 2.3; structurally depicted in Figure 2.2a), with the same caveat regarding the χ^2 test as with the one-factor mental rotation model. The test on fit difference of the nested-factor model and the one-factor model was not significant, indicating that both models fitted the data comparably well ($\Delta\chi^2(3) = 3.079$, $p = .380$). This indicated that estimating the additional parameters required for the more complex nested-factor model did not significantly improve model fit. Consequently, the more parsimonious one-factor model was accepted. Furthermore, the specific factor for complex skewed-axis rotation remained unidentified in the nested-factor model ($\lambda = .19$; $s^2 = 0.001$, $SE = 0.001$, $p = .262$). The lack of variance of the specific factor that can be accounted for in terms of reasoning and spatial visualization led to regression coefficients being virtually zero ($\beta_{vis} = .103$, $\beta_{reas} = .058$, ps > .770). Magnitudes of regression coefficients predicting the general factor for simple cardinal-axis rotation were similar in the nested-factor model and the one-factor model, indicating that the presence of the specific factor for complex skewed-axis rotation had negligible influence on the rest of the model.

These results led to the conclusion that the specific factor for complex rotation did not appear to have incremental value: Its inclusion did not significantly improve model fit and it captured a non-significant amount of variance that could not be accounted for either by reasoning or by spatial visualization. Therefore, we accepted the more parsimonious one-factor model with a general rotation factor depicted in Figure 2.5 as the final model. This means that complex skewed-axis rotations were found not to constitute an ability that could be distinguished from simple cardinal-axis rotation ability. These findings can be reconciled with the notion that the processes resulting in individual differences in mental rotation are either the same in both item types or at least substantially correlated.

Furthermore, regression coefficients for the general rotation factor indicated that spatial visualization still had incremental predictive value beyond the expectably strong reasoning factor.

Table 2.2

Fit Statistics of Measurement Models for Mental Rotation

Model	χ²	df	p	RMSEA	CFI	TLI	SRMR	AIC	BIC	aBIC
Nested-factor	20.97	21	0.460	0.00	1.00	1.00	0.03	-1715.46	-1691.95	-1710.98
1 factor	24.49	22	0.322	0.02	1.00	1.00	0.03	-1713.95	-1694.36	-1710.22

Note. aBIC = sample-size corrected BIC.

Table 2.3

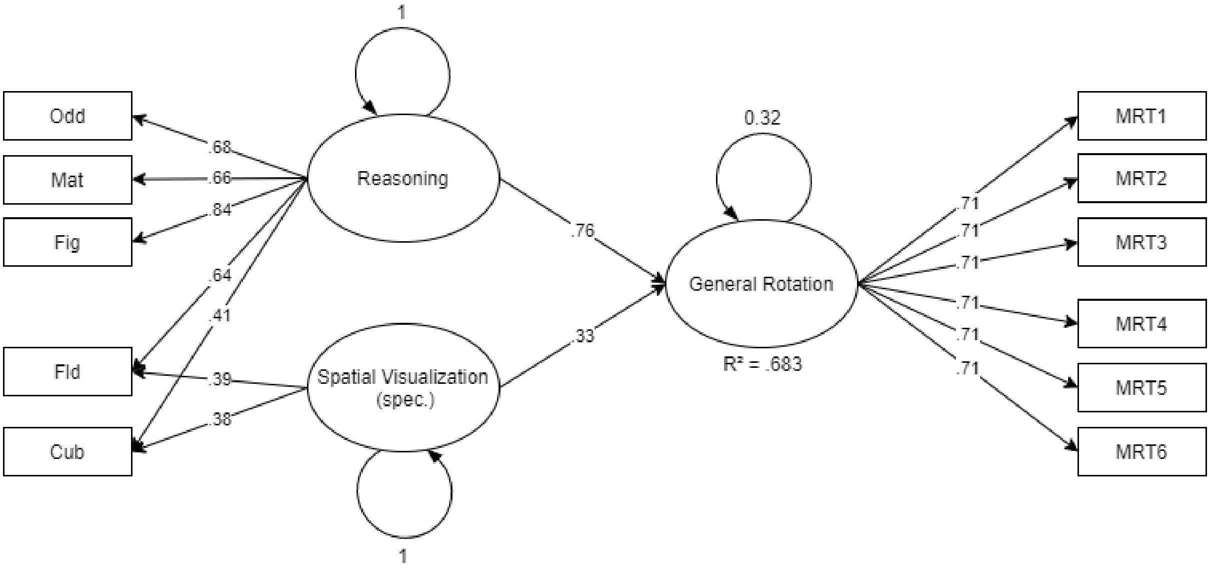
Fit Statistics of Prediction Models for Mental Rotation

Model	χ²	df	p	RMSEA	CFI	TLI	SRMR	AIC	BIC	aBIC
Nested-factor	83.97	51	0.002	0.04	0.98	0.98	0.04	-3135.11	-3033.22	-3115.71
1 factor	87.05	54	0.003	0.04	0.98	0.98	0.04	-3138.03	-3047.90	-3120.87

Note. Model names describe the measurement model used for mental rotation (as evaluated in Table 2.2) in the respective prediction model. aBIC = sample-size corrected BIC.

Figure 2.5

Predictive Model With General Rotation as a Single Factor



Note. Shown paths are standardized βs and are significant with α = .05.

2.5 Discussion

We investigated the popular Vandenberg and Kuse (S. G. Vandenberg & Kuse, 1978) mental rotation paradigm, which is frequently used for the assessment of individual differences in mental rotation ability. The test format builds on the R. N. Shepard and Metzler (1971) mental rotation test. However, most researchers compile their own tests using the Peters and Battista (2008) stimulus library. Consequently, test versions may differ in their processing requirements, and specific factors potentially affecting test performance are rarely considered in individual differences research.

Naturally, it could be hypothesized that similar factors would affect all forms of rotation test performance as those that were demonstrated for the R. N. Shepard and Metzler (1971) items (e.g., angular disparity, stimulus complexity or rotational complexity via rotational axes). This raises the question whether different test variants measure the same ability. A likely relevant factor in this respect is the number and type of rotational axes required. In typical mental rotation items (e.g., as summarized in S. Shepard & Metzler, 1988), stimuli need to be rotated around a single cardinal rotational axis. Usually, such a cardinal rotation axis can be easily derived either from the picture plane or from structural features of the stimulus (e.g., the edges of a cube figure). In such simple items, mental rotation around a cardinal axis may be the most central processing requirement. Conversely, more complex test items could be assumed to call for additional component processes, such as switching between different cardinal axes (i.e., the multiple-step rotation-by-dimension strategy) or searching for the one skewed rotational axis that allows for a one-step rotation to match the reference and target stimuli (i.e., complex skewed-axis rotation).

Previous studies testing the influence of rotational axes have revealed that complex skewed-axis rotations are more difficult, in terms of longer response times (Parsons, 1987); only high-spatial participants were able to use the skewed-axis rotation strategy (Just & Carpenter, 1985). However, the effect of rotational axes on the factor structure of mental rotation tests has not yet been investigated yet from an individual-differences perspective. This was the first goal of the present study.

To this end, we adopted a cognitive processes perspective and manipulated item characteristics accordingly (i.e., the required axes of rotation). Then, we tested differential effects of this manipulation using nested-factor analyses. This well-established method (Brunner et al., 2012; Chen et al., 2006) allows to test whether – from an individual-differences perspective – component processes or abilities can be distinguished from each other. Furthermore, nested-factor analyses allow inspecting differential relations of the separated factors with other variables in the nomological network of relevant constructs (Messick, 1989).

Our results show that there is no specific ability for items requiring complex (skewed-axis) rotation, over and above general rotation ability. Specifically, a general factor of mental rotation ability accounted for individual differences in test performance in all rotation items. Conversely, there was no incremental factor capturing the specific requirements of complex skewed-axis rotation. This suggests that complex skewed-axis rotation does not constitute an ability on its own. This finding is of practical relevance for test construction. It suggests that the type of rotation (simple cardinal-axis rotation or complex skewed-axis rotation) does not affect the construct validity of the test. Consequently, our original concern that different test versions using different subsets of the Peters and Battista (2008) stimuli may differ in processing requirements was not supported. Conversely, it appears that the same ability is measured regardless of the required rotation requirements.

Furthermore, we did not find a meaningful mean difference in accuracy between simple cardinal-axis rotation and complex skewed-axis rotation items. This supports the supposition that distinguishing between both item types is not mandatory for the assessment of individual differences in mental rotation ability. This finding is well in line with Just and Carpenter's (1985) finding that there is no difference between rotating a cube once by 180° around one cardinal axis and rotating it twice, by 90° each time, around two different cardinal axes. Our results support that this finding also applies to more complex objects and to rotations around skewed axes not perpendicular to any feature of the rotated stimulus. (In Just and Carpenter’s cubes, even skewed axes were always perpendicular to either faces, edges or corners of the cubes.)

However, our results did not confirm Parsons's (1987) important finding that rotations around skewed axes are more difficult, as reflected in slower reaction times in Parson’s study. It is possible that this difference between simple cardinal-axis rotation and complex skewed-axis rotation is confined to reaction time data and has no effect on accuracy in untimed conditions. This difference could be due to differences in sensitivity or validity of the respective performance metric for cognitive processing requirements. These possibilities should be addressed in future research.

The second goal of the current study was to evaluate whether the Vandenberg and Kuse mental rotation test indeed necessitates a specific spatial-visualization ability over reasoning ability. This could be argued to be a critical threat to construct validity as previous research has shown (e.g., Hegarty (2018) that use of reasoning-based strategies can lead to better performance in these tests than spatial strategies alone. In order to test incremental validity, we fitted another nested-factor model to the tests we used as predictors. This allowed modeling a general factor of reasoning and an incremental factor capturing visualization-specific task requirements.

Next, rotation ability as assessed by the Vandenberg and Kuse test was regressed onto these two factors. The general factor of reasoning substantially predicted mental rotation ability. This was to be predicted, as reasoning tests usually reveal a substantial positive manifold (Carroll, 1993). This also holds for spatial ability tests in particular (Lohman, 1996). It is important to note, however, that the specific spatial-visualization factor incrementally predicted mental rotation ability. This finding confirms that there is a spatial-visualization component in the Vandenberg and Kuse paradigm, and that this component can be demonstrated over and above the substantial relation with general (reasoning) ability that has been shown for the Shepard and Metzler paradigm (see Mary Hegarty's work, e.g., Boone & Hegarty, 2017; Hegarty, 2018). This indicates that items are not exclusively solved by using analytical strategies such as feature matching, and spatial strategies such as mental rotation are used instead, at least to some extent.

2.5.1 Limitations

In the Vandenberg and Kuse paradigm, participants know the number of correct stimuli per item. This led to two possible limitations in our study: First, responses to the two targets within each item were not fully independent of each other. This could have increased the estimated correlation between the two rotation factors. However, it could not account for their substantial relationship in the magnitude of identity. Nonetheless, this limitation could be alleviated in future research by presenting only one type of stimulus in each item (simple cardinal-axis or complex skewed-axis rotation).

A second limitation resulting from the number of correct answers per item being known is that it allowed solving an item by exclusion of distractors. If two distractors had been identified with sufficient certainty, participants did not need to look at the other two stimuli as these had to be the targets. This is especially problematic for the mental rotation test as a measure of spatial abilities as distractors (especially structure foils) can be detected via reasoning-based feature matching strategies more easily than targets, as it is easier to spot a single difference than to check the whole stimulus for identity. Therefore, distractors can be excluded by using both spatial strategies and analytical, reasoning-based strategies. Hegarty (2018) discussed this and questioned whether the mental rotation test following the Vandenberg and Kuse paradigm actually measures (exclusively) mental rotation ability: In her study she found that usage of some analytic strategies (namely what she called the "global shape strategy") also led to increased performance compared to most other reported strategies besides holistic rotation. Therefore, it may be possible that the test also measures the ability to find efficient analytical strategies which could account for its relations with reasoning abilities. Conversely, guessing based on an elimination strategy alone would be difficult to reconcile with the high internal consistencies (omega total $\omega_{simple} = .88$, $\omega_{complex} = .89$) observed in this study. Furthermore, as the specific factor for spatial visualization ability still predicted mental rotation incrementally over and above reasoning, supporting that the test taps a spatial ability. However, we cannot rule out that participants also used analytical strategies (such as feature matching) or response elimination to supplement the spatial strategies.

There was no evidence of a specific factor for complex skewed-axis rotation. However, even if it had emerged it could not be unraveled whether participants had used the expected one-step rotation around the skewed axis for those items or had applied a two-step rotation-by-dimension strategy around two separate cardinal axes. Both the search process for the skewed axis required for the one-step rotation strategy and the process of switching between cardinal rotation axes in the two-step rotation strategy could have explained the emergence of a separate ability factor. In future research, self-reporting of solving strategy might give clues as to the strategies used by participants.

2.5.2 Conclusion

The present study contributes to a better understanding of mental rotation ability as an individual-differences construct. Using a psychometric approach applied to a large sample of participants, we showed that items in a mental rotation test following the Vandenberg and Kuse paradigm were accounted for by a single general factor. Conversely, there was no evidence of a specific processing ability required for the skewed-axis rotation items. Hence, mental rotation appears to be a unitary factor. Furthermore, we confirmed that the Vandenberg and Kuse paradigm measures spatial-visualization ability over and above reasoning ability. The implication for assessment purposes is that the Vandenberg and Kuse test measures spatial-visualization ability. Further, this ability can be assessed comparably regardless of whether simple cardinal-axis rotation or complex skewed-axis rotation items are used.

2.6 References

Arendasy, M. E., & Sommer, M. (2013). Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence*, *41*(4), 234–243. https://doi.org/10.1016/j.intell.2013.03.006

Boone, A. P., & Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process! *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(7), 1005–1019. https://doi.org/10.1037/xlm0000370

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749.x

Bungart, M. (2018). *Entwicklung und Evaluation eines Manipulationsansatzes der Itemschwierigkeit in dem distraktorbasierten Format des Würfelkonstruktionstests* [Development and evaluation of a manipulation approach to item difficulty in the distractor-based format of the cube construction task] [Unpublished bachelor's thesis]. Universität des Saarlandes, Saarbrücken.

Burin, D. I., Delgado, A. R., & Prieto, G. (2000). Solution strategies and gender differences in spatial visualization tasks. *Psicológica*, *21*(2), 275–286. https://www.redalyc.org/articulo.oa?id=16921204

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology*, *7*(1), 20–43. https://doi.org/10.1016/0010-0285(75)90003-1

Doyle, R. A., & Voyer, D. (2018). Photographs of real human figures: Item types and persistent sex differences in mental rotation. *Quarterly Journal of Experimental Psychology (2006)*, *71*(11), 2411–2420. https://doi.org/10.1177/1747021817742079

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. https://doi.org/10.1207/s15328007sem0803_5

Flanagan, D. P., & Dixon, S. G. (2013). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Wiley online library. Encyclopedia of special education: A reference for the education of children adolescents and adults with disabilities and other exceptional individuals* (4th ed.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118660584.ese0431

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating “rotators” from “nonrotators” in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research*, *41*(3), 261–293. https://doi.org/10.1207/s15327906mbr4103_2

Glück, J., Machat, R., Jirasko, M., & Rollett, B. (2002). Training-related changes in solution strategy in a spatial test: An application of item response models. *Learning and Individual Differences*, *13*(1), 1–22. https://doi.org/10.1016/S1041-6080(01)00042-5

Hausmann, M., Slabbekoorn, D., van Goozen, S. H. M., Cohen-Kettenis, P. T., & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience*, *114*(6), 1245–1250. https://doi.org/10.1037//0735-7044.114.6.1245

Hegarty, M. (2018). Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychonomic Bulletin & Review*, *25*(3), 1212–1219. https://doi.org/10.3758/s13423-017-1347-z

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, *19*, 61–70. https://doi.org/10.1016/j.lindif.2008.04.006

Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology (2006)*, *61*(5), 683–689. https://doi.org/10.1080/17470210701822967

Heller, K. A., & Perleth, C. (2000). *KFT 4-12+ R - Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision* [KFT 4-12+ R - Cognitive ability test for 4th to 12th grades, revision]. Beltz.

Hirnstein, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences*, *19*(2), 225–228. https://doi.org/10.1016/j.lindif.2008.11.006

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review*, *92*(2), 137–172. https://doi.org/10.1037/0033-295X.92.2.137

Khooshabeh, P., & Hegarty, M. (2010). Representations of shape during mental rotation. In Association for the Advancement of Artificial Intelligence (Ed.), *2010 AAAI Spring Symposium Series* (pp. 15–20). AAAI Publications.

Krüger, J. K., & Suchan, B. (2016). You should be the specialist! Weak mental rotation performance in aviation security screeners – reduced performance level in aviation security with no gender effect. *Frontiers in Psychology*, *7*, 333. https://doi.org/10.3389/fpsyg.2016.00333

LimeSurvey GmbH. (2019). *LimeSurvey: An Open Source survey tool* (Version 3.20.2+191119) [Computer software]. Limesurvey GmbH. Hamburg, Germany. http://www.limesurvey.org

Lohman, D. F. (1996). Spatial ability and G. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Lawrence Erlbaum Associates, Inc.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement, 3rd ed* (pp. 13–103). American Council on Education.

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and bifactor-ESEM. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (4th ed., pp. 1044–1073). Wiley. https://doi.org/10.1002/9781119568124.ch51

Muto, H., & Nagai, M. (2020). Mental rotation of cubes with a snake face: The role of the human-body analogy revisited. *Visual Cognition*, *28*(2), 106–111. https://doi.org/10.1080/13506285.2020.1727598

Parsons, L. M. (1987). Visual discrimination of abstract mirror-reflected three-dimensional objects at many orientations. *Perception & Psychophysics*, *42*(1), 49–59. https://doi.org/10.3758/BF03211513

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, *57*(2), 176–184. https://doi.org/10.1016/j.bandc.2004.08.052

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260–264. https://doi.org/10.1016/j.bandc.2007.09.003

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Rogister, F., Pottier, L., El Haddadi, I., Monseur, J., Donneau, A.-F., Diep, A. N., Camby, S., Defaweux, V., Bonnet, P., Tombu, S., Lefebvre, P., & Poirrier, A.-L. (2021). Use of Vandenberg and Kuse mental rotation test to predict practical performance of sinus endoscopy. *Ear, Nose & Throat Journal*, 1455613211000599. https://doi.org/10.1177/01455613211000599

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Shepard, S., & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 3–11. https://doi.org/10.1037/0096-1523.14.1.3

Sorby, S. A., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences*, *67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Spearman, C. E. (1904). "General intelligence" - objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*(2), 219–234. https://doi.org/10.1016/j.learninstruc.2007.01.012

Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D., & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology*, *110*(8), 1160–1174. https://doi.org/10.1037/edu0000258

Thissen, A., Koch, M., Becker, N., & Spinath, F. M. (2018). Construct your own response: The cube construction task as a novel format for the assessment of spatial ability. *European Journal of Psychological Assessment*, *34*(5), 304–311. https://doi.org/10.1027/1015-5759/a000342

Thurstone, L. L. (1938). *Primary mental abilities*. The University of Chicago Press.

Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods*, *9*(2), 194–201. https://doi.org/10.1177/1094428105285506

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin & Review*, *18*(2), 267–277. https://doi.org/10.3758/s13423-010-0042-0

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Weiß, R. H. (2006). *CFT 20-R Grundintelligenztest Skala 2* [CFT 20-R Base intelligence test scale 2]. Hogrefe.

Wohlschläger, A [Andreas], & Wohlschläger, A [Astrid] (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 397–412. https://doi.org/10.1037/0096-1523.24.2.397

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19. https://doi.org/10.1162/jocn.2008.20013

Ziemek, T., Creem-Regehr, S., Thompson, W. L., & Whitaker, R. (2012). Evaluating the effectiveness of orientation indicators with an awareness of individual differences. *ACM Transactions on Applied Perception*, *9*(2), 1–23. https://doi.org/10.1145/2207216.2207218

3 Studie 2: Cross-Disciplinary Impact of Spatial Visualization Ability on Study Success in Higher Education

3.1 Abstract

Spatial ability emerged as a key factor for success in STEM fields. But so far there are no studies on differences in this effect between different STEM domains. The current study used content knowledge tests at the beginning (pretest) and the end of the semester (posttest) as comparable measure for study success in civil engineering (*n* = 247), chemistry (*n* = 425), and biology (*n* = 325) first-year university students from eight German universities. To evaluate the effect's specifity for STEM domains, we also included social science (*n* = 360) as a non-STEM domain.

A multiple regression analysis confirmed that spatial ability is an incremental predictor of posttest content knowledge beyond high school GPA, reasoning ability, and pretest content knowledge. Furthermore, multiple-group mediation modeling found that spatial ability is related to prior knowledge and high school GPA, resulting in substantial indirect effects that varied between domains while the direct effect was stable.

Our results add a direct comparison across different domains to the extensive body of literature on the effect of spatial ability on success in STEM domains: This effect is stronger in engineering (total *b* = .81) and chemistry (total *b* = .83) than in biology (total *b* = .68) and social science (total *b* = .51). However, the substantial effect even in the non-STEM domain of social science indicates that spatial ability might be relevant for higher education in general. Mathematical skills and visual model comprehension skills are discussed as possible connections between spatial ability and study success.

This manuscript is the preprint of the following publication:

Nolte, N., Fleischer, J., Spoden, C., & Leutner, D. (2024). Cross-disciplinary impact of spatial visualization ability on study success in higher education. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/edu0000847

3.2 Introduction

Science, Technology, Engineering, and Mathematics (STEM) study programs suffer from high dropout rates (e.g., Chen, 2009; Heublein et al., 2020; Price, 2010). Shortages of skilled workers in STEM fields could threaten economic growth and innovation (National Academies Committee on Science, Engineering, and Public Policy [COSEPUP], 2007) and provide an ongoing challenge for educational systems (Uttal & Cohen, 2012). Naturally, identifying factors relevant for study success in STEM fields, and potential ways of supporting such success, could prove essential to prevent such shortages, and should be a focus of educational research. According to Heublein and Schmelzer (2018), performance deficits are the main driving force behind dropout, as STEM students most frequently cite trouble in reaching the study program's performance requirements as their reason for dropping out. As the major outcome of academic study programs is the acquisition of domain-specific content knowledge (York et al., 2015), this variable is of primary interest as an indicator of study success. Identification of predictors for domain-specific content knowledge could help identify both exceptionally competent and promising students for academic promotion, as well as students at risk of dropout, so that tailored counseling and tutoring can be provided.

Intelligence and high school grade point average (GPA) are well-established as the main predictors of study success (e.g., Burton & Ramist, 2001; Neisser et al., 1996). Additionally, several studies have demonstrated the incremental validity of prior content knowledge as a predictor of study success (Formazin et al., 2011; Freyer et al., 2014; Hailikari, 2009). In some studies prior content knowledge outperformed intelligence (e.g., a prior knowledge test at the beginning of the semester accounted for 35% of the variance in exam grades compared to 10% for general intelligence; Formazin et al., 2011) and had a comparable effect to GPA (Hailikari, 2009).

In recent years, however, spatial ability has attracted increasing interest as another factor especially relevant for success in STEM programs (Shea et al., 2001; Wai et al., 2009) in general. Some research has also highlighted the effects of spatial ability on success in specific domains such as chemistry (e.g., Talley, 1973; Wu & Shah, 2004), engineering (e.g., Duesbury & O'Neil, 1996; Sorby et al., 2018), physics (e.g., Kozhevnikov et al., 2007 Pallrand & Seeber, 1984), or biology (e.g., Rochford, 1985; Russell-Gebbett, 1985). However, to our knowledge, there is no research so far comparing the effect of spatial ability on success in different STEM domains: Many of the previous studies focused on a single domain only and used widely different outcomes. The few more general studies used binary outcomes (e.g., later occupation in a STEM domain or in a non-STEM domain), without being able to compare different STEM domains, or only compared spatial ability levels between domains – in both cases there is no variance in measures of STEM success within the domains that could be accounted for in more detail by spatial ability.

Therefore, neither the domain-specific studies nor the general studies allow for the comparison of effects of spatial ability on STEM success between different STEM domains: Previous studies either allowed for one (analyses of effects within a single domain) or the other (comparison of ability levels between domains; predicting STEM vs. non-STEM occupations). To address this gap, the present study collected data from multiple STEM domains and a non-STEM domain and used comparable measures of study success in each of the domains.

Additionally, previous literature often used spatial ability as sole predictor of STEM success, with only a few studies adding further control variables such as intelligence. The present study extended previous research by also controlling for knowledge variables such as domain-specific prior content knowledge and high school GPA. This allows for drawing stronger conclusions on the strength of the effect of spatial ability on STEM success, and for comparing the effect of spatial ability with the effect of other performance-related variables.

3.2.1 Spatial Ability and STEM

Intelligence in general is known as a major predictor of educational outcomes, such as school performance or years of education (Neisser et al., 1996). Spatial ability has long been neglected as a relevant facet of intelligence for identifying gifted individuals (Lubinski, 2010), possibly based on beliefs such as Spearman’s (1950), that “spatial tests merely […][are] unreliable measures of G” (cited in Lohman, 1996, p. 2) and do not constitute a meaningful separate factor of intelligence. Later models such as Carroll's (1993) factor analytic classification of cognitive abilities acknowledge that ability tests share a general intelligence factor *g* but still include multiple primary factors below the general factor – one of these factors being spatial ability. In case of spatial ability, Lohman (1996) argues that the overlap between spatial ability and measures of *g* such as reasoning tests may stem from the working memory demands of maintaining and transforming images and creating mental models: Such demands are central to spatial tests, in common with most reasoning tests that commonly require creating and manipulating mental models of rules and patterns (Johnson-Laird, 1983; Kyllonen & Christal, 1990). On this basis, Lohman (1996) suggests that Spearman’s conclusion could be turned around just as well: “Measures of G are largely unreliable measures of the ability to generate and transform different types of mental models in working memory” (Lohman, 1996, p. 12).

In the present study we followed Carroll's (1993) factor analytic classification of spatial ability as a multi-faceted construct with five different factors. We focused on the spatial visualization factor, which is defined as “processes of apprehending, encoding, and mentally manipulating spatial forms” (Carroll, 1993, p. 309). This is an essential process in handling STEM fields’ typical problems using spatial models, such as comprehending molecular models in a chemical synthesis process (Wu & Shah, 2004), understanding cross-sectional diagrams of internal organs (Hegarty et al., 2009), or mentally reconstructing landscapes from visible eroded remains in geology (Hambrick et al., 2012). Indeed, previous research identified the spatial visualization facet as the most relevant facet for many STEM domains (e.g., Hsi et al., 1997; Lowrie et al., 2017; Miller & Halpern, 2013; Shea et al., 2001; Sorby, 2009; Wai et al., 2009; Wu & Shah, 2004; see Uttal & Cohen, 2012, for a review). The impact of spatial ability on success in STEM is not confined only to complex content at university level but has been shown in earlier stages of education as well – mainly for mathematics, as it is the most prevalent STEM subject in school. For example, Mix et al. (2021) found an effect of spatial ability training on mathematics achievement in first and sixth grade students. The effectiveness of this training in early stages of education is also in line with the conclusion from Uttal and Cohen's (2012) review that spatial ability predicts STEM learning best when students' level of expertise is low, and becomes less important with increased expertise.

In general, cognitive abilities are more important for performance when domain knowledge is low (Hambrick & Meinz, 2011). This also applies to spatial ability and STEM performance, as greater domain knowledge may allow people to use "mental representations […] to solve problems without having to use spatial thinking" (Uttal & Cohen, 2012, p. 148). For example, Hambrick et al. (2012) found an interaction of spatial ability (measured mainly via spatial visualization-related tests) and geological knowledge in predicting performance on a geological bedrock mapping test: Performance in experienced geologists with high domain knowledge differed only marginally between different levels of spatial ability, while performance in novices with low geological knowledge varied largely based on their spatial ability. This leads to the conclusion that, with respect to studying the effects of spatial ability on STEM study success, "earlier is better" (Uttal & Cohen, 2012), indicating that the study entry phase should be the focus. This matches conclusions drawn from studies on student dropout: Most students drop out in the first year (e.g., Aulck et al., 2016); this again points to the freshman year as a prime target for research and intervention.

Spatial ability interventions appear to be an approach worth pursuing, as the meta-analysis of Uttal, Meadow, et al. (2013) has demonstrated the malleability of spatial abilities in different interventions. The authors further concluded that applying a spatial ability intervention with an average effect size of Hedge's $g = 0.4$ found in the meta-analysis could double the number of people able to match the average spatial ability of engineering bachelor graduates at the population level. Although such a type of large scale training has not been tested yet, a few studies have applied spatial ability trainings to improve STEM performance – most notably the works of Mix and colleagues (2016; 2021) for mathematics and the works of Sorby and colleagues (2000; 2013; 2018) for engineering. These experimental studies support our interpretation of the effect of spatial ability on STEM performance as a causal relationship in at least some STEM domains.

Furthermore, there appears to be an “ability-preference fit with pursuing STEM careers” (Webb et al., 2007, p. 409) in individuals with high spatial abilities. In Wai et al.'s (2009) large longitudinal study, high school students, who achieved STEM degrees 11 years later, displayed almost exclusively above-average spatial abilities: Around 70% of STEM bachelor graduates, around 75% of STEM master graduates and almost 90% of STEM PhDs even scored in the study’s top quartile for spatial ability. In the same study, similar patterns were found in favorite school subjects, as students who chose math or science courses as their favorite displayed higher spatial ability than those who chose other courses. Accordingly, improving the spatial abilities of students might also increase their interest in STEM subjects in general.

Even though there appears to be a helpful effect of spatial ability on STEM outcomes in general, differences between STEM domains may exist: Shea et al. (2001) assessed patterns of cognitive ability dimensions (mathematical, verbal and spatial) between different bachelor’s degree groups and found substantial differences in spatial ability means, even between STEM fields. Engineering and physical science students had the highest average spatial ability scores, math and computer science students scored lower but still above average, whereas biology students scored below average, halfway between the math/computer science group and the non-STEM fields. Although these differences include strong effects of self-selection and include no detailed measure of study success other than the university degree, this suggests a “spatial ability hierarchy” across study programs. This pattern appears to be stable, as Shea et al. (2001) found similar patterns for subjects in high school and occupational groups at an age of 33 years.

3.2.2 The Present Research: Aims and Hypotheses

Previous research on the importance of spatial ability for performance in STEM domains mostly focused on single domains (e.g., engineering; Sorby et al., 2000; 2010; 2013; 2018) and used measures specifically aligned to the content of the particular domain (or even a particular study program), which makes it hard to compare effects between different domains. Research comparing effects across multiple domains has relied on more general (and often binary) and individual outcomes, such as graduating in a particular domain or starting a particular profession (e.g., Humphreys et al., 1993; Shea et al., 2001; Wai et al., 2009). This is a constraint against making exact comparisons across STEM domains. Therefore, the current study aimed to address these issues by studying *different* STEM domains with *comparable* and *performance-based* outcome variables rather than, e.g., exam grades, domain of graduation, or type of vocational profession.

Three different STEM domains were included in our sample of university students: Chemistry and Civil Engineering as “typical” STEM domains and Biology as an “atypical” domain regarding the aforementioned cognitive ability patterns (Shea et al., 2001; Wai et al., 2009). We further included

Social Science as a non-STEM domain in our sample, to evaluate the specificity of spatial ability for STEM fields. We used content knowledge as an indicator of study success because it allows for standardized testing at different points in time following a controlled pre-post design, as opposed to academic exams, which represent singular measurements that do not allow measuring prior knowledge before going into the semester. Using standardized content knowledge tests also allows to validly compare test results across different student cohorts, as exams vary greatly among universities and cohorts. Besides replicating previous studies observing spatial ability as a strong predictor of study success in STEM domains, our study aimed, using multiple regression, to consolidate its incremental predictive role beyond other cognitive and knowledge-based predictors of STEM performance such as prior content knowledge, high school GPA, and intelligence. Including measures of general intelligence also allowed testing whether the predictive value of spatial ability results from its overlap with general intelligence or whether the spatial component has incremental predictive value. As special emphasis was placed on differences between domains, these were likewise included in the regression models. Since spatial visualization was expected to affect content knowledge acquisition at university, it was also assumed to have affected the acquisition of prior content knowledge in high school. Effects on high school GPA were also expected, as spatial ability is documented as being related to performance in different high school subjects – most notably mathematics (e.g., Bruce & Hawes, 2015; Judd & Klingberg, 2021; Lowrie et al., 2019) but also in chemistry (Barke & Engida, 2001), physics (Delialioğlu & Aşkar, 1999), general physical sciences (Ganley et al., 2014), and even English (Satterly, 1976). As prior content knowledge and high school GPA are also known predictors of study success, we expected indirect effects of spatial ability on acquired content knowledge via both prior content knowledge and high school GPA. These effects can be tested in a mediation model and differentiated between domains by conducting multiple-group modelling. On the basis of the results of Shea et al. (2001) regarding cognitive ability patterns in different domains we expected substantial differences in the strength of effects between different domains. In summary, this results in the following hypotheses:

1. Spatial visualization ability is an incremental predictor of first-year university student's content knowledge acquired at the end of the first semester (posttest), beyond prior content knowledge (pretest), high school GPA, and intelligence. The strength of incremental prediction varies between domains of study.
2. Spatial visualization ability is a predictor of first-year university student's high school GPA and their content knowledge acquired at high school (pretest), resulting in indirect effects of spatial visualization ability on acquired content knowledge (posttest) through high school GPA and content knowledge (pretest) as mediators. The strength of indirect prediction varies between domains of study.

3.3 Methods and Data Sources

3.3.1 Transparency and Openness

The present research was conducted in accordance with the human subjects guidelines of national research committees in Germany, as well as the APA standards for Ethical Principles of Psychologists and Code of Conduct. At the time of data collection, institutional review board approvals were not required for this type of longitudinal study in Germany.

We report how we determined our sample size, all data exclusions, all manipulations, and all measures relevant for the study, and we follow JARS (Kazak, 2018). All data are available upon request from the corresponding author. The study design followed the grant proposal of the project this study was part of (see Fleischer et al., 2019, for a more detailed description), which we extended with a spatial visualization test. Study analyses were not preregistered. Code for the data analysis and non-licensed research materials are available upon request from the corresponding author.

3.3.2 Sample and Procedure

A sample of *N* = 1,394 first-year university students ($n_{\text{civil engineering}}$ = 247; $n_{\text{chemistry}}$ = 425; n_{biology} = 362; $n_{\text{social science}}$ = 360) from eight German universities was surveyed at two points of measurement (PoM) during their first semester in the winter semesters 2018/2019 and 2019/2020 in a nonexperimental, multiple-group comparison design. This study used the nonexperimental cohorts from the overall research project it was part of for the STEM samples. The social science sample came from a related project using the same study design (only university cohorts; Walpuski et al., 2021). Sample size for this study was defined by the data collection of the overall projects. Where possible, data collection took place in regular courses, otherwise timeslots that suited the participants' schedules were used. Only participants that studied either one of the four domains (civil engineering, chemistry, biology, social science) or a subdomain of the four domains (requiring to take most of the same courses in the first semester; e. g., biochemistry for chemistry, molecular and applied biotechnology for biology, or political science for social science) were included. Non-first-year students and students with two majors (such as teacher candidates) were excluded.

In total, 50% of participants identified themselves as female (engineering 36%, chemistry 44%, biology 64%, social science 55%), 80% were German native speakers (engineering 64%, chemistry 80%, biology 90%, social science 81%). The participants' average age was 20.0 years (*SD* = 2.68; range: 17-64; quantiles: 5% = 18, 95% = 24), with slight variation between the domains (*M*/*SD* engineering 20.2/3.09; chemistry 19.6/1.87; biology 19.7/1.77; social science 20.6/3.68).

PoM 1 took place at the beginning of the semester and included a demographic questionnaire asking, among others, for high school grades, cognitive ability tests for spatial visualization ability and reasoning ability, and domain-specific content knowledge tests. It should be noted that grades in Germany range from 1 to 6 (1 being the best). This results in negative coefficients for high school GPA in the models reported below. PoM 2 took place at the end of the semester and included the same test for domain-specific content knowledge. The demographics questionnaire, the reasoning ability test and the content knowledge tests were conducted as paper-pencil tests at the respective universities, while the tests of spatial visualization ability were conducted online using LimeSurvey (LimeSurvey GmbH, 2019).

3.3.3 Data Preparation

All data preparation and analyses were performed with R (version 4.1.2; R Core Team, 2022) using RStudio (version 2022.2.2.492; RStudio Team, 2023) and the tidyverse package environment (Wickham et al., 2019). All tests were scaled according to the Rasch model using the R package TAM (version 4.0-16; Robitzsch et al., 2022). Weighted likelihood estimates (WLE; Warm, 1989) were calculated as person ability indicators and used in the analyses. Participants were included in test scaling and scored if they completed at least half of the items of the respective tests to ensure a valid measurement. Prior to analysis, all variables were z-standardized within academic domains.

3.3.4 Materials

3.3.4.1 Spatial Visualization Ability

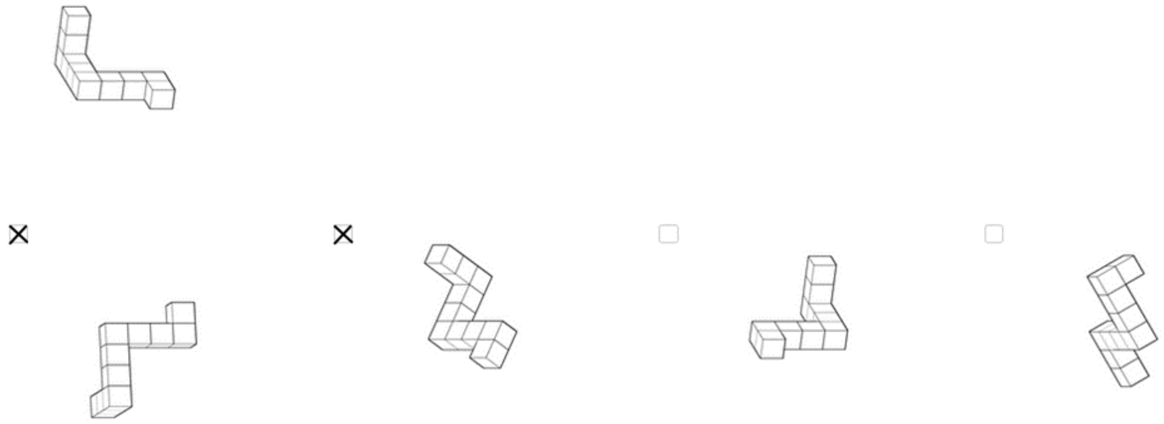
To assess spatial visualization ability, we used a mental rotation test comprising 24 items and following Vandenberg and Kuse's (1978) mental rotation paradigm (WLE reliability = .84). The test requires the participants to detect two rotated versions of a target stimulus out of four options (Figure 3.1). Participants could select exactly two options and items were scored as correct only if the participants selected both correct options. Stimuli were taken from the stimulus library of Peters and Battista (2008) and were selected similarly to Krüger and Suchan (2016). As the test requires rotation of complex 3D block figures across multiple axes, and was applied without a time limit, it could be expected to measure spatial visualization rather than spatial relation (Carroll, 1993). Previous studies showed that this particular test variant includes both a general intelligence component (due to positive manifold; e. g., Carroll, 1993; Spearman, 1904) and an incremental spatial visualization component (Nolte et al., 2022).

3.3.4.2 Content Knowledge

For each academic domain, content knowledge was measured at the beginning of the semester (“pretest content knowledge”) and at the end of the semester (“posttest content knowledge”) with the same domain-specific test. The tests for the different domains were developed in a parallel way in previous projects, following the same guidelines and a theoretical rationale based on Walpuski et al. (2010), making them comparable. They consisted of items both from domain-specific topics covered in high school and from domain-specific topics covered in introductory courses at the university (see Table 3.1). This made it possible to assess both prior content knowledge acquired before university, mostly from high school, at the first PoM, and content knowledge acquired during the first semester at the second PoM. All tests used a single-selection format with four options to choose from. Item difficulties for the content knowledge pretests in civil engineering, chemistry, and biology were fixed on the values obtained from previous test evaluation studies (Table 2.1) if the weighted mean square residuals (“infit”) with fixed difficulty were within acceptable limits (between 0.7 and 1.3; Bond & Fox, 2007). For the social science content knowledge test no reference values were available and therefore all items were estimated freely. Item parameter drift between pretest and posttest was tested by fixing posttest item difficulties to their pretest values and checking their infits. If a fixed item’s infit was acceptable (within the same limits as in the pretest) it was kept fixed; otherwise, its difficulty was estimated freely. Example items for each domain are shown in Figure 3.2.

Figure 3.1

Example Item of the Mental Rotation Test



Note. The participants were asked to select the two objects in the bottom row that were rotated versions of the object at the top.

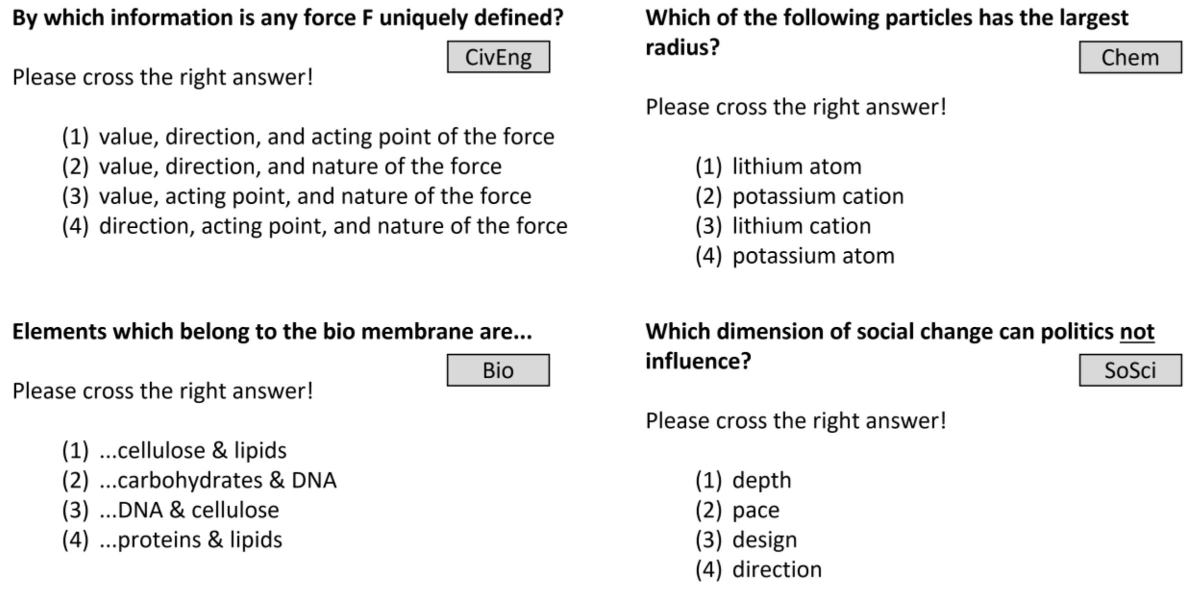
Table 3.1

Description of Domain-specific Content Knowledge Tests

Domain	Topic	Items	WLE reliability PoM 1/2	Reference
Civil engineering	Technical mechanics	36	.74/.77	Dammann (2016)
Chemistry	General chemistry	35	.75/.75	Averbeck (2021)
Biology	Botany, cell biology, zoology	36	.57/.58	Binder, Schmiemann, and Theyßen (2019); Binder, Sandmann, et al. (2019)
Social science	Sociology, political science	40	.74/.85	Walpuski et al. (2021)

Figure 3.2

Example Items of the Content Knowledge Tests for Each Domain



By which information is any force F uniquely defined?

CivEng

Please cross the right answer!

(1) value, direction, and acting point of the force
(2) value, direction, and nature of the force
(3) value, acting point, and nature of the force
(4) direction, acting point, and nature of the force

Which of the following particles has the largest radius?

Chem

Please cross the right answer!

(1) lithium atom
(2) potassium cation
(3) lithium cation
(4) potassium atom

Elements which belong to the bio membrane are...

Bio

Please cross the right answer!

(1) ...cellulose & lipids
(2) ...carbohydrates & DNA
(3) ...DNA & cellulose
(4) ...proteins & lipids

Which dimension of social change can politics <u>not</u> influence?

SoSci

Please cross the right answer!

(1) depth
(2) pace
(3) design
(4) direction

Note. Translation by the authors. CivEng = civil engineering, Chem = chemistry, Bio = biology, SoSci = social science

3.3.4.3 Reasoning

Reasoning ability is considered the closest proxy to general intelligence (Carroll, 1993), so we used reasoning ability tests to control for general intelligence. To measure reasoning ability, the object classification test from the nonverbal intelligence subscale of the *Kognitiver Fähigkeitstest* was used (Cognitive Abilities Test, KFT; Heller & Perleth, 2000; 15 items, WLE reliability = .73). The KFT is a battery of cognitive tests similar to the Cognitive Abilities Test of Lorge and Thorndike (1957). The subscale used required participants to choose one stimulus (out of five), correctly continuing a series of stimuli by detecting the underlying rule.

3.3.5 Analyses

In a first step, data were analyzed in a hierarchical multiple regression analysis to test the relevance of different predictors and to test for general domain differences in the predictors' effect on posttest content knowledge as the criterion. Pretest content knowledge, high school GPA, and reasoning ability, being control variables, were entered in this order in the model first. Spatial visualization ability, representing the focus of our study, was entered last. To test the effect of spatial visualization ability under strict conditions, the control variables were all kept in this model regardless of their significance. For all predictors, differences in effects between domains were tested by adding the interaction term of domain and the respective variable and then comparing the model's R^2 with and without the interaction. Interaction terms were kept in the model if they significantly improved the model's R^2.

In a second step, the final multiple regression model from Step 1 was converted into a multiple-group path model using the R-package lavaan (Rosseel, 2012, version 0.6-11) to be able to test the predictors' effects for significance within each domain. Path coefficients of predictors that had significant interactions with domain were estimated freely, while path coefficients for the other variables were fixed across domains. Parameter estimates were tested for statistical significance by bias-corrected bootstrapping using 10,000 bootstrap samples. Only complete cases ($n_{\text{civil engineering}} = 150$, $n_{\text{chemistry}} = 232$, $n_{\text{biology}} = 210$, $n_{\text{social science}} = 180$) were used in the regression models and the path models, as otherwise the sample would be different between different models of the hierarchical regression analysis, rendering model comparison tests impossible. All coefficients reported from Steps 1 and 2 are standardized βs as all variables were standardized within domains beforehand.

The third and last step was to conduct multiple-group structural equation modeling on the full sample using full information maximum likelihood estimation (FIML; Enders & Bandalos, 2001) to investigate indirect and total effects of spatial ability on posttest content knowledge via pretest content knowledge and high school GPA as mediators. If reasoning ability had previously emerged as a significant predictor of posttest content knowledge in the final multiple regression (Step 1) and path models (Step 2), it was included as a separate predictor in the structural equation model as well. For taking test reliability into account in the model, we used single indicators (Hayduk & Littvay, 2012) for the latent variables of all ability and knowledge tests (spatial visualization ability, pretest content knowledge, posttest content knowledge, and reasoning ability, if included) using their WLE abilities, WLE reliabilities, and WLE variance estimates. Parameter estimates were tested for statistical significance by bias-corrected bootstrapping using 10,000 bootstrap samples. Differences in path coefficients across the four academic domains were tested by imposing cross-group equality constraints on the path coefficients and comparing the constrained to the unconstrained model in terms of model fit: Constrained models were accepted if the χ^2 model comparison tests were not significant and the difference in comparative fit index (CFI) was smaller than .002 (Meade et al., 2008). Coefficients of the same paths were either fixed to be equal across all domains or were all free. Reported coefficients from Step 3 are unstandardized b's. These are close, but not identical to standardized βs as manifest variables (high school GPA and the indicators for the latent variables) were standardized within domains.

All analyses used $\alpha = .05$.

3.4 Results

3.4.1 Hierarchical Linear Regression

The hierarchical linear regression analysis, as the first step of analysis, tested the general relevance of our predictors pretest content knowledge, high-school GPA, reasoning ability, and spatial visualization ability (added to the model in this order) on posttest content knowledge. Also, the addition of interactions between predictors and domain to the model to indicate domain differences in effects was tested immediately after the predictor itself was added. In the resulting final model, pretest content knowledge was by far the strongest predictor of posttest content knowledge and accounted (interaction with domain included) for 23% of its variance beyond the other predictors (spatial visualization ability, high school GPA, and reasoning ability). High school GPA (interaction with domain included) accounted for 4% of posttest content knowledge variance beyond the other predictors, while reasoning ability accounted for less than 1% of the variance ($\Delta R^2 = 0.006$) beyond the other predictors. The addition of interactions with domain significantly improved the model for pretest content knowledge ($F(6, 764) = 2.33$, $p = .031$) and high school GPA ($F(3, 760) = 5.10$, $p = .002$), indicating differences in effects across domains, but not for reasoning ability ($F(3, 756) = 2.59$, $p = .052$). After adding all control variables and relevant interactions, results indicated that the addition of spatial visualization ability to the regression model improved its R^2 significantly ($\Delta R^2 = .04$, $F(1, 758) = 64.04$, $p < .001$), without interacting with domain ($F(3, 755) = 1.54$, $p = .357$), indicating a stable effect of spatial visualization ability across domains.

3.4.2 Multiple-Group Path Modeling

The multiple-group path model, as the second step of analysis, tested and compared direct effects of the predictors on posttest content knowledge across domains. Because, in the hierarchical regression analysis, prior content knowledge and high school GPA interacted with domain, the paths from prior content knowledge and high school GPA to posttest content knowledge were not fixed across domains in the multiple-group path model. Thus, domain-specific effects for prior content knowledge and high school GPA were tested while the effects of spatial visualization ability and reasoning ability were fixed across domains. The resulting path model fitted the data satisfactory ($\chi^2(6) = 12.13$, $p = .059$, CFI = 0.990, Tucker-Lewis-Index [TLI] = 0.972, root-mean-square error of approximation [RMSEA] = 0.073, standardized root-mean-square residual [SRMR] = 0.016) and is displayed in Figure 3.3.

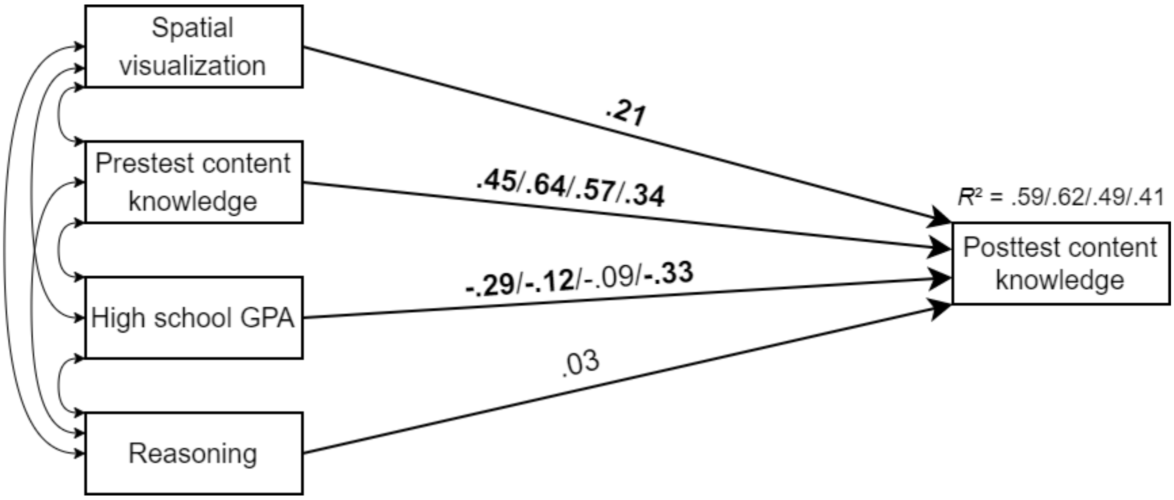
Prior content knowledge was the strongest predictor in all domains (Figure 3.3), even though absolute coefficients differed between domains. High school GPA was the second strongest predictor in engineering and social science, while in chemistry and biology the effect of high school GPA was trumped by spatial visualization ability. In contrast to previous studies (e.g., Burton & Ramist, 2001), high school GPA was not a significant predictor in all domains, as it was not significant in biology ($p = .101$). The effect of reasoning ability was not significant across domains after spatial visualization ability was added to the model ($p = .212$). Since reasoning ability had already not played a significant role in the hierarchical multiple regression model computed in the first step of data analysis ($\Delta R^2 = .006$, $F(1, 758) = 0.83$, $p = .362$), and the path could be set to 0 in all domains without significantly reducing model fit in the multiple-group path model ($\Delta\chi^2(1) = 1.456$, $p = .228$), reasoning ability was not considered in further analysis.

3.4.3 Multiple-Group Structural Equation Modeling

In the third and final step of analysis we estimated a constrained mediation model (Figure 3.4) that was based on the predictors and their domain differences identified in the first two steps. The intention was to compare not only direct effects on posttest content knowledge across domains but indirect effects as well. The model fitted the data well ($\chi^2(15) = 7.79$, $p = .931$; CFI = 1; TLI = 1.014; RMSEA = 0; SRMR = 0.019). As expected, the direct effects of the predictor (spatial visualization) and the two mediators (high school GPA and pretest content knowledge) on the criterion (posttest content knowledge) showed the same pattern regarding domain differences as in the first two steps of analysis: The path from spatial visualization ability to posttest content knowledge did not vary across domains and could be fixed without significantly reducing model fit ($\Delta\chi^2(3) = 5.85$, $p = .119$; ΔCFI = 0). Fixing the paths from pretest content knowledge ($\Delta\chi^2(3) = 27.29$, $p < .001$; ΔCFI = .019) and high school GPA ($\Delta\chi^2(3) = 11.80$ $p = .008$; ΔCFI = .001) to posttest content knowledge across domains significantly reduced model fit, so these were estimated freely. Paths specific to the mediation model from the predictor (spatial visualization ability) to the two mediators (pretest content knowledge and high school GPA) were tested for domain differences as well: The path from spatial visualization ability to pretest content knowledge could not be fixed between domains ($\Delta\chi^2(3) = 13.82$, $p = .003$; ΔCFI = .003) and was estimated freely, while the path from spatial visualization ability to high school GPA was did not vary across domains and could be fixed without significantly reducing model fit ($\Delta\chi^2(3) = 1.11$, $p = .775$; ΔCFI = 0). Because of the correlation of pretest content knowledge and high school GPA (which was stable across all domains; $\Delta\chi^2(3) = 0.83$, $p = .843$; ΔCFI = 0), there were four indirect paths of spatial visualization ability on posttest content knowledge (Wright, 1934). All of them were significant in at least one domain. Total indirect effects for the domains (all $p < .001$) were as follows: engineering .48 (*SE* = 09, 95% CI [.32, .66], chemistry .51 (*SE* = .07, 95% CI [.39, .65]), biology .36 (*SE* = .06, 95% CI [.24, 49]), and social science .19 (*SE* = .04, 95% CI [.10, .28]). Stronger effects in engineering and chemistry compared to social science (with biology in-between) supported the notion that spatial ability is especially important for STEM domains, even though the direct effect on content knowledge was the same across all domains.

Figure 3.3

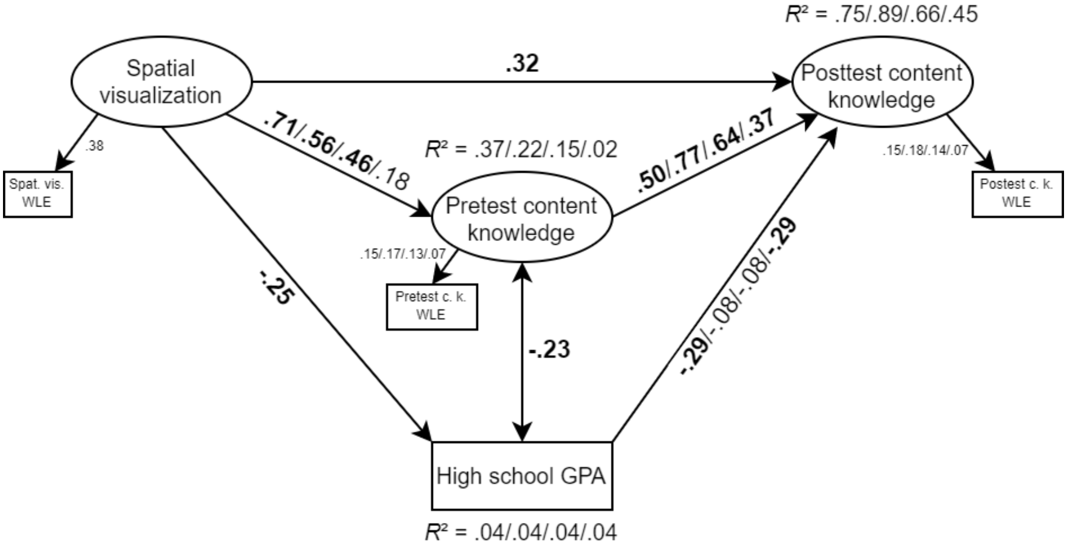
Multiple-group Path Model with Posttest Content Knowledge as Criterion



Note. Paths were fixed across domains if their interaction with domain did not significantly improve the hierarchical multiple regression model computed in the first step of data analysis. Coefficients are standardized βs. Bold coefficients are significant with α = .05, while non-bold coefficients are not significant. Varying coefficients are displayed in order of civil engineering/ chemistry/ biology/ social science.

Figure 3.4

Constrained Mediation Model of Spatial Visualization Ability on Posttest Content Knowledge



Note. Paths were fixed across domains if fixing did not significantly reduce the model's fit. Residual variances of all single indicators were fixed on (1 – WLE reliability) * WLE variance. Coefficients are unstandardized bs but comparable in size to standardized βs as all manifest variables (high school GPA and all indicators) were standardized within domains. Bold coefficients are significant with α = .05, while non-bold coefficients are not significant. Varying coefficients are displayed in order of civil engineering/ chemistry/ biology/ social science.

3.5 Discussion

We demonstrated that spatial visualization ability is predictive of content knowledge acquisition during students' first semester at university across different domains, including the STEM domains of civil engineering, chemistry, and biology, and the non-STEM domain of social science. The results are in line with earlier work on the role of spatial visualization ability in academic success in STEM fields (e.g., Sorby et al., 2018, for introductory course grades in engineering; Stieff et al., 2014, for chemistry content knowledge; Russell-Gebbett, 1985, for biology problem solving). We extended, however, this previous research by demonstrating and comparing the relevance of spatial visualization ability for the acquisition of content knowledge across multiple and heterogenous STEM domains, and even social science. Additionally, we even found this effect after controlling for several cognitive and knowledge variables which is a much stricter control than in previous studies on the effect of spatial ability on STEM success. In predicting posttest content knowledge (representing content knowledge acquired during the semester), spatial visualization ability outperformed reasoning ability and was on par with high school GPA, being trumped only by pretest content knowledge (representing prior content knowledge acquired mostly in high school). In fact, the effect of reasoning ability was no longer significant after spatial visualization ability was included in the model, so reasoning ability had no effect beyond the other predictors – most importantly, spatial visualization ability. Accordingly, spatial visualization ability accounted for most of the variance also explained by reasoning ability. This indicates that there is in fact an overlap between the reasoning test and the spatial visualization test that accounts for some variance in posttest content knowledge – most likely representing the general intelligence factor (e.g., Carroll, 1993; Nolte et al., 2022). But spatial visualization ability also accounted for additional posttest content knowledge variance beyond the variance shared with reasoning ability (as indicated by its significant partial regression coefficient) – we expected this to reflect the effect of the incremental spatial visualization ability factor included in the mental rotation test (Nolte et al., 2022). So, while general intelligence is certainly relevant for academic achievement in general, specific intelligence facets might be better suited to explaining achievement in different domains. In STEM, spatial ability (especially its sub-facet spatial visualization) appears to be one of these, while there may be others of relevance in other domains (e.g., verbal intelligence for language studies where many aspects of verbal intelligence might be relevant, such as a sensitivity for words, their direct and indirect meaning, and their sound, or the general ability to learn new languages; Gardner, 1991). However, the relevance of spatial ability for achievement in higher education might extend beyond STEM domains, as we found a substantial effect of spatial ability on posttest content knowledge in social science as well. Possible avenues for this relationship are discussed in the next chapter.

In contrast to previous studies (Hambrick et al., 2012; Uttal & Cohen, 2012) we did not find any interaction between expertise (reflected by pretest content knowledge) and spatial visualization ability that would indicate the decreasing importance of spatial abilities with increasing domain expertise. Potentially, this is due to our sample all being novices with varying degrees of novice-level content knowledge while, e.g., Hambrick et al. (2012) explicitly compared novices to experienced experts. Most likely, this interaction is more pronounced when great differences in domain knowledge are present, such as when comparing novices and experts, and it is harder to detect in samples with roughly the same level of expertise.

The mediation analysis showed that prior content knowledge was predicted by spatial visualization ability as well, resulting in a strong indirect effect on content knowledge. A similar but weaker result was found for high school GPA. Total indirect effects were quite substantial: As hypothesized, engineering and chemistry had by far the strongest total indirect effects of spatial ability (total $b_{\text{engineering}} = .48$; total $b_{\text{chemistry}} = .51$), followed by biology (total $b_{\text{biology}} = .36$), and social science (total $b_{\text{social science}} = .19$) trailing far behind. Interestingly, this hierarchy only showed through the indirect effects, most notably in the paths through prior content knowledge: The direct effect on posttest content knowledge was the same at *b* = .32 across domains. This indicates that spatial ability had a general effect on learning in higher education that is not domain-specific, and effects appear to be more diverse for the acquisition of prior content knowledge in (high) school.

3.5.1 Implications for Theory and Practice

Our result that only indirect effects of spatial visualization ability on posttest content knowledge varied between domains while the direct effect was the same across all four domains indicates that coursework in higher education in general might require spatial ability, regardless of domain, while in previous stages of education the effects vary more between domains (as indicated by the large differences in indirect effects of spatial ability on posttest content knowledge through pretest content knowledge). It is possible that mathematics may play a role in this: Effects of spatial ability have been shown most frequently for mathematics, compared to other domains (Bruce & Hawes, 2015; Cheng & Mix, 2014; Judd & Klingberg, 2021; Lowrie et al., 2017; Lowrie et al., 2019; Mix et al., 2016; Mix et al., 2021; Rodán et al., 2016; Satterly, 1976; Simms et al., 2016; Sorby et al., 2013; Wolfgang et al., 2003; for a review, see Mix & Cheng, 2012). All domains in the present study included math classes (including statistics). But mathematical skills do not seem to have the same importance in the respective “predecessor school subjects”: While math is very important in physics (which contains introductions to classical mechanics, which was the topic of our civil engineering content knowledge test) and is also frequently used in chemistry, it is less important in high school biology and basically nonexistent in high school social science classes in Germany. It is therefore possible that mathematics is the key skill that moderates the strength of spatial ability’s effect on achievement in STEM versus non-STEM domains. Further studies therefore could look to include mathematical skills as a possible mediator between spatial ability and content knowledge in different domains to evaluate this. Other skills possibly connecting spatial ability and content knowledge could be visual model comprehension skills (sometimes also called “representational competence skills”; Stieff et al., 2018), defined as skills to “handle and learn with visualizations” (Dickmann et al., 2019): The use of visual representations supports learning, especially in STEM domains (Rau, 2017), and making better use of visual representations is likely to be a critical skill for students. It’s possible that higher education uses more visual representations for complex topics even in non-STEM domains (such as social science in the present study, which, e.g., uses complex organizational diagrams) – this could also be an explanation for the stability of the effect of spatial ability on posttest content knowledge acquired in university compared to the effect varying across domains for pretest content knowledge acquired in high school.

Earlier studies showed that STEM exam grades (e.g., Sorby et al., 2018; for a review, see Uttal, Miller, & Newcombe, 2013) can be improved through spatial training. Our results demonstrated that general spatial abilities predict the acquisition of content knowledge (the central prerequisite for success in exams) in different domains. This could spark interest in developing spatial ability trainings in more domains than before (mainly math, e.g., Cheng & Mix, 2014, and engineering, e.g., Sorby et al., 2018). Such trainings do not necessarily need to be domain-specific, as in our study the same spatial ability test with content-neutral figures predicted the acquisition of content knowledge across all domains studied. Domain-specific trainings would probably result in stronger specific effects (as suggested in, e.g., Uttal, Miller, & Newcombe, 2013; for an example of a spatial training integrated in domain-specific content, see Bille et al., 2020), but a general spatial ability training with topic-independent tests (Maquet et al., 2022) could work on a broader field, which would make it easier to use in education, and earlier – as shown by, e.g., Cheng and Mix's (2014) mental rotation training to improve children’s mathematical ability. Such training could easily be introduced in schools to facilitate performance in multiple subjects, as it requires relatively little preparation and might positively affect achievement in a broad spectrum of subjects: Our results suggest that spatial visualization ability predicts the acquisition of content knowledge in both STEM domains and social science. Considering also the results of Satterly (1976), who found a connection between spatial ability and English comprehension, spatial ability might be relevant for quite a wide array of subjects in school. This would be an *explicit spatial training intervention*, defined as an intervention that consists of explicitly spatial tasks without relation to content (also called “direct practice”; Maquet et al., 2022). But, if schools cannot make room in their curricula for an ability training that is not directly tied to a subject, “spatializing” current curricula could also help in the form of *implicit spatial training intervention*, defined as training that is embedded organically in normal activities related to the topic of the lesson (Maquet et al., 2022).

There are two approaches to involving spatial ability in school curricula: On the one hand, teachers could try to include spatially demanding teaching materials in their classes, to train spatial ability within the context of the subject. On the other hand, teachers could adapt their materials in such a way as to reduce spatial ability demands by providing targeted assistance (see, e.g., the five principles for designing chemistry visualizations described in Wu & Shah, 2004). These adapted materials could, for example, include built models that students can physically explore and manipulate manually (e.g., ball-and-stick models for molecule structures in chemistry) instead of having to construct and manipulate a mental model from a description or printed depiction. This second approach is already integrated in some subjects – for example as part of mathematics education standards in primary school geometry lessons (The National Council of Teachers of Mathematics [NCTM], 2000). But there is still great potential for this approach in others, such as chemistry, where German teachers have traditionally followed the “historical way” (Barke & Engida, 2001, p. 228) of teaching chemical concepts of phenomena, molecular formulas and structural formulas in a stepwise fashion as separate concepts, instead of integrating them with each other to support holistic learning of chemical concepts. In practice, both approaches would most likely look quite similar; mostly they would differ in their demands on spatial ability. Combining the approaches sequentially by beginning with a high level of visual assistance and then slowly reducing it (e.g., using less-detailed built models with additional descriptions instead of physical details) could prove effective. This strategy could also help with the representation dilemma described by Rau (2017) that students' on the one hand need content knowledge to understand visual representations and on the other hand need the visual representations to learn the content: Easing the way into working with visual representations and developing representational competencies early could help with more complex representations in later stages of education. General benefits of and issues with using explicit or implicit spatial training interventions in school contexts are discussed in Maquet et al.'s (2022) review of spatial ability interventions in secondary level STEM education.

Modern technology also allows for new avenues in both implicit and explicit spatial training. For example, Roca-González et al. (2017) used an augmented reality environment and Molina-Carmona et al. (2019) used a virtual reality environment to train students' spatial abilities. Video games also could be used for (usually implicit) training, as these are known to train a variety of abilities helpful for education (Barr, 2017; Granic et al., 2014; for a review, see Aguilera & Mendiz, 2003 and Quiroga & Colom, 2019) including spatial ability (Murias et al., 2016; Shute et al., 2015; Uttal, Meadow, et al., 2013). Many studies have used commercial video games that are aimed at entertainment – because of this, such video game-based trainings could lead to higher engagement and motivation than would more "formal" interventions, especially in students usually not open to additional coursework.

Playing such video games and thereby improving their spatial abilities might ease a student's way into later STEM studies by facilitating the acquisition of prior content knowledge that is beneficial for university entry and sets the cornerstones for study success. Our results show that such effects are not exclusive to any academic domain but even extend beyond STEM domains that are already known for their connection to spatial ability. Mathematical skills and visual model comprehension skills are potential mediators for the effect of spatial ability on content knowledge acquisition in higher education, as they are strongly related to spatial ability and required for learning across both STEM and many non-STEM domains.

3.5.2 Strengths, Limitations, and Directions for Future Research

A main strength of the present study is the large and diverse sample including cohorts from different domains and different universities within each domain. This allows for stronger conclusions on the generalizability of the results, in comparison to the more usual studies focusing on a single cohort in which it is harder to rule out cohort-specific effects such as curriculum specifics. However, some cohorts in the present study are smaller than others and may by themselves be less representative, which is why we pooled cohorts from the same domain. Collecting data in regular courses proved the most effective way of recruitment in the present study but was not possible in all cohorts. Further studies should place strong emphasis on collecting data in regular courses to ensure samples are as complete and representative as possible for their respective cohorts.

Spatial ability was assessed at the same time as pretest content knowledge (representing prior knowledge), after participants had already completed their high school education. This is a limitation for the assumed direction of the causation of effects. As spatial abilities are malleable by training, life experience, or educational interventions (Uttal, Miller, & Newcombe, 2013) it cannot be ruled out that the assessment of spatial ability at the beginning of the participants' first semester did not accurately reflect their spatial ability levels during high school (which were relevant for acquisition of prior knowledge and for high school GPA). However, improvements of spatial ability still require an adequate intervention. Even though educational outcomes such as content knowledge acquisition are related to spatial ability, usual education processes do not necessarily improve spatial ability: In the present study, participants' spatial ability did not improve from pretest to posttest, even though spatial ability was strongly related to the content knowledge acquired during that period. We have no reason to expect that the high school curriculum might have an effect on spatial ability when university curricula do not. On the same note, we do not expect other activities that might affect spatial ability (such as spatially demanding hobbies, e.g., video games) to have changed between later high school years and the first semester of university. Taken together, this leads us to the conclusion that our measurement of spatial ability is valid to predict both pretest content knowledge and high school GPA even though it is measured at the same time. Nevertheless, a longitudinal study would be necessary to answer the question of how spatial ability develops during high school and university studies to ensure the validity of delayed testing of spatial ability such as in our study.

3.5.3 Conclusion

We demonstrated and compared the incremental value of spatial visualization ability for predicting academic achievement in higher education, beyond reasoning ability, high school GPA, and prior knowledge, in three STEM domains and one non-STEM domain. The total effect was stronger in the STEM domains, but the effect was nevertheless present in the non-STEM domain as well, suggesting the relevance of spatial abilities for STEM domains in particular but also for higher education in general, without regard to domain. Possible avenues of explanation might be mathematical skills which are important in many domains (sometimes in the form of statistics, e.g., in social science), or visual model comprehension skills (Rau, 2017), as both STEM and non-STEM domains in higher education work with complex visualizations (such as molecule models in chemistry or complex organizational diagrams in social science): Both skills are closely related to spatial ability. Accordingly, spatial visualization tests might prove useful for talent searches (e.g., allocation of study places or awarding scholarships; Lubinski, 2010) in many study programs, especially in STEM domains. Spatial visualization tests could also be used to identify at-risk students with deficits in spatial ability in those domains and support them with interventions. These interventions could be domain-specific and could combine spatial training directly with content knowledge relevant to the field (for an example of an intervention in engineering, see Sorby et al., 2018). But interventions could also be unspecific with regard to the domain (for example, as part of a cross-faculty preparation course), as we found effects of spatial visualization ability on content knowledge using a spatial ability test with content-neutral stimuli; previous studies have shown the effectiveness of such direct trainings (e.g., for mathematical ability, Cheng & Mix, 2014).

3.6 References

Aguilera, M. de, & Mendiz, A. (2003). Video games and education: Education in the face of a “parallel school”. *Computers in Entertainment*, *1*(1), 1–10. https://doi.org/10.1145/950566.950583

Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016, June 21). *Predicting student dropout in higher education*. #Data4Good: Machine Learning in Social Good Applications, New York. http://arxiv.org/pdf/1606.06364v4

Averbeck, D. (2021). *Zum Studienerfolg in der Studieneingangsphase des Chemiestudiums: Der Einfluss kognitiver und affektiv-motivationaler Variablen*. Dissertation [Regarding study success in the study entry phase of chemistry study programs: The role of cognitive and affective-motivational variables]. *Studien zum Physik- und Chemielernen: Vol. 308*. Logos.

Barke, H.-D., & Engida, T. (2001). Structural chemistry and spatial ability in different cultures. *Chemistry Education Research and Practice*, *2*(3), 227–239. https://doi.org/10.1039/B1RP90025K

Barr, M. (2017). Video games can develop graduate skills in higher education students: A randomised trial. *Computers & Education*, *113*, 86–97. https://doi.org/10.1016/j.compedu.2017.05.016

Bille, V., Rumann, S., Roelle, J., Opfermann, M., & Schmuck, C. (2020). Förderung des ikonischen Modellverständnisses in Chemiestudiengängen [Support of iconic model comprehension in chemistry study programs]. In S. Habig (Chair), *Jahrestagung.* Symposium conducted at the meeting of Gesellschaft für Didaktik der Chemie und Physik (GDCP), Vienna.

Binder, T., Sandmann, A., Sures, B., Friege, G., Theyßen, H., & Schmiemann, P. (2019). Assessing prior knowledge types as predictors of academic achievement in the introductory phase of biology and physics study programmes using logistic regression. *International Journal of STEM Education*, *6*(1), 162. https://doi.org/10.1186/s40594-019-0189-9

Binder, T., Schmiemann, P., & Theyßen, H. (2019). Knowledge acquisition of biology and physics university students—the role of prior knowledge. *Education Sciences*, *9*(4), 281. https://doi.org/10.3390/educsci9040281

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Erlbaum.

Bruce, C. D., & Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: What is it? Why does it matter? And what can we do about it? *ZDM Mathematics Education*, *47*(3), 331–343. https://doi.org/10.1007/s11858-014-0637-4

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980*. Research Report No. 2001-2. College Entrance Examination Board.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Chen, X. (2009). Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education. Stats in Brief. NCES 2009-161. *National Center for Education Statistics*.

Cheng, Y.-L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*, 2–11. https://doi.org/10.1080/15248372.2012.725186

Dammann, E. (2016). *Entwicklung eines Testinstruments zur Messung fachlicher Kompetenzen in der Technischen Mechanik bei Studierenden ingenieurwissenschaftlicher Studiengänge* [Development of a test measuring engineering students' expertise in technical mechanics] [Dissertation]. Universität Stuttgart. https://doi.org/10.18419/opus-9073

Delialioğlu, Ö., & Aşkar, P. (1999). Contribution of students' mathematical skills and spatial ability of achievement in secondary school physics. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *16*(16), 34–39.

Dickmann, T., Opfermann, M., Dammann, E., Lang, M., & Rumann, S. (2019). What you see is what you learn? The role of visual model comprehension for academic success in chemistry. *Chemistry Education Research and Practice*, *20*(4), 804–820. https://doi.org/10.1039/C9RP00016J

Duesbury, R. T., & O'Neil, H. F. (1996). Effect of type of practice in a computer-aided design environment in visualizing three-dimensional objects from two-dimensional orthographic projections. *Journal of Applied Psychology*, *81*, 249–260. https://doi.org/10.1037/0021-9010.81.3.249

Enders, C., & Bandalos, D. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. https://doi.org/10.1207/s15328007sem0803_5

Fleischer, J., Leutner, D., Brand, M., Fischer, H. E., Lang, M., Schmiemann, P., & Sumfleth, E. (2019). Vorhersage des Studienabbruchs in naturwissenschaftlich-technischen Studiengängen [Prediction of student drop-out in STEM study programs]. *Zeitschrift für Erziehungswissenschaft*, *22*(5), 1077–1097. https://doi.org/10.1007/s11618-019-00909-w

Formazin, M., Schroeders, U., Köller, O., Wilhelm, O., & Westmeyer, H. (2011). *Studierendenauswahl im Fach Psychologie* (Vol. 62). https://doi.org/10.1026/0033-3042/a000093

Freyer, K., Epple, M., Brand, M., Schiebener, J., & Sumfleth, E. (2014). Studienerfolgsprognose bei Erstsemesterstudierenden in Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, *20*(1), 129–142. https://doi.org/10.1007/s40573-014-0015-3

Ganley, C. M., Vasilyeva, M., & Dulaney, A. (2014). Spatial ability mediates the gender difference in middle school students' science performance. *Child Development*, *85*(4), 1419–1432. https://doi.org/10.1111/cdev.12230

Gardner, H. (1991). *Abschied vom IQ: Die Rahmen-Theorie der vielfachen Intelligenzen* [Frames of mind: The theory of multiple intelligences]. Klett-Cotta.

Granic, I., Lobel, A., & Engels, R. C. M. E. (2014). The benefits of playing video games. *The American Psychologist*, *69*(1), 66–78. https://doi.org/10.1037/a0034857

Hailikari, T. (2009). *Assessing university students' prior knowledge: Implications for theory and practice*. Helsinki University Print. https://helda.helsinki.fi/handle/10138/19841

Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., Turner, S. P., Rench, T. A., & LaDue, N. D. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General*, *141*, 397–403.

Hambrick, D. Z., & Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science*, *20*(5), 275–279. https://doi.org/10.1177/0963721411422061

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*(1), 159. https://doi.org/10.1186/1471-2288-12-159

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, *19*, 61–70. https://doi.org/10.1016/j.lindif.2008.04.006

Heller, K. A., & Perleth, C. (2000). *KFT 4-12+ R - Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision* [KFT 4-12+ R - Cognitive ability test for 4th to 12th grades, revision]. Beltz.

Heublein, U., Richter, J., & Schmelzer, R. (2020). *Die Entwicklung der Studienabbruchquoten in Deutschland.* https://doi.org/10.34878/2020.03.DZHW_BRIEF

Heublein, U., & Schmelzer, R. (2018). Die Entwicklung der Studienabbruchquoten an den deutschen Hochschulen: Berechnungen auf Basis des Absolventenjahrgangs 2016 [The development of drop-out from higher education at German colleges: Calculations based on 2016 graduate class]. *DZHW-Projektbericht*.

Hsi, S., Linn, M. C., & Bell, J. E. (1997). The role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education*, *86*, 151–158.

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology*, *78*, 250–261. https://doi.org/10.1037//0021-9010.78.2.250

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (1 [Dr.]. *Cognitive science series: Vol. 6*. Havard University Press.

Judd, N., & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour.* Advance online publication. https://doi.org/10.1038/s41562-021-01118-4

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *The American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31*, 549–579. https://doi.org/10.1080/15326900701399897

Krüger, J. K., & Suchan, B. (2016). You should be the specialist! Weak mental rotation performance in aviation security screeners – reduced performance level in aviation security with no gender effect. *Frontiers in Psychology*, *7*, 333. https://doi.org/10.3389/fpsyg.2016.00333

Kyllonen, P. C., & Christal, R. E. (1990). *Reasoning ability is (little more than) working-memory capacity?!* (Vol. 14). https://doi.org/10.1016/S0160-2896(05)80012-1

LimeSurvey GmbH. (2019). *LimeSurvey: An Open Source survey tool* (Version 3.20.2+191119) [Computer software]. Limesurvey GmbH. Hamburg, Germany. http://www.limesurvey.org

Lohman, D. F. (1996). Spatial ability and G. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 97–116). Lawrence Erlbaum Associates, Inc.

Lorge, I., & Thorndike, R. L. (1957). *The Lorge-Thorndike intelligence tests: Technical manual*. Houghton-Mifflin.

Lowrie, T., Logan, T., & Hegarty, M. (2019). The influence of spatial visualization training on students' spatial reasoning and mathematics performance. *Journal of Cognition and Development*, *20*(5), 729–751. https://doi.org/10.1080/15248372.2019.1653298

Lowrie, T., Logan, T., & Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *The British Journal of Educational Psychology*, *87*, 170–186. https://doi.org/10.1111/bjep.12142

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*, 344–351. https://doi.org/10.1016/j.paid.2010.03.022

Maquet, L., Dunbar, R., Buckley, J., Seery, N., & Sorby, S. A. (2022). A review of the literature to inform the efficacy of a spatial skills intervention for secondary level STEM education. In D. Gill, J. Tuff, T. Kennedy, S. Pendergast, & S. Jamil (Eds.), *PATT39 Conference 2022. PATT on the edge: Technology, innovation and education* (pp. 496–507).

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of Applied Psychology*, *93*, 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences*, *26*, 141–152. https://doi.org/10.1016/j.lindif.2012.03.012

Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: Developmental and Educational Implications. In J. B. Benson (Ed.), *Advances in child development and behavior* (Vol. 42, pp. 197–243). Elsevier. https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Mix, K. S., Levine, S. C., Cheng, Y.-L., Stockton, J. D., & Bower, C. (2021). Effects of spatial training on mathematics in first and sixth grade children. *Journal of Educational Psychology*, *113*(2), 304–314. https://doi.org/10.1037/edu0000494

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology: General*, *145*, 1206–1227. https://doi.org/10.1037/xge0000182

Molina-Carmona, R., Pertegal-Felices, M. L., Jimeno-Morenilla, A., & Mora-Mora, H. (2019). Assessing the Impact of Virtual Reality on Engineering Students' Spatial Ability. In A. Visvizi, M. D. Lytras, & L. Daniela (Eds.), *Emerald studies in higher education innovation and technology. The future of innovation and technology in education: Policies and practices for teaching and learning excellence* (pp. 171–185). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78756-555-520181013

Murias, K., Kwok, K., Castillejo, A. G., Liu, I., & Iaria, G. (2016). The effects of video game use on performance in a virtual navigation task. *Computers in Human Behavior*, *58*, 398–406. https://doi.org/10.1016/j.chb.2016.01.020

National Academies Committee on Science, Engineering, and Public Policy. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. National Academies Press.

The National Council of Teachers of Mathematics. (2000). *Standards for school mathematics*. NCTM.

Neisser, U., Bodoo, G., Bouchard, T. J., JR., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., Urbina, S., & Boodoo, G. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Nolte, N., Schmitz, F., Fleischer, J., Bungart, M., & Leutner, D. (2022). Rotational complexity in mental rotation tests: Cognitive processes in tasks requiring mental rotation around cardinal and skewed rotation axes. *Intelligence*, *91*, 101626. https://doi.org/10.1016/j.intell.2022.101626

Pallrand, G. J., & Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching*, *21*, 507–516. https://doi.org/10.1002/tea.3660210508

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260–264. https://doi.org/10.1016/j.bandc.2007.09.003

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*(6), 901–910. https://doi.org/10.1016/j.econedurev.2010.07.009

Quiroga, M. Á., & Colom, R. (2019). Intelligence and video games. In R. J. Sternberg (Ed.), *The Cambridge handbook of intelligence* (pp. 626–656). Cambridge University Press. https://doi.org/10.1017/9781108770422.027

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3

Robitzsch, A., Kiefer, T., & Wu, M. (2022). *TAM: Test Analysis Modules* (Version 4.0-16) [Computer software]. https://CRAN.R-project.org/package=TAM

Roca-González, C., Martín-Gutiérrez, J., GarcÍa-Dominguez, M., & Del Mato Carrodeguas, M. C. (2017). Virtual Technologies to Develop Visual-Spatial Ability in Engineering Students. *Eurasia Journal of Mathematics, Science and Technology Education*, *13*(2), 441–468. https://doi.org/10.12973/eurasia.2017.00625a

Rochford, K. (1985). Spatial learning disabilities and underachievement among university anatomy students. *Medical Education*, *19*(1), 13–26. https://doi.org/10.1111/j.1365-2923.1985.tb01134.x

Rodán, A., Contreras, M. J., Elosúa, M. R., & Gimeno, P. (2016). Experimental but not sex differences of a mental rotation training program on adolescents. *Frontiers in Psychology*, *7*, 1050. https://doi.org/10.3389/fpsyg.2016.01050

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

RStudio Team. (2023). *RStudio. Integrated Development Environment for R* (Version 2023.03.0.386) [Computer software]. RStudio, PBC. Boston, MA. http://www.rstudio.com/

Russell-Gebbett, J. (1985). Skills and strategies – pupils' approaches to three-dimensional problems in biology. *Journal of Biological Education*, *19*(4), 293–298. https://doi.org/10.1080/00219266.1985.9654755

Satterly, D. J. (1976). Cognitive styles, spatial ability, and school achievement. *Journal of Educational Psychology*, *68*(1), 36–42. https://doi.org/10.1037/0022-0663.68.1.36

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604–614. https://doi.org/10.1037//0022-0663.93.3.604

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, *80*, 58–67. https://doi.org/10.1016/j.compedu.2014.08.013

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuomotor integration and visuospatial skills. *Journal of Experimental Child Psychology*, *145*, 22–33. https://doi.org/10.1016/j.jecp.2015.12.004

Sorby, S. A. (2009). Developing spatial cognitive skills among middle school students. *Cognitive Processing*, *10*, 312-315. https://doi.org/10.1007/s10339-009-0310-y

Sorby, S. A., & Baartmans, B. J. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year eengineering students. *Journal of Engineering Education*, *3*(89), 301–307.

Sorby, S. A., Casey, M. B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, *26*, 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sorby, S. A., & Veurink, N. (2010). Long-term results from spatial skills intervention among first-year engineering students. In American Society for Engineering Education (Chair), *Proceedings of the 65th midyear meeting of the engineering design graphics division of ASEE*.

Sorby, S. A., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is ‘yes’. *Learning and Individual Differences*, *67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Spearman, C. E. (1904). “General intelligence” - objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spearman, C. E., & Wynn Jones, L. (1950). *Human ability: A continuation of “The abilities of man”*. Macmillan.

Stieff, M., Dixon, B. L., Ryu, M., Kumi, B. C., & Hegarty, M. (2014). Strategy training eliminates sex differences in spatial problem solving in a stem domain. *Journal of Educational Psychology*, *106*(2), 390–402. https://doi.org/10.1037/a0034823

Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D., & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology*, *110*(8), 1160–1174. https://doi.org/10.1037/edu0000258

Talley, L. H. (1973). The use of three-dimensional visualization as a moderator in the higher cognitive learning of concepts in college level chemistry. *Journal of Research in Science Teaching*, *10*(3), 263–269. https://doi.org/10.1002/tea.3660100311

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 147–181). Elsevier. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking. *Current Directions in Psychological Science*, *22*(5), 367–373. https://doi.org/10.1177/0963721413484756

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835. https://doi.org/10.1037/a0016127

Walpuski, M., Fischer, V., Lang, M., Leutner, D., Manzel, S., & Sumfleth, E. (2021). Chemie, Sozialwissenschaften und Ingenieurwissenschaften: Studienerfolg und Studienabbruch [Chemistry, social science, and engineering: Study success and dropout]. In M. Neugebauer, H.-D. Daniel, & A. Wolter (Eds.), *Studienerfolg und Studienabbruch* (pp. 151–175). Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-32892-4_7

Walpuski, M., Kauertz, A., Kampa, N., Fischer, H. E., Mayer, J., Sumfleth, E., & Wellnitz, N. (2010). ESNaS – Evaluation der Standards für die Naturwissenschaften in der Sekundarstufe I [ESNaS – Evaluation of standards for the natural sciences in secondary level I]. In A. Gehrmann, U. Hericks, & M. Lüders (Eds.), *Bildungsstandards und Kompetenzmodelle: Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht* (pp. 171–184). Klinkhardt.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450. https://doi.org/10.1007/BF02294627

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420. https://doi.org/10.1037/0022-0663.99.2.397

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wolfgang, C., Stannard, L., & Jones, I. (2003). Advanced constructional play with LEGOs among preschoolers as a predictor of later school achievement in mathematics. *Early Child Development and Care*, *173*(5), 467–475. https://doi.org/10.1080/0300443032000088212

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*(3), 161–215.

Wu, H.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492. https://doi.org/10.1002/sce.10126

York, T. T., Gibson, C., & Rankin, S. (2015). *Defining and measuring academic success.* https://doi.org/10.7275/HZ5X-TX03

4 Studie 3: Evaluating a Video Game Training for Spatial Ability and Transfer in Engineering Studies

4.1 Abstract

The usefulness of spatial ability training for performance (e.g., in exams) has been shown in several STEM disciplines. However, only two of the three types of training has been evaluated (as classified by Uttal et al., 2013) – spatial task training and course training. The goal of the present study was to evaluate the potential of the third type, video game training, for content knowledge acquisition and exam grades in 68 first-year engineer students. Participants played either the 3D puzzle game “Portal 2” or the adventure game “Sherlock Holmes vs. Jack the Ripper” for 8h.

In total, playing a video game led to improved spatial abilities, with no significant interaction with group even though the within-group training effect was only significant in the Portal 2 group. Training gains did not significantly drop until the end of the semester two months later. Prior video game experience might affect the effectiveness of training with complex video games and should be considered in future studies. Posttest spatial ability predicted acquired content knowledge at the end of the semester at exam grades beyond reasoning ability and prior knowledge. Mediation analysis showed that the effect of spatial ability on exam grades was fully mediated by acquired content knowledge.

In conclusion, the present study first showed the potential of using video games as a readily available training tool of spatial ability to improve content knowledge acquisition and exam grades in engineering students. However, further research is needed to identify which type of games serve this purpose best.

Nolte, N., Rahimi, S., Shute, V. J., & Leutner, D. (2023). Ev*aluating a video game training for spatial ability and transfer in engineering studies* [manuscript in preparation]. Faculty of Educational Sciences, University of Duisburg-Essen.

4.2 Introduction

Uttal et al.'s (2013) meta-analysis convincingly showed the malleability of spatial ability through a number of different types of training, namely spatial task training (e.g., Judd & Klingberg, 2021), course training (e.g., Sorby et al., 2018), and video game training (e.g., Achtman et al., 2008). The authors further described the potential of such interventions for STEM education. In the past 20 years, more literature has emerged that highlighted the impact of spatial ability on success in STEM fields (for a review, see Wai et al., 2009; for a comparison between disciplines, see Nolte et al., 2023).

Currently, STEM programs suffer from high dropout rates, and have so for years (e.g., Chen, 2009; Heublein et al., 2020; Price, 2010). This constitutes a major risk factor for a skilled worker shortage in STEM fields (CBI, 2011), which has been called out as a threat to economic growth and innovation (National Academies Committee on Science, Engineering, and Public Policy [COSEPUP], 2007) and costs businesses billions each year (CBI, 2019). Preventing this shortage is a challenge for the education system, and spatial ability training applied in viable ways could provide a piece for this puzzle. As a thought experiment, Uttal et al. (2013) proclaimed that spatial ability training employed at the population level with a conservatively estimated effect size of Hedge's $g = 0.40$ could *double* the number of people with spatial abilities equal or higher to those displayed by current engineers. This is of course not a realistic prospect but nevertheless demonstrates the potential that lies in spatial ability training applied in viable places.

4.2.1 Spatial Ability and STEM

Wai et al. (2009) collected evidence from decades of research on the connection between spatial ability and STEM-related outcomes. They demonstrated that this relationship ranges from subject preferences in high school, to college majors, and up to occupations in later years. People with higher levels of spatial ability more frequently chose STEM subjects, majors, and occupations, with higher academic degrees also displaying higher spatial ability across the board (Lubinski, 2010). These findings highlight the importance of spatial ability across different stages of education.

Several studies investigated this connection in different STEM disciplines, such as engineering (e.g., Duesbury & O'Neil, 1996; Sorby et al., 2018), chemistry (e.g., Talley, 1973; Wu & Shah, 2004), mathematics (e.g., Hegarty & Kozhevnikov, 1999; Mix & Cheng, 2012), physics (e.g., Kozhevnikov et al., 2007; Pallrand & Seeber, 1984), or biology (e.g., Rochford, 1985; Russell-Gebbett, 1985). Nolte et al. (under review) conducted a comparison of the effect of spatial ability on success in engineering, chemistry, biology, and social science for first-year university students. They found a consistent direct effect on university students' content knowledge acquisition, suggesting that spatial ability affects learning across disciplines. However, they also found indirect effects through prior knowledge and high school GPA that varied across disciplines, with the highest indirect effects for engineering and chemistry, a weaker effect for biology and the weakest (but still significant) for social sciences, suggesting that in earlier stages of education effects of spatial ability on learning appear to be STEM specific.

4.2.2 Training of Spatial Abilities for STEM

Uttal et al. (2013) demonstrated the possibility to train spatial abilities using different kinds of interventions. Many studies have applied this work and the established connection between spatial ability and STEM outcomes. The common goal was to facilitate learning and improve various STEM outcomes such as specific test performance, exam grades, or study program selection and retention by applying spatial ability training (for a review, see Maquet et al., 2022). Such interventions are most promising when applied early in learning. For instance, in the review by Uttal and Cohen (2012), the authors concluded that spatial ability is more important to those with lower knowledge in a STEM area. That is, experts can use their acquired knowledge and established mental representations to reduce potential negative effects of low spatial ability, while novices require spatial ability to understand course material and construct those mental representations.

Training for spatial ability in STEM contexts so far appeared in two of the three types described in Uttal et al. (2013). The first form includes conditions where participants practice with common spatial tests such as mental rotation tests or spatial reasoning tests (e.g., Lowrie et al., 2019; Mix et al., 2021), while the second form of training involves participants enrolling in a course to engage in substantial spatial thinking, often related to the studied discipline (e.g., Molina-Carmona et al., 2019; Sorby et al., 2018). But the third type of spatial ability training, discussed in Uttal et al.'s (2013) meta-analysis, appears to be underutilized in research so far: playing video games – even though video game training resulted in the largest average effect size in the meta-analysis (Hedge’s $g = 0.54$, compared to $g = 0.48$ for spatial task training and $g = 0.41$ for course training).

Video games have also been shown to be able to train different cognitive abilities (for a metaanalysis, see Bediou et al., 2023). Earlier studies have focused on action video games and their effects on (visual) attention and (visual) working memory (e.g., Boot et al., 2008; Green & Bavelier, 2003, 2006) – action video games require a multitude of different skills for successful play, which they in turn train while playing (for a review, see Spence & Feng, 2010). On the one hand, this is a compelling argument for using video games to train multiple abilities at once. But on the other hand, this imposes difficulties for research purposes if a single set of abilities is the focus. More recent studies have branched out from using the action video game genre. One example of such a training study was reported by Shute et al. (2015) who used Portal 2 (Valve, 2011) as the training game for different aspects of spatial ability and problem solving. This game has an advantage over action video games commonly used previously as its gameplay is almost purely spatial and problem solving-related. This helped to ensure that these particular abilities were practiced and no complementary effects were transpiring from things such as improved attentional capabilities commonly occurring after action video game play. Although training multiple abilities at once seems desirable, it occludes the mechanisms affecting secondary variables the training should transfer to (e.g., STEM outcomes).

Additionally, using video games as training has several practical advantages in both educational and research contexts. For instance, it can be expected that getting teachers to buy into new methods of teaching is difficult, especially if it requires additional work for them in terms of preparation of materials and acquisition of necessary skills. Furthermore, it can be expected that it’s not an easy task to motivate students for additional extracurricular activities. Video games as training tools circumvent both problems, as they are available and ready-to-use without much preparation by the teachers, and they provide an activity that many students participate in as leisure activity anyway (around 60% of North Americans and around 65% of Europeans played digital games in the past six months; Newzoo, 2022). Various theories offer explanations for the intrinsic motivation for playing video games. One is Self-Determination Theory (SDT), as described in Rahimi et al. (2021). Achieving high levels of competence and autonomy while simultaneously being able to develop social bonds over games have been described as mechanisms for motivation in video games based on SDT (Przybylski et al., 2010). The participants’ higher intrinsic motivation for participation in the intervention is a major advantage compared to other types of interventions to reduce drop-out due to disinterest and to improve compliance, which are common problems in longitudinal studies.

4.3 The Present Research: Aims and Hypotheses

The present study aimed to close the gap in terms of the types of interventions used to train spatial ability and thereby facilitate learning in STEM disciplines. The study combined previous findings on the viability of spatial training for improvement in STEM outcomes and findings on the potential of video games as a training avenue for spatial ability. Specifically, we employed two commercial video games: (1) “Portal 2” as a game with high spatial complexity in gameplay demands for the experimental group and (2) “Sherlock Holmes vs. Jack the Ripper” as a game with lower spatial complexity in gameplay demands for the active control group. We expected playing the spatially more complex game to result in a stronger training effect on spatial ability. First-year civil engineering students were the target group as Sorby et al. (2018) demonstrated the effectiveness of spatial ability training for study-related outcomes in this group (using a course training).

One particular facet of spatial ability that has stood out as the most important for STEM disciplines and engineering in particular is spatial visualization ability (for a review, see Uttal & Cohen, 2012), defined as “processes of apprehending, encoding, and mentally manipulating spatial forms” (Carroll, 1993, p. 309). Many common tasks in STEM disciplines require performing such processes.

For example, engineers need to be able to visualize how different parts fit, move, and interact with each other, how to most efficiently assemble components with limited room, or read, visualize, and create topographical maps (Sorby et al., 2013). Because of this, the present study focused on the spatial visualization facet of spatial ability. Students' spatial visualization ability, in the present study expected to be improved by the training, should facilitate learning in introductory civil engineering courses, leading to improved content knowledge acquisition and better grades. The study focused on *technical mechanics* as it is a common introductory course with high spatial demands in taught concepts and used representations.

This led to the following *primary* hypotheses:

1. Video game training improves spatial visualization ability. This effect will be more pronounced in the Portal 2 group than in the Sherlock Holmes group as Portal 2's gameplay has higher spatial demands.
2. Spatial visualization ability measured after the training (posttest) predicts:
    a) total content knowledge at the end of the semester and
    b) exam grades,

    even when controlling for (prior) content knowledge and reasoning ability.

A few participants reported struggling with their assigned game, especially in the Portal 2 group. They mentioned that their struggles were related to unfamiliarity with the game's controls due to their lack of previous experience with video games. This led to the addition of the following *exploratory* hypothesis:

3. Previous video game experience affects training gains, especially in the Portal 2 group.

4.4 Materials and Method

The present research was conducted in accordance with the human subjects guidelines of national research committees in Germany, as well as the APA standards for Ethical Principles of Psychologists and Code of Conduct. At the time of data collection, institutional review board approvals were not required for this type of study in Germany.

We report how we determined our sample size, all data exclusions, all manipulations, and all measures relevant for the study, and we follow JARS (Kazak, 2018). The study was not preregistered. All data, code for the data analysis, and non-licensed research materials are available upon request from the corresponding author.

4.4.1 Sample and Procedure

Data was collected in an optional course on multimedia learning for designing presentations for civil engineering first-year students at a German university during winter semester 2021/2022. Only civil engineering students in their first semester could participate, other students were excluded. Based on the average effect size of spatial training from Uttal et al.'s (2013) meta-analysis (Hedge’s g = 0.47), a sample of 30 participants per group would be required to detect the desired training effect in each group with 80% power and α = .05 (calculated using G*Power; Faul et al., 2007). To detect a mean difference of the same size between two groups (assuming no effect in the control group), a sample of 57 participants per group would be required. We aimed to match these sample sizes but were limited to the size of the available cohort. With the limited number of participants, we assigned more participants to the experimental group (54 in the experimental group, 32 in the control group) to be able to reach at least the sample size required to reliably detect the expected training effect within the experimental group even with substantial dropout. In total, 86 participants initially took part in the course, of which 69 completed at least 4 hours of their respective training and were included in the analyses. Of these, 43% identified as female. Participants were, on average, 19.6 years old (*SD* = 1.9, range = 17-26).

The study consisted of two groups and three points of measurement. The pretest took place in the first two weeks of the semester and consisted of a demographics questionnaire including questions about available devices and about video game experience, a spatial visualization ability test, a content knowledge test on technical mechanics, and two tests for reasoning ability. Afterwards, there were four weeks of training followed by the posttest that consisted of the same tests used in the pretest and a short feedback form about the training. The delayed posttest took place at the end of the semester and again included the same test materials as the two previous tests.

The training consisted of playing one of two video games at home: “Portal 2” (Valve, 2011) for the experimental group (n = 42) and “Sherlock Holmes vs. Jack the Ripper” (“Holmes”; n = 27; Frogwares, 2009) for the control group. The games were distributed as gifts using the gaming platform Steam (Valve Corporation, 2021). Participants were assigned to the two groups first based on previous experience and available devices; that is, participants who played either game before were assigned to the other group. MacOS users were assigned to the experimental group as Portal 2 had a MacOS client available while Sherlock Holmes did not. No MacOS users had played Portal 2 before so there were no conflicts between these rules. Afterwards, all other participants were assigned randomly to the groups in a way that resulted in similar average pretest levels of spatial visualization ability in the two groups. Because not all participants were assigned to their experimental groups completely at random, the study had a quasi-experimental design.

Participants were told the purpose of the study would be training of various aspects of problem-solving ability to facilitate learning in their introductory courses. Participants were tasked to play for 8 hours during the training period but were included in the data analyses if they played at least 4 hours. Most participants played at least the required 8 hours (3 between 4 and 7 hours, 4 between 7 and 8 hours; $M = 8.69$, $SD = 1.54$). Participants were also asked to upload their save game file after the training to verify their playtime and to allow for analyses of game progress. Afterwards, grades from the technical mechanics exam were collected from the teacher ($n_{\text{Portal}} = 26$ and $n_{\text{Holmes}} = 17$ took part in the exam) as a criterion directly related to the participants' study progress. The proportion of correct answers (i.e., ranging from 0 to 1) in the exam were used as detailed indicators of performance in the exam and are called "exam grades" in later sections for ease of use.

Participants were compensated with 50 € if they completed all parts of the study and 25 € if they lacked either one of the tests or the training (as indicated by submitting a save game file).

4.4.2 Data Preparation

All data was prepared and analyzed using R (version 4.2.2; R Core Team, 2022) with RStudio (version 2023.03.0.386; RStudio Team, 2023) and the tidyverse package environment (Wickham et al., 2019). Participants were scored if they completed at least half of the items of the respective tests to ensure a valid measurement. All tests were scaled according to the Rasch model using the R package TAM (version 4.0-16; Robitzsch et al., 2022). Weighted likelihood estimates (WLE; Warm, 1989) were calculated as person ability indicators and used in the analyses. Outliers, defined as scores that lie more than 1.5 interquartile ranges (IQR) beyond the first or third quartile, were winsorized to the most extreme value not yet defined as an outlier, so 1.5 IQR below the first or 1.5 IQR above the third quartile.

Item difficulties for all tests were fixed on the pretest scores (unless noted otherwise) if the infit mean square residuals ("infit") with fixed difficulty were within acceptable limits (between 0.7 and 1.3; Bond & Fox, 2007). If an item's infit was outside those limits, it displayed item parameter drift and its difficulty was not fixed to its pretest score.

Self-reported playtime with the game was verified by comparing the self-reported playtime to the playtime logged by the gaming platform Steam. Participants were defined as having previous video game experience if they reported that they either currently played more than 5 hours of video games per week or had been playing video games for at least 10 hours per week in the past.

Table 4.1

Description of Tests Used

Ability/skill	Test content	Number of Items	Time limit	WLE reliability (pre/post/delayed)	Reference
Spatial visualization	Mental rotation	24	None	.79/.80/.83	Vandenberg and Kuse (1978)
Content knowledge	Technical mechanics	36	40 min	.41/.59/.79	Dammann (2016)
Reasoning	Odd-one-out Matrices	15 15	5 min 6 min	.67/.73/.71	Weiß (2006)

4.4.3 Materials

All tests and questionnaires were conducted within the open-source learning management system Moodle (Dougiamas, 2019) where most of the course work took place as well. Participants worked on the tests and questionnaires during course hours in a seminar room. Table 4.1 gives an overview of the tests used.

4.4.3.1 Spatial Visualization Ability

We used a mental rotation test to assess spatial visualization ability following Vandenberg and Kuse's (1978) mental rotation paradigm with no time limit. Average time of completion across time was 15.9 minutes (pretest: 17.2 min; posttest: 17.7 min; delayed posttest: 13.3 min). Participants were asked to mark the two rotated versions (targets) of a reference stimulus out of four options (Figure 4.1). An item was scored as correct if a participant correctly marked both rotated targets. If more or less than two stimuli were marked, the item was always scored as wrong. Participants were not scored if they consistently marked only one option instead of the required two, because this is most likely an issue with the instruction and not the participants’ spatial abilities. This was the case for one participant on the pretest and two different participants on the delayed posttest. Stimuli were taken from the stimulus library of Peters and Battista (2008). Stimuli were rotated around one or two cardinal axes in relation to the reference stimulus, distractors depicted mirrored versions of the same object and were also rotated. Targets were rotated from the reference stimulus by 55 to 180 degrees (via direct rotation which could be around a cardinal axis or a skewed axis). To reduce ceiling effects after training, 11 items that were solved correctly by at least 75% of the participants in the pretest were replaced for the posttest and delayed posttest.

According to Carroll (1993), mental rotation tests requiring the rotation of complex stimuli (such as 3D objects like the block figures used in this study) measure the spatial visualization facet of spatial ability if they are conducted without a time limit. Previous studies showed that this test includes a clear spatial visualization component (Nolte et al., 2022) beyond the general intelligence component most ability tests share due to a positive manifold structure (e. g., Carroll, 1993; Spearman, 1904).

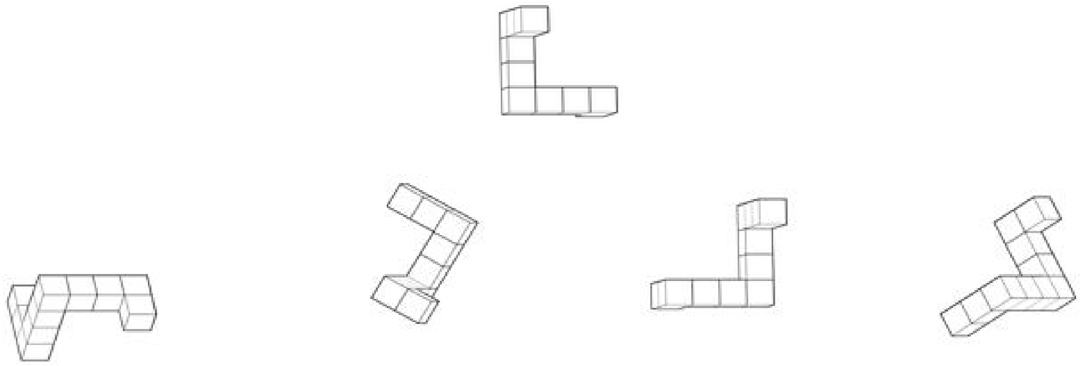
Of the 13 repeated items, five displayed item parameter drift at the posttest and three did at the delayed posttest and therefore did not have their difficulties fixed to their values from pretest.

4.4.3.2 Content Knowledge

We used a content knowledge test (Dammann, 2016) to assess the participants knowledge in Technical Mechanics taught in introductory courses at the university. At pretest, test results represented prior knowledge that participants had acquired before attending the university, most likely during high school. At the posttest and delayed posttest, test results included content knowledge acquired in their university courses on Technical Mechanics. The average completion time was much shorter than the time limit with 23.3 min across time (pretest: 28.4 min; posttest: 22.5 min; delayed posttest: 18.1 min). The test used a single-select format with four options to choose from (Figure 4.2). As participants had little prior knowledge of the topic, parameter estimation at the pretest had low reliability and the data from the delayed posttest was used as reference for fixing instead. Of the 36 items, twelve in the pretest and one item in the posttest displayed item parameter drift and therefore did not have their difficulties fixed to their difficulties from delayed posttest.

Figure 4.1

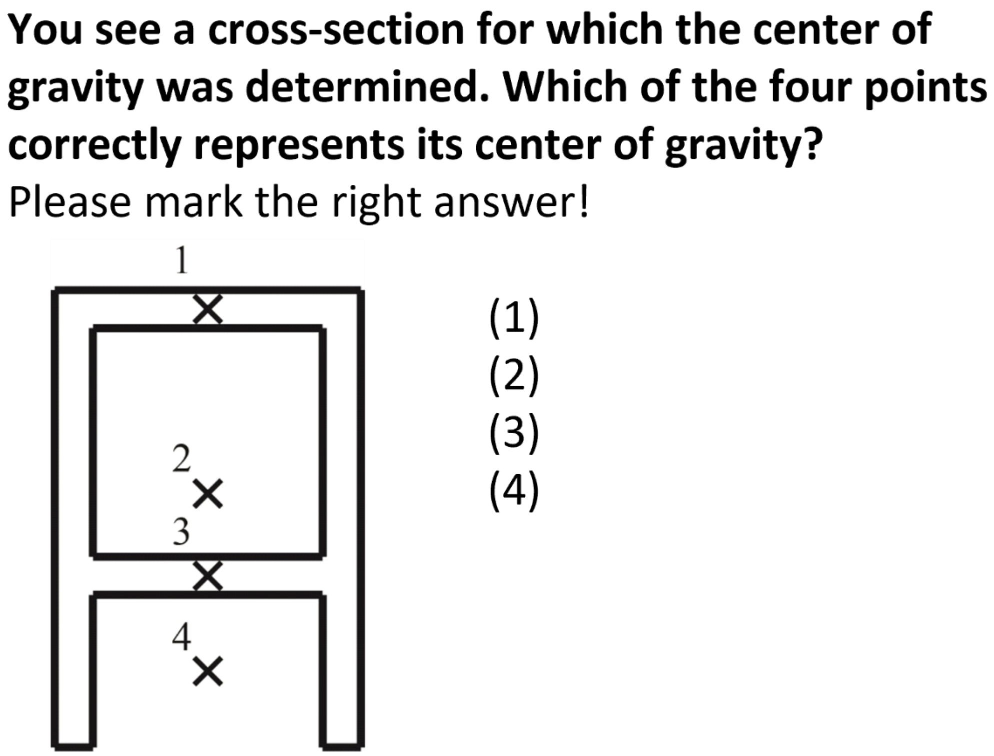
Example Item of the Mental Rotation Test



Note. The participants were asked to pick the two objects in the bottom row that were rotated versions of the objective at the top.

Figure 4.2

Example Item of the Content Knowledge Test



You see a cross-section for which the center of gravity was determined. Which of the four points correctly represents its center of gravity?
Please mark the right answer!

(1)
(2)
(3)
(4)

Note. Translation by the authors.

4.4.3.3 Reasoning Ability

To measure reasoning ability, we administered two tests from the Culture Fair Test battery (Weiß, 2006): In the first test, the “Odd-Man-Out” task (“OOO”, Subscale 2), participants are asked to mark one out of five objects that does not fit the others by detecting the rule that connects the other four objects (e.g., a triangle among circles). The second test, the Figural Matrices Task (“MAT”, Subscale 3), is similar to Raven’s matrices (Raven, 2000): Participants see a matrix of differently patterned tiles in which one tile is empty. Participants then need to analyze the rules that guide the tiles’ patterns and use this to infer the pattern of the missing tile, selecting the missing tile from five options. An example for such a rules system could be as follows: “in each column, triangles are either right- or left-sided, and in each row they are either white or colored”, which leads to “the triangle on the missing tile needs to have the same orientation as the ones in the tiles above it and the same color as the ones next to it”. As the two tests both measure reasoning ability with very similar task requirements (“recognize and apply pattern rules”), they were scaled together in a single Rasch model. An analysis of dimensionality confirmed that scaling them together in a one-dimensional Rasch model does not significantly reduce model fit compared to the two-dimensional model ($\chi^2(2) = 5.35$, $p = .07$). For scaling at the posttest, four items from the OOO task and five items from the MAT task did not have their difficulties fixed to the values from pretest due to item parameter drift. At the delayed posttest, two items from the OOO task and five items from the MAT task did not have their difficulties fixed to their pretest values.

4.4.3.4 Training Video Games

We used two commercially available video games as immersive and enjoyable training for both groups. Paired with the training taking place at the participants’ homes, this increased ecological validity for studying the effects of everyday video game play for entertainment on cognitive abilities compared to training in a lab with games especially created with the aim to train abilities (such as “brain training games”) instead of entertainment and recreation.

4.4.3.4.1 Portal 2

Portal 2 is a three-dimensional first-person puzzle game by Valve (2011) in which the player navigates through so-called “test chambers”. In each chamber, the player needs to reach the exit into the next area, overcoming different hurdles using (often interacting) tools of increasing complexity. The main tool of the game is the namesake portal gun, which can create two portals that are interconnected, similar to a wormhole: Everything, including the player, moving into one portal exits from the other, preserving momentum and (relative) direction. This creates many possibilities for spatially demanding strategies by itself such as the “fling” depicted in Figure 4.3 where the player jumps into a portal placed on a lower floor to be catapulted out from the second portal, crossing both distances and heights otherwise impossible to overcome. Many other tools interact with these portals (and each other), such as weighted cubes, lasers, light bridges, or tractor beams, which can be directed through portals to activate buttons, laser detectors, or to reach new areas.

Portal 2’s gameplay offers a substantial advantage over commonly used training games from the action video game genre: Portal 2 requires very little mechanical video game skills (such as reaction speed, fast hand-eye-coordination, or remembering complex mission requirements). Most of its gameplay demands are purely spatial and spatial problem solving-related. Previous research showed succinct training effects of Portal 2 on these abilities, even beyond what playing a so-called “brain training game” could accomplish (Shute et al., 2015). Some researchers also evaluated the game relative to rules application as an element of problem solving abilities and fluid intelligence (Foroughi et al., 2016; Shute & Wang, 2015), using either the base game’s levels or levels specifically created for this purpose with the game’s level editor (“Puzzle Creator”).

4.4.3.4.2 Sherlock Holmes vs. Jack the Ripper

In the active control group, we used Sherlock Holmes vs. Jack the Ripper ("Holmes", Frogwares, 2009) as the training game. It is a game in the setting of the popular Sherlock Holmes novels and the fifth installation of the Sherlock Holmes series of adventure games. To our knowledge, there had been no previous research on cognitive effects of games from this series. We used the Holmes game because it displays little spatial requirements in its core gameplay, which is based on piecing together clues and logically deducting the next step in solving the case of Jack the Ripper. The strong emphasis on logical deduction in gameplay enabled the cover story of training logical problem solving (compared to spatial problem solving, as advertised for the Portal group), reducing harmful motivational effects of being in an obvious control group. Example screenshots showcasing the gameplay are displayed in Figure 4.4.

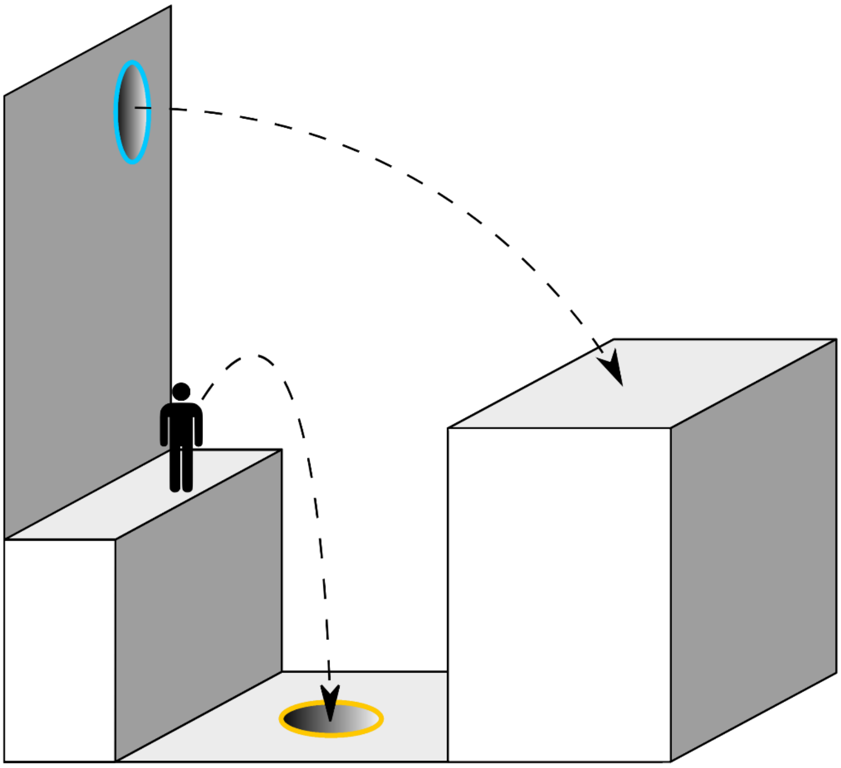
4.4.4 Analyses

To ensure that training effects did not vary between groups because of training duration (as 8 hours were required but participants could voluntarily play more than that) or enjoyment of their training game (self-reported after the training period on a scale from 1 “no fun” to 5 “lots of fun”), playtime and enjoyment were compared between groups beforehand.

For the intervention check (Hypothesis 1), we conducted a mixed ANOVA with repeated-measures with spatial visualization ability as the dependent variable, point of measurement (i.e., pretest, posttest, delayed posttest) as the within-subject independent variable, and group as the between-subjects independent variable. Post-hoc two-sided t-tests for main effects and interactions relevant to the hypotheses were conducted using Holm correction of p values. If the main effect or interaction was not significant in the RM-ANOVA, the post-hoc tests were interpreted as explorative.

Figure 4.3

Depiction of the "Fling" Technique as an Example of Portal 2's Gameplay



Note. Own image, adapted from https://commons.wikimedia.org/w/index.php?curid=5377237

Figure 4.4

Example Screenshot of Sherlock Holmes Versus Jack the Ripper

Note. Own image.

For Hypothesis 2, hierarchic multiple regression analyses with either delayed posttest content knowledge or exam grades as criteria and posttest spatial visualization ability as the main predictor were used. Prior knowledge and posttest reasoning ability were added first as the control variables in both analyses, as prior knowledge and intelligence (of which reasoning ability is a strong indicator) are known as a strong predictors of performance in such tests and exams (e.g., Burton & Ramist, 2001; Freyer et al., 2014; Hailikari, 2009; Neisser et al., 1996). In the exam grade model, delayed posttest content knowledge was added as another control variable before entering posttest spatial visualization ability in the model. Model comparison tests were used to assess whether the inclusion of additional variables significantly improved the model's R^2, in which case the variable was kept in the model. Furthermore, the regression models for delayed posttest content knowledge and exam grades were combined into a mediation analysis to evaluate whether spatial visualization ability affects exam grades beyond facilitating content knowledge acquisition (represented by the indirect effect via delayed posttest content knowledge as mediator). In the mediation model, parameter estimates were tested for statistical significance by bias-corrected bootstrapping using 10,000 bootstrap samples. To test whether it is a partial or full mediation, the direct path from spatial visualization ability to exam grades was fixed to zero in a separate model. This constrained model was then compared to the unconstrained model in terms of model fit: The constrained model was accepted if the χ^2 model comparison test was not significant and the difference in comparative fit index (CFI) was smaller than .002 (Meade et al., 2008). If this path could be set to 0, then the effect of spatial visualization ability was fully mediated by delayed posttest content knowledge, otherwise it is a partial mediation (Cheon & MacKinnon, 2012). The same procedure was applied for other paths that had significant bivariate correlations but were not significant anymore in the multiple regression to reduce the model's complexity.

To translate the effect of spatial visualization ability on exam grade into practice-oriented measures, we took the total effect of spatial visualization ability in the mediation model as a regression coefficient and the average training gains in spatial visualization ability for each group as predictors in predicting average gains in exam grades (percentage correct) for each group. For example, if the total effect of spatial visualization ability were $b = .05$, the Portal 2 group showed an average training gain of 1 and the Holmes group showed an average training gain of 0.2, this would mean that participants in the Portal 2 group improved their proportion of correct answers in the exam on average by $0.05*1 = 0.05$ (which is five percentage points) based on the training and the Holmes group improved their exam grade by $0.05*0.2 = 0.01$ (which is one percentage point).

For Hypothesis 3, we conducted two-sample t-tests to compare participants with and without previous video game experience within both the Portal 2 group and the Holmes group to detect differences in their improvements in spatial visualization ability during the training period (calculated as difference between pretest and posttest).

All analyses used $\alpha = .05$. Confidence intervals for Cohen's *d* were created using bootstrapping using 10,000 bootstrap samples.

4.5 Results

Both groups reported similar enjoyment of their training ($M_{\text{Portal 2}} = 3.3$, $M_{\text{Holmes}} = 3.1$; $t(51.23) = 0.79$, $p = .432$, 95% CI of difference [-0.3, 0.7]). Playtime did not differ between groups as well ($M_{\text{Portal 2}} = 8.4$h, $M_{\text{Holmes}} = 9.0$h; $t(48.82) = -1.34$, $p = .185$, 95% CI of difference [-1.4, 0.3]).

4.5.1 Hypothesis 1: Video Game Training Improves Spatial Ability

Average WLE person abilities for the groups across time are displayed in Table 4.2. A single participant's low score (Portal 2 group, posttest) in the mental rotation test was declared as an outlier and henceforth winsorized (i.e., set to the most extreme value that was not yet considered an outlier). The assumption of sphericity was met in all within-subjects factors (time and the interaction of time*group; Mauchly's $W = .92$, $p = .075$), but Greenhouse-Geisser correction was applied regardless. Normality assumption did hold for all combinations of group*point of measurement (Shapiro-Wilk's $W \geq .95$, $p > .174$) except for the Holmes group at pretest (Shapiro-Wilk's $W = .91$, $p = .027$). The QQ plot showed that this was a small violation most likely due to a ceiling effect, as seven participants from that group answered all MRT items correctly which interfered with the assumed normality. Apparently, replacing the easiest items in the pretest for later tests helped, as the normality assumption did hold for posttest and delayed posttest. The assumption of variance homogeneity was violated at the pretest (Levene's $L = 4.06$, $p = .047$) but did hold at the posttest and delayed posttest (Levene's $L \leq 1.15$, $p > .295$). There was homogeneity of covariances (Box's $M = 1.34$, $p = .241$). As the present violations of assumptions were small and only in the Holmes pretest, we performed the planned mixed RM-ANOVA, as it is usually quite robust against such violations (Blanca et al., 2023).

In the RM-ANOVA, there was a significant main effect of time, indicating differences in mental rotation ability across the points of measurement ($F(1.85,116.66) = 4.21$, $p = .020$, partial $\eta^2 = .063$). The main effect of group was not significant, albeit marginal ($F(1,63) = 3.85$, $p = .054$, partial $\eta^2 = .058$). The interaction of time and group was not significant ($F(1.85,116.66) = 0.85$, $p = .422$, partial $\eta^2 = .013$), indicating that the groups did not differ significantly in their changes in spatial visualization ability across three points of measurement, including training gains between pretest and posttest.

To get a clearer picture of the hypothesis, we conducted post-hoc t-tests for both the significant main effect of time and the interaction of time with group. Figure 4.5 displays the distributions of spatial visualization ability across time and groups, including the comparisons most relevant for the hypothesis.

Concerning time, there was a significant gain in spatial visualization ability from pretest to posttest ($\Delta M = 0.49$, $t(66) = 3.84$, $p_{\text{holm}} < .001$, $d = 0.29$), which hints at a training effect across groups. Spatial ability levels dropped slightly from posttest to delayed posttest, but not significantly ($\Delta M = -0.17$, $t(64) = -1.08$, $p_{\text{holm}} = .284$, $d = -0.10$). This led to a non-significant difference between pretest and delayed posttest across groups as well ($\Delta M = 0.31$, $t(65) = 1.79$, $p_{\text{holm}} = .155$, $d = 0.18$). An explorative look into changes across time separately within each group revealed that the significant main effect of time was mainly driven by the Portal 2 group's training effect: The increase from pretest to posttest was only significant in the Portal 2 group ($\Delta M = 0.61$, $t(41) = 3.81$, $p_{\text{holm}} = .001$) with a small to medium effect size of $d = 0.44$, but not in the Holmes group ($\Delta M = 0.30$, $t(24) = 1.40$, $p_{\text{holm}} = .528$, $d = 0.12$). This indicated that the Portal 2 training, compared to the Holmes training, was more effective in training spatial visualization ability, even though the difference in the gains from pretest to posttest was not significant ($\Delta M = 0.30$, $t(50.13) = 1.1$, $p = .276$, $d = 0.28$; conducted separately). However, as spatial visualization ability decreased slightly (but not significantly) between posttest and delayed posttest within both groups, the total increase in spatial visualization ability across the semester (meaning, between pretest and delayed posttest) was not significant even in the Portal 2 group ($\Delta M = 0.45$, $t(39) = 1.95$, $p_{\text{holm}} = .117$) which initially showed a significant training effect, even though it was still of a small effect size of $d = 0.29$. Neither of the other pairwise comparisons between the three test times within a group was significant (all $p > .480$) or of relevant effect size (all $|d| < 0.11$).

Table 4.2

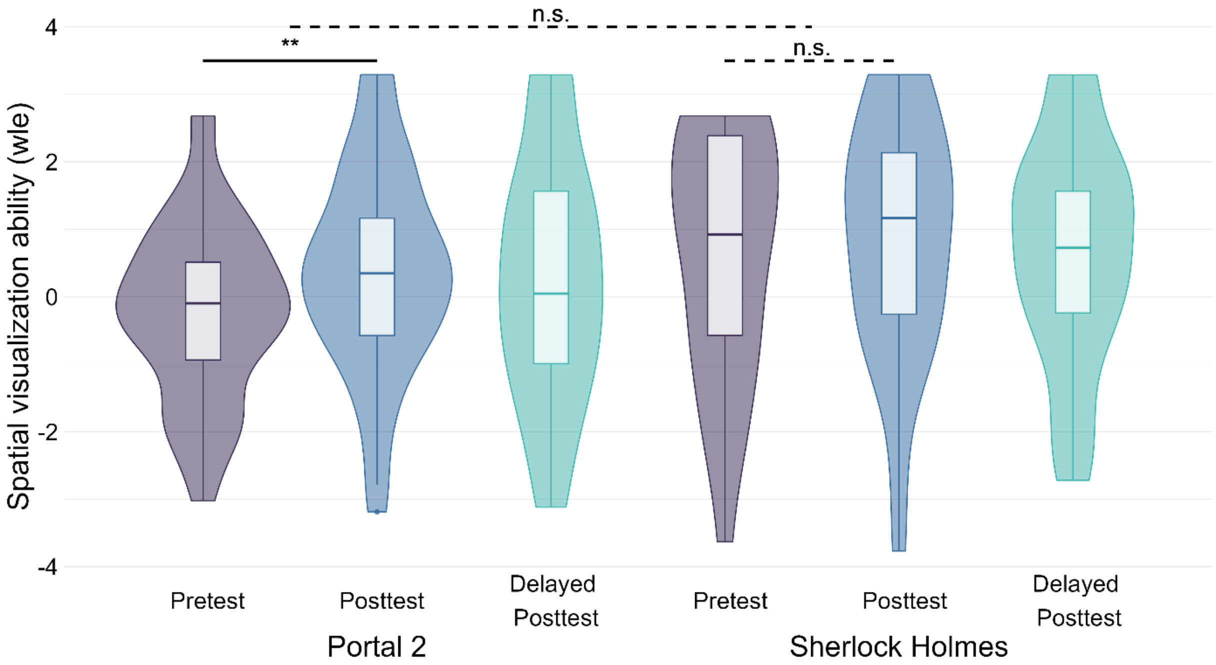
Means and Standard Deviations of Relevant Variables

Time	Spatial visualization				Content knowledge				Reasoning			
	Portal 2		Holmes		Portal 2		Holmes		Portal 2		Holmes	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pretest	-0.21	1.31	0.62	1.82	-0.56	0.49	-0.57	0.52	0.01	1.02	0.25	0.87
Posttest	0.41	1.47	0.85	1.80	-0.04	0.50	-0.09	0.58	0.14	0.94	0.56	1.36
Delayed Posttest	0.24	1.80	0.66	1.67	-0.05	0.81	0.22	0.75	0.42	1.07	0.39	1.13

Note. Content knowledge measured at pretest represented prior knowledge. All measures are in the WLE scale from their respective Rasch models which can become negative.

Figure 4.5

Distributions of Mental Rotation Ability Across Groups and Time Including Pairwise Comparisons



Note. Boxplots display medians and 25% and 75% quantiles as the box and have a maximum whisker length of 1.5 IQR. The figure displays the tests for training gains and the comparison of training gains across groups as these were the comparisons most relevant for testing the hypothesis. None of the other comparisons across time within each group or across groups at the same time were significant.

** *p* < .01

4.5.2 *Hypothesis 2: Spatial ability predicts study-related outcomes*

4.5.2.1 Content Knowledge

Hierarchical linear regression analysis showed that participants' spatial visualization ability had incremental value in predicting the delayed posttest content knowledge beyond prior knowledge (pretest content knowledge) and reasoning ability (see Table A6.6 in the Appendix), explaining 4% variance beyond the other predictors. Prior knowledge was the strongest predictor and explained just under 9% variance beyond the other predictors. Reasoning ability explained just above 1% variance when spatial visualization ability was included in the model and had no significant effect on delayed posttest content knowledge beyond the other predictors. Therefore, reasoning ability was excluded from the model without significantly reducing its total R^2.

4.5.2.2 Exam Grades

Only the delayed posttest content knowledge ($r = .70$) and spatial visualization ability ($r = .38$) displayed significant bivariate correlations with exam grades, which is reflected in negligible effects and ΔR^2 for prior knowledge and reasoning ability in the hierarchical regression analysis for predicting exam grades (see Table A6.7 in the Appendix). Interestingly though, spatial visualization ability had no significant effect on exam grades when controlling for the delayed posttest content knowledge, and its addition did not significantly improve the model's R^2 ($\Delta R^2 < .01$), indicating an overlap in explained variance between delayed posttest content knowledge and spatial visualization ability, which suggested a mediation effect as spatial visualization ability predicted delayed posttest content knowledge as seen in the previous models (see Section 4.5.1). This left the delayed posttest content knowledge as the sole significant predictor of exam grades, explaining 49% variance.

In a subsequent mediation model, all significant bivariate correlations were included as estimated paths, while paths between variables displaying no significant bivariate correlation were set to 0. The model revealed that the effect of spatial visualization ability appears to be fully mediated by content knowledge at the delayed posttest, as the direct path was not significant (in line with the multiple regression beforehand) and could be fixed to 0 without significantly reducing the model‘s fit ($\Delta\chi^2(1) = 0.57$, $p = .452$; ΔCFI = .004). Subsequently, the non-significant direct path from reasoning ability to content knowledge was also fixed to 0 without significantly reducing the model's fit ($\Delta\chi^2(1) = 1.23$, $p = .268$; ΔCFI = .002), resulting in the final model displayed in Figure 4.6. The final model fitted the data well ($\chi^2(4) = 5.09$, $p = .278$; CFI = .989, TLI = .972, RMSEA = .063, SRMR = .039). The total (indirect) effect of spatial visualization ability on exam grades was $b = .07$ ($\beta = .33$, $SE = .02$, $p = .004$, 95% CI [.026, .116]), mediated by pretest content knowledge and delayed posttest content knowledge.

Using this total effect of *b* = 0.07 as regression coefficient and the average training gains in spatial visualization ability for each group ($\Delta M_{\text{PrePost}}$ Portal 2 = 0.60, Holmes = 0.28) as predictors, the Portal 2 group improved their exam grade by 0.04 (translating to 4 percentage points) and the Holmes group by 0.02 (translating to 2 percentage points).

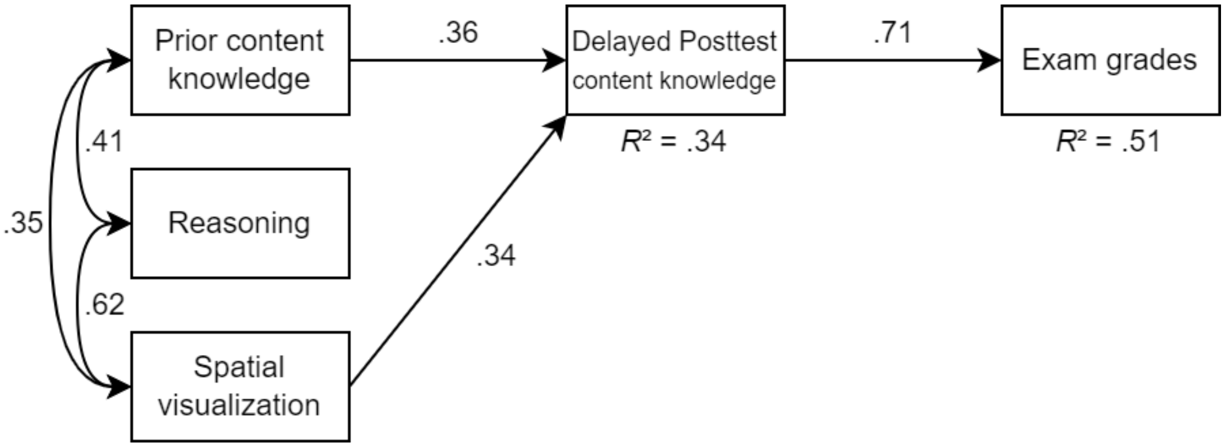
4.5.3 Hypothesis 3: Previous video game experience affects training gains

Previous video game experience was reported by 12 participants in the Portal 2 group and by 16 participants in the Holmes group.

There was no significant difference in training gain between participants with and without previous video game experience in both groups, although there was a trend in the Portal 2 group: Participants with previous video game experience improved their spatial visualization ability slightly more than participants without previous video game experience ($t(19.27) = 2.65$, $p_{\text{holm}} = .058$), with a moderate to strong effect size of $d = 0.70$. In the Holmes group, participants without previous video game experience had slightly but not significantly higher training gains ($t(18.32) = -1.05$, $p_{\text{holm}} = .310$, $d = -0.43$).

Figure 4.6

Mediation Model for Exam Grades



Note. Displayed path coefficients are standardized βs. All displayed paths are significant with $p < .01$.

4.6 Discussion

The first hypothesis was partially confirmed: There was a significant main effect of time on participants' spatial visualization ability. Post-hoc tests showed that there was an improvement from pretest to posttest but not from posttest to delayed posttest, as well as not in total from pretest to delayed posttest. The significant increase from pretest to posttest indicates that the video game training improved participants' spatial visualization ability without resulting from a simple retest effect – as in that case there should have been an improvement from posttest to delayed posttest as well. However, the non-significant total improvement when comparing the pretest at the beginning and the delayed posttest at the end of the semester might indicate that the achieved training effect was not stable. This does not contradict decent stability of training effects for spatial ability interventions reported by Uttal et al. (2013) as they only drew conclusions for delays up to one month (as very few studies in their metaanalysis had a longer delay) and there were two months between the posttest and the delayed posttest in the present study. An alternative explanation for the reduced delayed posttest scores might be that data collection for the delayed posttest was at the end of the semester which overlapped with exam preparations for the participants. This might have led to reduced motivation and thereby concentration when working on the tests. Anecdotal evidence from a few participants' feedback pointed in this direction (six participants admitted to having clicked a few items at random after being questioned on them "finishing" the tests within seconds – those participants were asked to complete the tests again). That the reduction in test performance was small and not significant could be interpreted as the training effect counteracting the losses due to reduced motivation and concentration, resulting in a non-significant total change. Future research in similar contexts should incorporate measures of (test) motivation to be able to evaluate this.

Regarding the expected stronger effect of the spatially more complex game Portal 2 on spatial visualization ability, the intervention check was inconclusive: Post-hoc tests showed that the significant pretest-posttest improvements (part of the main effect of time) mainly resulted from significant improvements in the Portal 2 group. But the interaction of time and group was not significant in the RM-ANOVA, which rendered the post-hoc tests investigating group differences exploratory in nature. This limits the conclusion that playing a spatially complex game such as Portal 2 results in a stronger training effect than a spatially simpler game such as Holmes does. There are several potential explanations for this. From a theoretical point of view, it is quite likely that the Holmes group had a slight training effect as well due to them still playing a 3D game: Earlier studies used 2D games such as Tetris or brain training games with 2D minigames as comparison games to detect training advantages of spatially more complex games (e.g., Adams & Mayer, 2012; Gagnon, 1985; Shute et al., 2015). Holmes has a first-person view for navigating through the world, even though the quite simple navigation component requires little cognitive involvement. But even just having gameplay that requires navigating through a virtual 3D world might already influence spatial ability, even if the required navigation is neither complex nor central to the intended main challenges of the game. The supposed small training effect in the Holmes group reduced the expected difference in the training gains between the groups. This further complicated finding a significant interaction by reducing the effect size of the interaction. The significance test for the interaction had a sufficient power of .79, so not finding the interaction most likely was due to the difference in training effects between the two groups being too small to bring the whole interaction term into significance when all other differences tested by the interaction (i.e., between measurements within a group or between groups at the same measurement) were negligible. Future studies should most likely use a 2D comparison game as active control group to evaluate more complex video games as intervention while still providing a plausible cover story for participants to reduce potential placebo and nocebo effects.

The second hypothesis was confirmed: For the delayed posttest content knowledge, there was a significant and substantial effect of spatial visualization ability even after controlling for prior content knowledge and reasoning ability, with the latter not being significant anymore after spatial visualization ability was added to the model. This indicated that spatial visualization ability explained unique variance beyond those control variables, even though it shared a lot of variance with reasoning ability (which is to be expected for cognitive ability tests due to their positive manifold structure, Carroll, 1993; Spearman, 1904). For exam grades, the partial effect of spatial visualization ability was not significant anymore after controlling for the delayed posttest content knowledge, for which the previous regression offered an explanation that could be confirmed by the subsequent mediation analysis: The effect of spatial visualization ability (that was significant in a bivariate analysis) was fully mediated by the delayed posttest content knowledge. In the application context, this is quite plausible: spatial visualization ability supported the previous acquisition of content knowledge and had no effect on performance in the exam situation beyond that, as all of the exam's spatial demands stemmed from the content it was about and not from the way the exam situation worked.

Participants from the Portal 2 group achieved 4 percentage points more in the exam than they would have without the training, while participants from the Holmes group achieved 2 percentage points more. Grades in German universities range from 1 to 6 in steps of thirds, with 1 being the best. In this exam, 4.4 percentage points equaled one step (e.g., from 2.3 to 2.0). So based on the model the Portal 2 group scored almost one grade step better in their exam than they would have without the training, while the Holmes group improved by almost half of a grade step. This is a visible improvement in practical terms after only 8 hours of play.

The analysis on the exploratory third hypothesis, sparked by participant feedback, showed no significant differences in training gains between participants with and without previous video game experience in either group. However, the bootstrapped confidence interval for Cohen's *d* in the Portal 2 group did not include 0, which indicated that there might be an effect that was not detected because of small sample size (power of the t-test was .51). This suggests that potentially for complex games such as Portal 2 previous video game experience might be helpful or even necessary to be able to profit from training, but further research is needed to confirm this.

4.6.1 Limitations

The main limitation of the present study is the inconclusive intervention check: The interaction of group and time was not significant, rendering the post-hoc test exploratory that showed a significant improvement in spatial ability from pretest to posttest in the Portal 2 group but not the Holmes group. This is most likely due to multiple factors, most importantly that Holmes was too strict of a control condition as a 3D game itself and therefore the effect size of the group difference in training gains was too small. The sample size was sufficient to detect the interaction with a power of .79, but as the only expected group difference in the comparisons included in the interaction was in the pretest-posttest difference, the effective power for the one relevant comparison was likely smaller (as indicated by the a priori power analysis using a two-sample t-test on the difference in training gains reported in section 3.1).

Another limitation is that group assignment could not be completely random due to technical constraints or previous experience with either training game. Because of this, group assignment was only randomized for matching pretest spatial visualization ability after these conditions were applied. Additionally, dropout from the study affected average spatial visualization ability differently between groups: In the Portal 2 group, dropped-out participants showed similar pretest spatial visualization ability to the ones that did not drop out (dropouts scored 69% and recurring participants scored 73% in the MRT), while in the Holmes group mostly participants with low spatial visualization ability dropped out (43% in dropouts compared to 78% in recurring participants). However, the resulting difference in pretest spatial visualization ability was not significant at any point (strongest at pretest, $d = .51$, $p = .051$; weaker at the posttest and delayed posttest, $d < .27$, $p \geq .298$).

4.6.2 Conclusion and Implications

The present study suggested a general effect of playing 3D games on spatial visualization ability. Spatially more complex games most likely result in a stronger effect, but more research is needed to conclusively answer the question concerning which aspects of a game specifically improve which aspects of spatial ability. Furthermore, it could be shown that spatial visualization ability affects the acquisition of content knowledge in engineering, which is in line with previous research (e.g., Nolte et al., 2023; Sorby et al., 2018). The improved acquisition of content knowledge also led to improved exam grades in Technical Mechanics: Training spatial visualization ability improved civil engineering students' exam grades in Technical Mechanics by four percentage points (almost one grade step) in the experimental group and half as much in the control group, mediated by the enhanced content knowledge acquisition.

The present study thereby demonstrated the potential for using video games to improve student success in engineering in the form of exam grades. Video games offer a training tool that is readily available, requires little effort from teachers and already is an enjoyable leisure time activity for many students. However, previous video game experience might affect the effectiveness of training when using complex games and should be considered when selecting training video games.

4.7 References

Achtman, R. L., Green, C. S., & Bavelier, D. (2008). Video games as a tool to train visual skills. *Restorative Neurology and Neuroscience*, *26*(4-5), 435–446.

Adams, D. M., & Mayer, R. E. (2012). Examining the connection between dynamic and static spatial skills and video game performance. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *34th Annual meeting of the Cognitive Science Society 2012 (CogSci 2012): Building bridges across cognitive sciences around the world* (Vol. 34, pp. 1254–1259). Curran Associates, Inc.

Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2023). Effects of action video game play on cognitive skills: A meta-analysis. *Technology, Mind, and Behavior*, *4*(1). https://doi.org/10.1037/tmb0000102

Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal data in repeated measures ANOVA: Impact on type I error and power. *Psicothema*, *35*(1), 21–29. https://doi.org/10.7334/psicothema2022.292

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Erlbaum.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, *129*(3), 387–398. https://doi.org/10.1016/j.actpsy.2008.09.005

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980*. Research Report No. 2001-2. College Entrance Examination Board.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

CBI. (2011). *Building for growth: Business priorities for education and skills*. https://help.open.ac.uk/students/_data/documents/careers/building-for-growth-cbi-education-and-skills-survey-2011.pdf

CBI. (2019). *Getting young people 'work ready'*. https://www.cbi.org.uk/media/2960/cbi_work-readiness.pdf

Chen, X. (2009). Students who study science, technology, engineering, and mathematics (STEM) in postsecondary education. Stats in Brief. Nces 2009-161. *National Center for Education Statistics*.

Cheon, J., & MacKinnon, D. P. (2012). Mediation/indirect effects in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 417–435). Guilford Press.

Dammann, E. (2016). *Entwicklung eines Testinstruments zur Messung fachlicher Kompetenzen in der Technischen Mechanik bei Studierenden ingenieurwissenschaftlicher Studiengänge*

[Development of a test measuring engineering students‘ expertise in technical mechanics] [Dissertation]. Universität Stuttgart. https://doi.org/10.18419/opus-9073

Dougiamas, M. (2019). *Moodle* (Version 3.5.5+) [Computer software]. Moodle Pty Ltd. West Perth, Australia. https://moodle.org/

Duesbury, R. T., & O’Neil, H. F. (1996). Effect of type of practice in a computer-aided design environment in visualizing three-dimensional objects from two-dimensional orthographic projections. *Journal of Applied Psychology*, *81*, 249–260. https://doi.org/10.1037/0021-9010.81.3.249

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146 (version 3.1.9.4).

Foroughi, C. K., Serraino, C., Parasuraman, R., & Boehm-Davis, D. A. (2016). Can we create a measure of fluid intelligence using Puzzle Creator within Portal 2? *Intelligence*, *56*, 58–64. https://doi.org/10.1016/j.intell.2016.02.011

Freyer, K., Epple, M., Brand, M., Schiebener, J., & Sumfleth, E. (2014). Studienerfolgsprognose bei Erstsemesterstudierenden in Chemie. *Zeitschrift für Didaktik der Naturwissenschaften*, *20*(1), 129–142. https://doi.org/10.1007/s40573-014-0015-3

Frogwares. (2009). *Sherlock Holmes versus Jack the Ripper* [Computer software]. Frogwares; Focus Home Interactive. https://store.steampowered.com/app/11190/Sherlock_Holmes_versus_Jack_the_Ripper/

Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. *ECTJ*, *33*(4), 263–275. https://doi.org/10.1007/BF02769363

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. https://doi.org/10.1038/nature01647

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*(1), 217–245. https://doi.org/10.1016/j.cognition.2005.10.004

Hailikari, T. (2009). *Assessing university students’ prior knowledge: Implications for theory and practice*. Helsinki University Print. https://helda.helsinki.fi/handle/10138/19841

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689. https://doi.org/10.1037/0022-0663.91.4.684

Heublein, U., Richter, J., & Schmelzer, R. (2020). *Die Entwicklung der Studienabbruchquoten in Deutschland.* https://doi.org/10.34878/2020.03.DZHW_BRIEF

Judd, N., & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour.* Advance online publication. https://doi.org/10.1038/s41562-021-01118-4

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *The American Psychologist*, *73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31*, 549–579. https://doi.org/10.1080/15326900701399897

Lowrie, T., Logan, T., & Hegarty, M. (2019). The influence of spatial visualization training on students' spatial reasoning and mathematics performance. *Journal of Cognition and Development*, *20*(5), 729–751. https://doi.org/10.1080/15248372.2019.1653298

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*, 344–351. https://doi.org/10.1016/j.paid.2010.03.022

Maquet, L., Dunbar, R., Buckley, J., Seery, N., & Sorby, S. A. (2022). A review of the literature to inform the efficacy of a spatial skills intervention for secondary level STEM education. In D. Gill, J. Tuff, T. Kennedy, S. Pendergast, & S. Jamil (Eds.), *PATT39 Conference 2022. PATT on the edge: Technology, innovation and education* (pp. 496–507).

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of Applied Psychology*, *93*, 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: Developmental and Educational Implications. In J. B. Benson (Ed.), *Advances in child development and behavior* (Vol. 42, pp. 197–243). Elsevier. https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Mix, K. S., Levine, S. C., Cheng, Y.-L., Stockton, J. D., & Bower, C. (2021). Effects of spatial training on mathematics in first and sixth grade children. *Journal of Educational Psychology*, *113*(2), 304–314. https://doi.org/10.1037/edu0000494

Molina-Carmona, R., Pertegal-Felices, M. L., Jimeno-Morenilla, A., & Mora-Mora, H. (2019). Assessing the Impact of Virtual Reality on Engineering Students' Spatial Ability. In A. Visvizi, M. D. Lytras, & L. Daniela (Eds.), *Emerald studies in higher education innovation and technology. The future of innovation and technology in education: Policies and practices for teaching and learning excellence* (pp. 171–185). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78756-555-520181013

National Academies Committee on Science, Engineering, and Public Policy. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. National Academies Press.

Neisser, U., Bodoo, G., Bouchard, T. J., JR., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., Urbina, S., & Boodoo, G. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Newzoo. (2022). *Global players per region*. https://newzoo.com/key-numbers

Nolte, N., Fleischer, J., Spoden, C., & Leutner, D. (2023). *Cross-disciplinary impact of spatial visualization ability on study success in higher education* [Manuscript submitted for publication]. Department of Educational Sciences, University of Duisburg-Essen.

Nolte, N., Schmitz, F., Fleischer, J., Bungart, M., & Leutner, D. (2022). Rotational complexity in mental rotation tests: Cognitive processes in tasks requiring mental rotation around cardinal and skewed rotation axes. *Intelligence*, *91*, 101626. https://doi.org/10.1016/j.intell.2022.101626

Pallrand, G. J., & Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching*, *21*, 507–516. https://doi.org/10.1002/tea.3660210508

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260–264. https://doi.org/10.1016/j.bandc.2007.09.003

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*(6), 901–910. https://doi.org/10.1016/j.econedurev.2010.07.009

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*(2), 154–166. https://doi.org/10.1037/a0019440

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2) [Computer software]. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/

Rahimi, S., Shute, V. J., Kuba, R., Dai, C.-P., Yang, X., Smith, G., & Alonso Fernández, C. (2021). The use and effects of incentive systems on learning and performance in educational games. *Computers & Education*, *165*, 104135. https://doi.org/10.1016/j.compedu.2021.104135

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Robitzsch, A., Kiefer, T., & Wu, M. (2022). *TAM: Test Analysis Modules* (Version 4.0-16) [Computer software]. https://CRAN.R-project.org/package=TAM

Rochford, K. (1985). Spatial learning disabilities and underachievement among university anatomy students. *Medical Education*, *19*(1), 13–26. https://doi.org/10.1111/j.1365-2923.1985.tb01134.x

RStudio Team. (2023). *RStudio. Integrated Development Environment for R* (Version 2023.03.0.386) [Computer software]. RStudio, PBC. Boston, MA. http://www.rstudio.com/

Russell-Gebbett, J. (1985). Skills and strategies – pupils' approaches to three-dimensional problems in biology. *Journal of Biological Education*, *19*(4), 293–298. https://doi.org/10.1080/00219266.1985.9654755

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, *80*, 58–67. https://doi.org/10.1016/j.compedu.2014.08.013

Shute, V. J., & Wang, L. (2015). Measuring problem solving skills in Portal 2. In P. Isaías, J. M. Spector, D. Ifenthaler, & D. G. Sampson (Eds.), *E-learning systems, environments and approaches* (pp. 11–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-05825-2_2

Sorby, S. A., Casey, M. B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, *26*, 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sorby, S. A., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences*, *67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Spearman, C. E. (1904). "General intelligence" - objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92–104. https://doi.org/10.1037/a0019491

Talley, L. H. (1973). The use of three-dimensional visualization as a moderator in the higher cognitive learning of concepts in college level chemistry. *Journal of Research in Science Teaching*, *10*(3), 263–269. https://doi.org/10.1002/tea.3660100311

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 147–181). Elsevier. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Valve. (2011). *Portal 2* [Computer software]. Valve. https://store.steampowered.com/app/620/Portal_2/

Valve Corporation. (2021). *Steam* [Computer software]. https://store.steampowered.com/

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835. https://doi.org/10.1037/a0016127

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450. https://doi.org/10.1007/BF02294627

Weiß, R. H. (2006). *CFT 20-R Grundintelligenztest Skala 2* [CFT 20-R Base intelligence test scale 2]. Hogrefe.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wu, H.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492. https://doi.org/10.1002/sce.10126

5 Zusammenfassende Diskussion

Zentrales Thema der vorliegenden Arbeit war die Untersuchung der Bedeutung räumlicher Fähigkeiten für den Studienerfolg in MINT-Fächern, die bereits grundlegend in zahlreichen Studien gezeigt werden konnte (für ein Review siehe Maquet et al., 2022). Die Literatur weist allerdings eine Reihe Lücken auf, die die vorliegende Arbeit helfen sollte zu schließen. So sollte in der vorliegenden Arbeit zuerst der Effekt räumlicher Fähigkeiten von dem der generellen Intelligenz *g* abgegrenzt werden, da sie eng zusammenhängen und gemeinsam in die Leistung in räumlichen Tests hineinspielen (Carroll, 1993). Diese Problematik wurde in der bisherigen Forschung kaum berücksichtigt, was die Interpretierbarkeit bisheriger Befunde einschränkt. Diese Abgrenzung wurde sowohl mittels einer psychometrischen Analyse eines für Studien zum Zusammenhang räumlicher Fähigkeiten mit MINT-Leistungen typischen Tests räumlicher Fähigkeiten, dem mentalen Rotationstest (Vandenberg & Kuse, 1978), untersucht als auch durch statistische Kontrolle von *g* in folgenden Studien.

Weiterhin sollten unterschiedliche Fächer hinsichtlich der Bedeutung räumlicher Fähigkeit für fachspezifischen Erfolg verglichen werden, da deskriptive Studien stark variierende räumliche Fähigkeiten zwischen unterschiedlichen Fächern fanden (Shea et al., 2001), was auf variierende Bedeutung für fachlichen Erfolg hinweist. Bisherige Studien ermöglichten allerdings keine direkten Vergleiche des Zusammenhangs räumlicher Fähigkeiten mit fachspezifischen Erfolgsmaßen zwischen unterschiedlichen Fächern. Hierbei sollte auch die Spezifität räumlicher Fähigkeiten für MINT-Fächer geprüft werden, da die bisherige Literatur Zusammenhänge vor allem in MINT-Fächern fand.

Abschließend sollte der korrelative Zusammenhang mittels eines Videospieltrainings auf Kausalität geprüft und in der Praxis genutzt werden. Das Videospieltraining wurde hierbei gewählt, um die bisherige Befundlage zu Interventionsstudien für räumliche Fähigkeiten im MINT-Kontext um Daten zu Videospieltrainings als dritter Trainingsart neben der Übung räumlicher Aufgaben und kursbasierten Trainings mit meist fachlichem Bezug (Uttal et al., 2013) zu erweitern.

Im Folgenden werden die zentralen Ergebnisse der in der vorliegenden Arbeit enthaltenen Studien zusammengefasst und anschließend eingeordnet.

5.1 Zentrale Ergebnisse

5.1.1 Studie 1

Die erste Studie diente im Kontext der vorliegenden Arbeit vor allem der psychometrischen Absicherung des in den darauffolgenden Studien verwendeten mentalen Rotationstests nach Vandenberg und Kuse (1978).

Hierbei wurde zuerst untersucht, ob die Art der benötigten Rotation (um eine Kardinalachse der Bildebene bzw. des Objekts) einen Einfluss auf die Dimensionalität des gemessenen Konstrukts mentaler Rotationsfähigkeit hat, indem diese komplexere Art der Rotation bei der Bearbeitung zusätzliche kognitive Anforderungen an die Teilnehmenden stellt. Dies hätte Einfluss auf die Interpretation der Ergebnisse von Tests in anderen Studien, die Items mit sowohl Kardinalachsenrotation als auch schiefer Rotation beinhalten. Zur Prüfung dieser Frage wurde zuerst ein Bifaktor-Messmodell mit einem generellen Faktor mentaler Rotationsfähigkeit, auf dem alle Items luden, und einem geschachtelten spezifischen Faktor, auf dem nur Items luden, die Rotationen um eine schiefe Rotationsachse erforderten, aufgestellt. Dieses Messmodell wurde anschließend mit einem Messmodell mit nur einem Faktor verglichen, auf dem alle Items luden. Dabei zeigte sich, dass die Aufnahme des geschachtelten spezifischen Faktors komplexer schiefer Rotation in das Messmodell dieses nicht signifikant gegenüber dem Messmodell mit nur dem gemeinsamen Faktor verbesserte. Außerdem war der spezifische Faktor nicht adäquat statistisch identifiziert, da seine Varianz nicht signifikant sowie seine standardisierten Ladungen sehr gering waren. Sowohl der Modellvergleich als auch die mangelnde Identifikation des spezifischen Faktors für komplexe schiefe Rotation ließen darauf schließen, dass die Differenzierung zwischen Items mit Kardinalachsenrotation und Items mit schiefer Rotation nicht notwendig ist, sondern dass beide gemeinsam denselben Faktor mentaler Rotationsfähigkeit gleichermaßen (also mit vergleichbar starken Ladungen im einfaktoriellen Modell) abbilden.

Die zweite Fragestellung der Studie untersuchte, inwieweit räumliche Visualisierungsfähigkeit den Rotationsfaktor (bzw. die Rotationsfaktoren, im geschachtelten Modell) über generelle Intelligenz *g* (operationalisiert durch Reasoning) hinaus vorhersagt, um so auch Aussagen darüber zu treffen, ob bei der Lösung der Rotationsaufgaben anzunehmenderweise räumliche Prozesse beteiligt waren bzw. räumliche Strategien eingesetzt wurden, um die Konstruktvalidität des Tests abzusichern. Von besonderem Interesse wären hier Unterschiede in der Vorhersage des spezifischen Faktors für komplexe schiefe Rotation gegenüber dem allgemeinen Rotationsfaktor gewesen, da dies auf Unterschiede in den angewandten Strategien hindeuten würde. Da der spezifische Faktor allerdings nicht identifiziert war und entsprechend auch seine Prädiktion mangels Varianz nicht möglich war, entfiel dieser Teil der Fragestellung. Zur Vorhersage des einfaktoriellen Modells mentaler Rotationsfähigkeit wurde auf Prädiktorseite ebenfalls ein Bifaktormodell verwendet. Der allgemeine Faktor, auf dem sowohl die Reasoningtests als auch die räumliche Visualisierungsfähigkeitstests luden, sollte *g* abbilden, während auf dem geschachtelten spezifischen Faktor räumliche Visualisierungsfähigkeit nur die verwendeten räumlichen Tests luden. Hier zeigte sich, dass der spezifische Faktor räumliche Visualisierungsfähigkeit den Faktor der mentalen Rotationsfähigkeit substanziell vorhersagte, obwohl die *g* messenden Anteile aus den räumliche Visualisierungsfähigkeitstests herausgezogen waren. Dadurch ergab sich die Schlussfolgerung, dass zur Lösung von mentalen Rotationsaufgaben räumliche Visualisierungsfähigkeit benötigt wird und die

Testleistung nicht allein über die auch in räumlichen Tests enthaltene *g*-Komponente erklärt werden kann. Entsprechend bestätigen die Ergebnisse dieser Studie die Eignung des Vandenberg und Kuse (1978) mentalen Rotationstests für die Messung räumlicher Visualisierungsfähigkeit. Allerdings sollte bei der Untersuchung des Zusammenhangs der Testleistung mit anderen Variablen wie Studienleistungen *g* immer mit kontrolliert werden, da der Test auch eine substanzielle *g*-Komponente enthält und sonst unklar wäre, ob die Vorhersagekraft der Testleistung durch die *g*-Komponente oder die räumliche Visualisierungsfähigkeitskomponente kommt.

5.1.2 Studie 2

Im Zentrum der zweiten Studie stand der Vergleich des Zusammenhangs zwischen räumlichen Fähigkeiten und Studienerfolg in unterschiedlichen Fächern, um differentielle Aussagen darüber treffen zu können, in welchen Fächern räumliche Fähigkeiten von besonderer Bedeutung sind und in welchen weniger. Aufgrund der von Shea et al. (2001) dokumentierten durchschnittlichen räumlichen Fähigkeiten von Bachelor-Absolventinnen und -Absolventen der entsprechenden Fächer wurden in den vier untersuchten Studiengängen die stärksten Zusammenhänge in den drei MINT-Fächern erwartet – die stärksten hierbei im Bauingenieurwesen und in der Chemie sowie ein schwächerer, aber nach wie vor substanzieller Zusammenhang in der Biologie. In den Sozialwissenschaften als nicht-MINT-Studiengang wurde hingegen kein relevanter Zusammenhang zwischen räumlichen Fähigkeiten und Fachwissen erwartet, was die MINT-Spezifität räumlicher Fähigkeiten zeigen sollte. Das bis zum Ende des ersten Semesters erworbene Fachwissen stellte hier die abhängige Variable dar und wurde mittels parallel entwickelter Fachwissenstests vergleichbar erhoben.

Mit einer multiplen Regression wurde zuerst die Relevanz der Kontrollvariablen (fachliches Vorwissen, Abiturnote und Reasoning) und der räumlichen Visualisierungsfähigkeit als zentralem Prädiktor geprüft. Hierbei wurden mittels Interaktionen auch bereits Unterschiede zwischen den Fächern untersucht und anschließend in einer Mehrgruppen-Pfadanalyse die einzelnen fachspezifischen Effekte ebenfalls auf Signifikanz geprüft. Hier fiel bereits auf, dass Reasoning (als Indikator für *g*) in keinem Fach mehr von signifikanter Bedeutung war, sobald räumliche Visualisierungsfähigkeit ins Modell aufgenommen wurde. Da räumliche Visualisierungsfähigkeit auch unter Kontrolle der anderen Variablen weiterhin einen signifikanten Effekt auf das Fachwissen aufwies, zeigte diese Analyse bereits, dass räumliche Visualisierungsfähigkeit für den Fachwissenserwerb aller untersuchter Fächer von Bedeutung war – und zwar nicht nur aufgrund der dem verwendeten mentalen Rotationstest eigenen *g*-Komponente (siehe Studie 1), sondern auch durch die räumliche Komponente des Tests, da der mentale Rotationstest nicht nur dieselbe Varianz aufklärte wie der Reasoning-Test, sondern noch inkrementelle Varianz darüber hinaus. Auffällig war zudem, dass, entgegen der Hypothese, der Effekt räumlicher Visualisierungsfähigkeit auf das erworbene Fachwissen in allen erhobenen Fächern gleich war statt für die MINT-Fächer größer als für die Sozialwissenschaften. Insgesamt zeigte sich in der multiplen Regression sowie der dazugehörigen Multigruppen-Pfadanalyse, dass fachliches Vorwissen durchweg der beste Prädiktor für zum Ende des Semesters erworbenes Fachwissen war, wobei die Fächer hier betragsmäßig variierten. Zweitstärkster Prädiktor im Ingenieurwesen und den Sozialwissenschaften war die Abiturnote, die in Chemie und Biologie hingegen von geringerer Bedeutung war, weswegen dort räumliche Visualisierungsfähigkeit den zweitstärksten Prädiktor darstellte.

Im nächsten Schritt der Studie wurden die in diesen Modellen als relevant identifizierten Prädiktoren (also fachliches Vorwissen, Abiturnote und räumliche Visualisierungsfähigkeit) in ein Mediationsmodell aufgenommen, um die Annahme zu prüfen, dass fachliches Vorwissen und Abiturnote den Zusammenhang zwischen räumlicher Visualisierungsfähigkeit und Fachwissen mediieren könnten. Diese Annahme fußt auf früheren Befunden, die Zusammenhänge zwischen räumlichen Fähigkeiten und Schulleistungen in unterschiedlichen Fächern zeigten, beispielsweise in Mathematik (bspw. Cheng & Mix, 2014). Weiterhin wurden in diesem Mediationsmodell die aus Tests abgeleiteten Variablen (also räumliche Visualisierungsfähigkeit, Vorwissen und Fachwissen) anhand eines Single-Indicator-Modells latent modelliert (Hayduk & Littvay, 2012), um die gefundenen Zusammenhänge von Messfehlern zu bereinigen. Hierbei zeigten sich substanzielle partielle Mediationen des Effekts von räumlicher Visualisierungsfähigkeit auf bis zum Ende des Semesters erworbenes Fachwissen. Die indirekten Effekte variierten hierbei deutlich zwischen den Fächern, was die fachspezifischen Unterschiede hinsichtlich der Bedeutung fachlichen Vorwissens und der Abiturnote für den universitären Fachwissenserwerb sowie der Bedeutung räumlicher Fähigkeiten für den Vorwissenserwerb und die Abiturnote abbildet. Der direkte Effekt blieb hingegen weiterhin stabil zwischen den Fächern, was darauf hindeutet, dass räumliche Fähigkeiten fächerübergreifend den gleichen Effekt auf Fachwissen über Vorwissen und Abiturnote (inklusive mediierter Effekte) hinaus haben. Insgesamt klärte das Modell einen beträchtlichen Teil der Varianz im Fachwissen auf – von 89% in Chemie bis immerhin 45% in Sozialwissenschaften. Die größten Unterschiede lagen in der Mediation durch das Vorwissen – während die MINT-Fächer hier βs zwischen .24 und .35 aufwiesen, lag das β in den Sozialwissenschaften nur bei .06. Gemeinsam mit dem zwischen den Fächern gleichen direkten Effekt räumlicher Fähigkeiten auf das Fachwissen deutet dies darauf hin, dass die MINT-Spezifität räumlicher Fähigkeiten für fachliche Leistungen für vor dem Studium (also vorwiegend in der Schule) erworbenes Vorwissen deutlich stärker ausgeprägt ist (was sich in den variierenden indirekten Effekten bzw. dem der Mediation zugrunde liegenden Zusammenhang von räumlicher Visualisierungsfähigkeit und Vorwissen zeigte) als für an der Universität erworbenes Fachwissen (was durch den stabilen direkten Effekt abgebildet wird). Dies eröffnet die Frage, was universitäres Lernen in Hinblick auf die Nutzung räumlicher Fähigkeiten von schulischem Lernen unterscheidet – besonders in der durch die Sozialwissenschaften vertretenen Gruppe der nicht-MINT-Fächer, da bei diesen räumliche Fähigkeiten keinen Zusammenhang mit Vorwissen aufwiesen, für den Fachwissenserwerb im Studium dann aber an Bedeutung gewannen. Die Studie diskutierte zwei Fähigkeiten als mögliche Verbindungen: Die erste mögliche Verbindung stellen mathematische Fähigkeiten dar, für die hinlänglich belegt ist, dass sie eng mit räumlichen Fähigkeiten zusammenhängen (bspw. Judd & Klingberg, 2021). In den MINT-Fächern sind mathematische Fähigkeiten sowohl in der Schule (mit Abstrichen in der Biologie) als auch im Studium von Bedeutung, werden in den Sozialwissenschaften aber erst im Studium (vorwiegend in Form von Statistik) relevanter Teil des Curriculums. Als zweite mögliche Verbindung kommt visuelles Modellverständnis (in manchen Quellen auch „Repräsentationskompetenz“ genannt; Stieff et al., 2018) in Frage als die Fähigkeit, mit (meist fachbezogenen, gedruckten oder als physische Modelle vorliegenden) Visualisierungen umzugehen und zu lernen (Dickmann et al., 2019), deren Definition deutliche Ähnlichkeiten zur räumliche Visualisierungsfähigkeitsfacette räumlicher Fähigkeiten nach Carroll (1993) aufweist. Entsprechend hängt das Konstrukt des visuellen Modellverständnisses eng mit räumlichen Fähigkeiten im Allgemeinen und räumlicher Visualisierungsfähigkeit im Besonderen zusammen. In MINT-Fächern mit ihrer Vielzahl komplexer Visualisierungen wie Schaubildern und Diagrammen ist diese Fähigkeit als lernförderlich bekannt (bspw. Dickmann et al., 2019; Rau, 2017; Stieff et al., 2018). Es ist möglich, dass universitäre Lehre (im Vergleich zu schulischer Lehre) nun auch fächerübergreifend verstärkt komplexe Visualisierungen nutzt, was den über die untersuchten Fächer hinweg stabilen Zusammenhang zwischen räumlichen Fähigkeiten und erworbenem Fachwissen ebenfalls erklären könnte.

Insgesamt ist der Befund zur Bedeutung räumlicher Fähigkeiten für fachliche Leistung konsistent mit früheren Befunden in den einzelnen untersuchten MINT-Fächern (Ingenieurwesen bspw. Sorby et al., 2018; Chemie bspw. Stieff et al., 2014; Biologie bspw. Russell-Gebbett, 1985). Die vorliegende Studie ermöglichte nun erstmalig den direkten Vergleich dieser fachspezifischen Zusammenhänge, zudem unter Kontrolle weiterer relevanter Kontrollvariablen. Neu ist die Abgrenzung zu einem nicht-MINT-Fach, bei dem sich zeigte, dass räumliche Visualisierungsfähigkeit für den Fachwissenserwerb an der Universität ebenfalls von Bedeutung ist. Der Einbezug von Mediationseffekten ermöglichte weiterhin einen Vergleich von universitärem und vorwiegend in der Schule stattfindendem Fachwissenserwerb. Dies eröffnet neue Forschungsrichtungen zur Untersuchung von Unterschieden in Hinblick auf die Bedeutung räumlicher Fähigkeiten zwischen schulischem und universitärem Lernen. Zudem eröffnete der durchgeführte Fächervergleich Einblicke darin, wo sich aufgrund starker Zusammenhänge mit Fachwissenserwerb räumliche Trainings zur Förderung besonders lohnen würden: An der Universität vermutlich fächerübergreifend, in der Schule könnten vor allem die MINT-Fächer von räumlichen Trainings profitieren, beispielsweise auch durch stärker räumlich gestaltete Kursmaterialien im Sinne eines Kurstrainings nach Uttal et al. (2013), wie es in einigen (vorwiegend universitären) Studien bereits angewandt wurde (bspw. Bruce & Hawes, 2015; Grzybowski et al., 2014; Sorby et al., 2018; Walton, S., Patrick et al., 2015). Die gezielte Förderung räumlicher Fähigkeiten (gerade im Schulkontext) könnte auch zur Förderung des Interesses an MINT-Fächern im Sinne der Fähigkeits-Präferenz-Passung (Webb et al., 2007) beitragen.

5.1.3 Studie 3

Im Kern der dritten Studie stand die Evaluation eines videospielbasierten Trainings räumlicher Fähigkeiten zur Verbesserung des Fachwissenserwerbs bei Bauingenieurs-Erstsemesterstudierenden. Videospieltrainings wurden bisher in der Forschung als Interventionsmöglichkeit vernachlässigt, obwohl die Metaanalyse von Uttal et al. (2013) ihre Effektivität zur Verbesserung räumlicher Fähigkeiten eindeutig belegte. Frühere Studien, so auch die in der vorliegenden Arbeit vorgestellte Studie 2, belegten die Relevanz räumlicher Fähigkeiten für den Fachwissenserwerb in ingenieurwissenschaftlichen Fächern (für ein Review siehe Ha & Fang, 2016); Sorby et al. (2018) belegten auch bereits den kausalen Nutzen von (kursbasierten) Trainings in diesem Kontext.

Das achtstündige Videospieltraining führte insgesamt zu einer signifikanten Verbesserung von räumlicher Visualisierungsfähigkeit. Die räumliche Visualisierungsfähigkeit der Teilnehmenden reduzierte sich zudem bis zur verzögerten erneuten Messung am Ende des Semesters nicht signifikant, was einerseits darauf hindeutete, dass der Trainingseffekt grundsätzlich zeitlich stabil war und andererseits tatsächlich ein Trainingseffekt durch die Videospiele vorlag statt nur ein Retest-Effekt durch Übung mit den Testverfahren (da sonst eine zusätzliche Verbesserung von Posttest zu verzögertem Posttest zu erwarten gewesen wäre). Der Trainingseffekt ging vorrangig auf die Gruppe zurück, die das räumlich anspruchsvolle Spiel „Portal 2“ spielte: Nur in dieser Gruppe war der Zuwachs räumlicher Visualisierungsfähigkeit signifikant, während sich in der Gruppe, die das räumlich simplere Spiel „Sherlock Holmes jagt Jack the Ripper“ spielte, keine signifikante Verbesserung zeigte. Allerdings ließ sich dieser Unterschied nur explorativ zeigen, da der Unterschied im Trainingszuwachs zwischen den Gruppen nicht stark genug war, um die Interaktion von Gruppe und Messzeitpunkt in der mixed RM-ANOVA signifikant werden zu lassen (trotz zufriedenstellender Power von .79), weswegen diese Post-hoc-Tests nur explorativ aufgrund ihrer Hypothesenrelevanz durchgeführt wurden. Somit ließ sich der Vorteil des räumlich anspruchsvolleren Videospiels gegenüber dem weniger räumlich anspruchsvollen Vergleichsspiel nicht eindeutig belegen. Der signifikante gruppenübergreifende Prä-Post-Unterschied deutet allerdings darauf hin, dass generell das Spielen eines in einer 3D-Welt stattfindenden Videospiels bereits die räumliche Visualisierungsfähigkeit trainierte, obwohl die Dreidimensionalität der Spielwelt bei Portal 2 zentrales Spielelement, beim Sherlock Holmes-Spiel für den eigentlichen Spielablauf hingegen von geringer Relevanz war.

Im zweiten Schritt der Analyse wurde auch in dieser Stichprobe der Effekt räumlicher Visualisierungsfähigkeit auf den Fachwissenserwerb geprüft, der sich hier in ähnlichem Maße zeigte wie bereits in Studie 2 für die dortigen Bauingenieursstudierenden. Eine Besonderheit der Analyse in Studie 3 ist darüber hinaus der Einbezug der Klausurnoten der Einführungsvorlesung „Technische Mechanik“, deren Inhalt maßgeblich im verwendeten Fachwissenstest enthalten war. Hierbei wurde gezeigt, dass die trainierte räumliche Visualisierungsfähigkeit auch auf die Klausurnote einen signifikanten Effekt hatte, dieser allerdings vollständig durch das bis zum Ende des Semesters erworbene Fachwissen mediiert wurde: Das Training von räumlicher Visualisierungsfähigkeit verbesserte also die erreichte Klausurnoten explizit durch die Unterstützung des vorherigen Fachwissenserwerbs. Auf Basis des totalen Effekts räumlicher Fähigkeiten auf die Klausurnote aus dem Mediationsmodell (das, neben dem Fachwissen als zentralem Mediator, zudem fachliches Vorwissen und Reasoning als mit räumlicher Visualisierungsfähigkeit korrelierende Prädiktoren enthielt) als Regressionskoeffizienten und dem trainingsbedingten Zuwachs der räumlichen Visualisierungsfähigkeit als Prädiktor ergab sich eine Verbesserung der Klausurnote um vier Prozentpunkte in der Portal 2-Gruppe und zwei Prozentpunkte in der Sherlock Holmes-Gruppe. Dies entsprach einer nicht zu vernachlässigenden Verbesserung um knapp eine bzw. eine halbe Drittelnote in den beiden Gruppen durch acht Stunden nicht fachbezogenes Videospieltraining, das von den Teilnehmenden durchschnittlich als neutral bis unterhaltsam bewertet wurde.

Abschließend wurde vorherige Videospielerfahrung als möglicher Einflussfaktor auf den Trainingserfolg untersucht, da einzelne Probanden aufgrund mangelnder Vorerfahrung Schwierigkeiten mit der Steuerung ihres Trainingsspiels Portal 2 berichteten. Hierbei zeigten sich in keiner Gruppe signifikante Unterschiede im Zuwachs der räumlichen Visualisierungsfähigkeit zwischen Teilnehmenden mit und ohne vorherige Videospielerfahrung. Allerdings schloss das Konfidenzintervall der gemessenen Effektstärke des Unterschieds in der Portal 2-Gruppe 0 nicht ein, was darauf hindeutet, dass hier möglicherweise doch ein Unterschied vorlag, der nur aufgrund geringer Power (die für diesen Vergleich bei .51 lag) knapp nicht signifikant wurde. Somit lässt sich die Relevanz von vorheriger Videospielerfahrung für den Nutzen eines videospielbasierten Trainings nicht eindeutig belegen. Allerdings bietet die vorliegende Studie erste Hinweise darauf, dass zumindest für Spiele mit komplexerer Steuerung wie Portal 2 Videospielvorerfahrung eine Rolle spielen könnte. Da sich die Videospielvorerfahrung mit einem einfachen Selbstbericht aufwandsarm erheben lässt, erscheint die Aufnahme eines solchen Items als zu prüfende Kontrollvariable in die Demografie- bzw. Screening-Fragebögen von Videospieltrainingsstudien empfehlenswert.

5.2 Limitationen und Ausblick

Alle drei Studien in der vorliegenden Arbeit weisen einige Limitationen auf, die bei der Interpretation ihrer Ergebnisse sowie bei der Durchführung von Folgestudien berücksichtigt werden sollten.

5.2.1 Studie 1

Zentrale Limitation der ersten Studie war, dass im Vandenberg und Kuse-Paradigma die einzelnen Zielstimuli nicht unabhängig voneinander bearbeitet werden konnten. Einerseits war der zu rotierende Referenzstimulus innerhalb eines Items konstant, egal mit welchem der vier zu beurteilenden Vergleichsstimuli des Items er verglichen wurde. Andererseits war die Anzahl korrekter Vergleichsstimuli im Vorhinein bekannt –dadurch ist auch Auswahl per Ausschluss möglich: Wenn Teilnehmende beispielsweise einen korrekten und zwei falsche Vergleichsstimulus bereits identifiziert hatten, konnten sie daraus ableiten, dass der vierte ebenfalls korrekt sein musste, ohne ihn mittels mentaler Rotation prüfen zu müssen (vorausgesetzt, sie vertrauten ihren vorherigen Rotationsprüfungen). Durch die dadurch vorliegende Abhängigkeit der einzelnen bearbeiteten Stimuli voneinander könnte der Zusammenhang der beiden Rotationsfaktoren (allgemein/kardinal und spezifisch/schief) vergrößert worden sein. Auch wenn nicht davon auszugehen ist, dass nur hierdurch die Faktoren zur psychometrischen Überlappung gebracht wurden, sollten zukünftige Studien zur Untersuchung von Itemcharakteristika in mentalen Rotationstests diese Problematik berücksichtigen und die zum Abgleich benötigten Rotationen aller Stimuli (korrekte wie Distraktoren) innerhalb eines Items gleich halten – also beispielsweise nur kardinale oder nur schiefe Rotationen. Die Variation der Anzahl korrekter Stimuli pro Item (0-4 statt immer genau 2) könnte hingegen ebenso dabei helfen, die Nutzung von Ausschlussstrategien zu reduzieren, wie die ausschließliche Verwendung gespiegelter Distraktoren, da diese keine strukturellen Unterschiede zum Referenzstimulus aufweisen und daher eher räumliche Lösungsstrategien verwendet werden müssen (auch wenn der Abgleich von einzelnen Charakteristika wie Knicke in unterschiedliche Richtungen nach wie vor genutzt werden könnten, allerdings komplizierter sind als beispielsweise das Zählen von Segmentlängen). Die hohen internen Konsistenzen der Rotationsfaktoren in dieser Studie deuten allerdings darauf hin, dass nicht von variierenden Strategien innerhalb der Items auszugehen ist. Dass diese Form des mentalen Rotationstest tatsächlich räumliche Visualisierungsfähigkeit misst, wird zusätzlich unterstützt durch einerseits die Vorhersage der Testleistung durch den spezifischen räumliche Visualisierungsfähigkeitsfaktor in dieser Studie über *g* hinaus und andererseits durch die Prädiktionskraft der Testleistung für die Vorhersage von erworbenem Fachwissen über Reasoning hinaus, die in Studie 2 und Studie 3 gezeigt wurde.

5.2.2 Studie 2

Eine große Stärke der zweiten Studie ist gleichzeitig auch eine Schwäche: Pro Fach sind mehrere (3-4) Universitäten und teilweise auch zwei Jahrgänge erhoben worden. Einerseits stärkt das die Verallgemeinerbarkeit der Ergebnisse (gerade im Vergleich zu früheren Studien, die typischerweise ein Fach an einem Standort erhoben haben), aber andererseits war nicht in allen Kohorten die Repräsentativität der Stichprobe für Studierende dieses Fachs, Jahrgangs und Standorts gewährleistet. Die Stichproben einzelner Kohorten waren teilweise sehr klein (besonders im Bauingenieurwesen war dies ein Problem), sodass in der vorliegenden Studie die einzelnen erhobenen Kohorten der Fächer zur weiteren Analyse zusammengelegt werden mussten, ohne dass umfassendere Kohortenvergleiche möglich gewesen wären. Die Erfahrungen aus dieser Studie haben klar gezeigt, dass diejenigen Datenerhebungen, die im Rahmen regulärer Lehrveranstaltungen durchgeführt wurden, deutlich höhere Ausschöpfungsquoten erreichten und entsprechend repräsentativer für die jeweilige Kohorte waren als die Datenerhebungen, die in eigenen Zeitslots abseits regulärer Veranstaltungen stattfanden. Zukünftige Studien sollten daher großen Wert auf die Unterstützung lokaler Lehrender legen, die Datenerhebungen als Teil regulärer Lehrveranstaltungen zu ermöglichen, statt nur freie Zeitslots im Stundenplan zur Erhebung zu nutzen, um Stichproben so groß und repräsentativ wie möglich zu erhalten.

Weiterhin ist anzumerken, dass die zeitgleiche Messung von räumlicher Visualisierungsfähigkeit und Vorwissen formal keine kausalen Schlüsse zulässt. Gerade unter Berücksichtigung der Tatsache, dass räumliche Fähigkeiten grundsätzlich veränderbar sind (bspw. durch räumlich anspruchsvolle Kurse) ist es möglich, dass die durchgeführte Messung räumlicher Fähigkeiten nicht akkurat das tatsächliche Niveau abbildet, das während des Vorwissenserwerbs vorlag. Da sich aber weder in dieser Studie noch in der dritten Studie Verbesserungen räumlicher Fähigkeiten ohne gezieltes Training zeigten, ist davon auszugehen, dass reguläre schulische und universitäre Lehre zwar räumliche Fähigkeiten nutzen mag, diese aber nicht per se auch trainiert. Daraus kann abgeleitet werden, dass die hier durchgeführte Messung zumindest näherungsweise auch das früher vorliegende Niveau räumlicher Fähigkeiten abbildete und somit eine Vorhersage des zeitgleich erhobenen, aber früher erworbenen Vorwissens gerechtfertigt werden kann. Nichtsdestotrotz wäre es wünschenswert, diese Annahmen zur (interventionsfreien) Entwicklung räumlicher Fähigkeiten vor allem in der Schule, aber auch im Studium längsschnittlich zu untersuchen, um die Zusammenhänge räumlicher Fähigkeiten mit fachspezifischen Leistungen zu unterschiedlichen Zeitpunkten der Bildungslaufbahn zu untersuchen, beispielsweise mittels einer Cross-Lagged-Panel-Studie.

5.2.3 Studie 3

Die größte Limitation der dritten Studie ist der nicht hypothesenkonforme Interventionscheck. Es konnte zwar gezeigt werden, dass es insgesamt einen Zuwachs während der Trainingsperiode gab, aber keine Veränderung über den nicht mehr intervenierten Rest des Semesters, was auf einen generellen Trainingseffekt durch 3D-Videospiele hinweist. Allerdings konnte nicht gezeigt werden, dass die als Experimentalgruppe angelegte Gruppe mit dem räumlich anspruchsvolleren Spiel Portal 2 einen signifikant höheren Trainingseffekt aufweist als die Kontrollgruppe mit dem räumlich weniger anspruchsvollen Sherlock Holmes-Spiel – die erwartete Interaktion von Gruppe und Messzeitpunkt wurde, trotz ausreichender Power, nicht signifikant. Hauptgrund hierfür ist vermutlich die zu strikte Kontrollbedingung anhand eines weiteren 3D-Spiels – frühere Studien, die signifikante Unterschiede im Trainingseffekt fanden, verwendeten typischerweise 2D-Spiele als Kontrollbedingung (bspw. Shute et al., 2015). Zukünftige Studien sollten bei der Auswahl der Trainingsspiele weitere Eigenschaften der untersuchten Spiele berücksichtigen, die Einfluss auf die zu erwartenden Effektstärken(-unterschiede) haben könnten, um die Stichprobenplanung entsprechend anzulegen. Studien, die gezielt Eigenschaften von Spielen variieren, könnten dabei helfen, bei zukünftigen Studien eine informierte Entscheidung für wünschenswerte Eigenschaften der Interventions- und Kontrollspiele treffen zu können.

Weiterhin brachten die Nutzung privater Endgeräte für die Intervention sowie der Einschluss von Teilnehmenden mit vorherigen Videospielerfahrungen Probleme bei der Randomisierung der Gruppen mit sich, da hierdurch harte Kriterien entstanden, die eine Zuordnung zur einen oder anderen Gruppe erzwangen. Aus methodischer Perspektive wäre es wünschenswert, allen Teilnehmenden für die Intervention nutzbare Endgeräte zur Verfügung zu stellen, die Teilnehmenden in festen Zeitintervallen unter kontrollierten Bedingungen spielen zu lassen und nur Teilnehmende ohne Erfahrung mit den Trainingsspielen aller Gruppen zuzulassen – oder gleich mit gar keiner vorherigen Videospielerfahrung. Diese Kriterien würden allerdings die ökologische Validität der Intervention enorm untergraben, die für die mögliche Implementierung in die Praxis von großer Bedeutung wäre. Gerade der Ausschluss von Personen mit Videospielvorerfahrung würde zudem die mögliche Stichprobe drastisch reduzieren – nicht nur schlösse man dadurch einen sehr großen Anteil der potenziell in Frage kommenden Teilnehmenden aus, sondern auch noch genau den Anteil, der an einer solchen Studie besonders teilnahmeinteressiert wäre. Entsprechend bleibt die Entscheidung über Einschlusskriterien für eine videospielbasierte Trainingsstudie ein Abwägen zwischen Kontrollierbarkeit und ökologischer Validität.

Außerdem gibt es noch keinen Goldstandard für die notwendige Dauer und Staffelung eines Videospieltrainings, um messbare und stabile Verbesserungen räumlicher Fähigkeiten hervorzurufen. Bisherige Studien nutzten unterschiedliche Dauern, die von 45 Minuten (Adams & Mayer, 2012) bis

12 Stunden (Cohen et al., 2007) reichten. Die vorliegende Studie verwendete eine Trainingsdauer von 8 Stunden, da eine frühere Studie von Shute et al. (2015) für dasselbe Trainingsspiel Portal 2 zeigen konnte, dass diese Dauer zu einem signifikanten Trainingseffekt auf räumliche Fähigkeiten führt. Es ist allerdings möglich, dass eine längere Dauer oder eine andere Staffelung (bspw. regelmäßige Intervalle statt freier Spielzeiteinteilung über einen kürzeren oder längeren Zeitraum als in der vorliegenden Studie) zu stärkeren oder stabileren Trainingseffekten führen könnte. Für zukünftige Videospieltrainingsstudien wäre eine systematische Evaluation unterschiedlicher Trainingsdauern und -staffelungen von großem Interesse.

5.3 Fazit

Die vorliegende Arbeit hatte das Ziel, mehrere Lücken in der Forschung zum Zusammenhang räumlicher Fähigkeiten mit und dem Potenzial räumlicher Fähigkeiten für den Studienerfolg insbesondere in MINT-Fächern zu füllen.

Die erste herausgearbeitete Lücke war die mangelnde Differenzierbarkeit zwischen räumlichen Fähigkeiten und *g* in bisherigen Studien, obwohl die Überlappung der beiden Konstrukte hinlänglich bekannt ist. Die vorliegende Arbeit konnte sowohl mittels einer psychometrischen Analyse eines typischen Tests (Studie 1) als auch durch Kontrolle von Reasoning als Indikator für *g* in Analysen des Zusammenhangs der Facette räumlicher Visualisierungsfähigkeit der räumlichen Fähigkeiten mit im ersten Semester erworbenem Fachwissen unterschiedlicher Fächer sowie Klausurnoten klar zeigen, dass räumliche Fähigkeiten im Allgemeinen und räumlicher Visualisierungsfähigkeit im Besonderen für den Fachwissenserwerb in vier exemplarischen Fächern (die drei MINT-Fächer Bauingenieurwesen, Chemie und Biologie sowie das nicht-MINT-Fach Sozialwissenschaften; Studie 2) sowie für Klausurnoten (exemplarisch im Fach Bauingenieurwesen; Studie 3) auch unter Kontrolle von *g* von Bedeutung sind.

Die zweite Lücke war der Mangel an Vergleichbarkeit der Befunde zum Zusammenhang zwischen räumlicher Visualisierungsfähigkeit und fachlichen Leistungen zwischen unterschiedlichen Fächern, die eine differentielle Aussage dazu, in welchen Fächern räumliche Fähigkeiten von größerer Bedeutung sind als in anderen, nicht zuließen. Auch dieser Lücke wurde in der vorliegenden Arbeit exemplarisch anhand dreier MINT-Fächer und einem nicht-MINT-Fach begegnet (Studie 2). Dabei zeigte sich die auf Basis der von Shea et al. (2001) dargestellten durchschnittlichen räumlichen Fähigkeiten unterschiedlicher Fächer erwartete Rangfolge in Bezug auf die Stärke des Zusammenhangs räumlicher Fähigkeiten mit fachlicher Leistung: die Ingenieurwissenschaften und *physical sciences* (anhand der Chemie) an der Spitze, die Biologie mit etwas Abstand dahinter und die Sozialwissenschaften als nicht-MINT-Fach als Schlusslicht. Zudem wurde eine interessante Perspektive für zukünftige Forschung eröffnet, da die hier vorgestellte Studie Hinweise darauf fand, dass sich schulisches und universitäres Lernen in den einzelnen Fächern in Hinblick auf die Bedeutung räumlicher Fähigkeiten unterscheiden könnte.

Die dritte Lücke stellte die Vernachlässigung von Videospieltrainings zur Verbesserung räumlicher Fähigkeiten im Anwendungskontext dar, während die Effektivität der beiden anderen Trainingstypen nach Uttal et al. (2013), der Übung räumlicher Aufgaben und kursbasierten Trainings, im MINT-Kontext bereits gezeigt wurde. Die in der vorliegenden dritten Studie verwendeten Videospiele führten zu einem signifikanten Trainingseffekt, der sich zwischen den beiden Gruppen nicht unterschied. Die trainierten räumlichen Fähigkeiten standen auch in dieser Studie in engem Zusammenhang mit sowohl dem Fachwissenserwerb als auch den Klausurnoten der Teilnehmenden. Entsprechend lässt sich sagen, dass videospielbasierte Trainings räumlicher Fähigkeiten im MINT-Kontext Potenzial zeigen, auch wenn weitere Forschung notwendig ist, um die hierfür geeignetsten Spiele und Trainingsstrukturen herauszuarbeiten.

Insgesamt konnte die vorliegende Arbeit die drei Kernfragestellungen beantworten:

1. Räumliche Fähigkeiten haben besondere Prädiktionskraft über *g* (meist operationalisiert durch Reasoning) für sowohl die Testleistung im mentalen Rotationstest als auch für den Fachwissenserwerb und Klausurerfolg in unterschiedlichen Fächern.
2. Der Zusammenhang von räumlichen Fähigkeiten mit fachspezifischen Leistungsmaßen ist in MINT-Fächern stärker als in nicht-MINT-Fächern, wobei es innerhalb der MINT-Fächer weitere Abstufungen gab. Fachunterschiede zeigten sich vor allem im Zusammenhang mit Vorwissen.
3. Videospieltraining war generell effektiv zur Verbesserung räumlicher Fähigkeiten und stand im Zusammenhang mit besserem Fachwissenserwerb und besseren Klausurnoten. Es zeigten sich allerdings keine signifikanten Unterschiede im erzielten Trainingseffekt zwischen den verwendeten Spielen, obwohl das eine Spiel räumlich deutlich anspruchsvoller war als das andere.

Somit leistet diese Arbeit einen substanziellen Beitrag zum Verständnis der Bedeutung räumlicher Fähigkeiten für den Studienerfolg, besonders in MINT-Fächern.

5.4 Literaturverzeichnis

Adams, D. M. & Mayer, R. E. (2012). Examining the connection between dynamic and static spatial skills and video game performance. In N. Miyake, D. Peebles & R. P. Cooper (Hrsg.), *34th Annual meeting of the Cognitive Science Society 2012 (CogSci 2012): Building bridges across cognitive sciences around the world* (Vol. 34, S. 1254–1259). Curran Associates, Inc.

Bruce, C. D. & Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: What is it? Why does it matter? And what can we do about it? *ZDM Mathematics Education*, *47*(3), 331–343. https://doi.org/10.1007/s11858-014-0637-4

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Cheng, Y.-L. & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*, 2–11. https://doi.org/10.1080/15248372.2012.725186

Cohen, J., Green, C. S. & Bavelier, D. (2007). Training visual attention with video games: Not all games are created equal. In H. F. O'Neil & R. Perez (Hrsg.), *Computer games and team and individual learning* (Bd. 9, S. 205–227).

Dickmann, T., Opfermann, M., Dammann, E., Lang, M. & Rumann, S. (2019). What you see is what you learn? The role of visual model comprehension for academic success in chemistry. *Chemistry Education Research and Practice*, *20*(4), 804–820. https://doi.org/10.1039/C9RP00016J

Grzybowski, D. M., Stavridis, O., Barclay, L. A., Abrams, L., Sorby, S. A., Thomas, J. & Merril, J. A. (2014). Impact of optional supplemental course to enhance spatial visualization skills in first-year engineering students. In American Society for Engineering Education (Vorsitz), *2014 ASEE Annual Conference*.

Ha, O. & Fang, N. (2016). Spatial ability in learning engineering mechanics: critical review. *Journal of Professional Issues in Engineering Education and Practice*, *142*(2), 4015014. https://doi.org/10.1061/(ASCE)EI.1943-5541.0000266

Hayduk, L. A. & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*(1), 159. https://doi.org/10.1186/1471-2288-12-159

Judd, N. & Klingberg, T. (2021). Training spatial cognition enhances mathematical learning in a randomized study of 17,000 children. *Nature Human Behaviour.* Vorab-Onlinepublikation. https://doi.org/10.1038/s41562-021-01118-4

Maquet, L., Dunbar, R., Buckley, J., Seery, N. & Sorby, S. A. (2022). A review of the literature to inform the efficacy of a spatial skills intervention for secondary level STEM education. In D. Gill, J. Tuff, T. Kennedy, S. Pendergast & S. Jamil (Hrsg.), *PATT39 Conference 2022. PATT on the edge: Technology, innovation and education* (S. 496–507).

Rau, M. A. (2017). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, *29*(4), 717–761. https://doi.org/10.1007/s10648-016-9365-3

Russell-Gebbett, J. (1985). Skills and strategies – pupils' approaches to three-dimensional problems in biology. *Journal of Biological Education*, *19*(4), 293–298. https://doi.org/10.1080/00219266.1985.9654755

Shea, D. L., Lubinski, D. & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*(3), 604–614. https://doi.org/10.1037//0022-0663.93.3.604

Shute, V. J., Ventura, M. & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, *80*, 58–67. https://doi.org/10.1016/j.compedu.2014.08.013

Sorby, S. A., Veurink, N. & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences*, *67*, 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Stieff, M., Dixon, B. L., Ryu, M., Kumi, B. C. & Hegarty, M. (2014). Strategy training eliminates sex differences in spatial problem solving in a stem domain. *Journal of Educational Psychology*, *106*(2), 390–402. https://doi.org/10.1037/a0034823

Stieff, M., Origenes, A., DeSutter, D., Lira, M., Banevicius, L., Tabang, D. & Cabel, G. (2018). Operational constraints on the mental rotation of STEM representations. *Journal of Educational Psychology*, *110*(8), 1160–1174. https://doi.org/10.1037/edu0000258

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C. & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Vandenberg, S. G. & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Walton, S., Patrick, Urban-Lurain, M., Idema, A., Hinds, Timothy, J. & Briedis, D. (2015). Spatial visualization skills intervention for first-year engineering students: everyone's a winner! In American Society for Engineering Education (Vorsitz), *2015 ASEE Annual Conference & Exposition*.

Webb, R. M., Lubinski, D. & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420. https://doi.org/10.1037/0022-0663.99.2.397

6 Anhang

6.1 Anhang Studie 1

6.1.1 Descriptive Data for the Mental Rotation Test Items

Table A6.1

Means and Standard Deviations of the Mental Rotation Task Items

Item	Simple rotation		Complex rotation		Classic scoring	
	M	*SD*	*M*	*SD*	*M*	*SD*
MRT_01	0.90	0.31	0.88	0.32	0.8	0.4
MRT_02	0.90	0.30	0.89	0.31	0.83	0.38
MRT_03	0.78	0.41	0.84	0.37	0.65	0.48
MRT_04	0.91	0.29	0.51	0.5	0.45	0.5
MRT_05	0.82	0.39	0.85	0.36	0.69	0.46
MRT_06	0.72	0.45	0.62	0.49	0.39	0.49
MRT_07	0.73	0.44	0.62	0.49	0.42	0.49
MRT_08	0.60	0.49	0.67	0.47	0.38	0.49
MRT_09	0.70	0.46	0.55	0.5	0.36	0.48
MRT_10	0.74	0.44	0.76	0.43	0.53	0.5
MRT_11	0.76	0.43	0.87	0.34	0.65	0.48
MRT_12	0.82	0.38	0.89	0.31	0.74	0.44
MRT_13	0.88	0.32	0.84	0.36	0.76	0.43
MRT_14	0.88	0.32	0.84	0.37	0.77	0.42
MRT_15	0.70	0.46	0.76	0.43	0.52	0.5
MRT_16	0.45	0.50	0.82	0.38	0.36	0.48
MRT_17	0.78	0.42	0.85	0.36	0.66	0.47
MRT_18	0.65	0.48	0.79	0.41	0.49	0.5
MRT_19	0.53	0.50	0.58	0.49	0.24	0.43
MRT_20	0.77	0.42	0.72	0.45	0.56	0.5
MRT_21	0.75	0.43	0.48	0.5	0.29	0.46
MRT_22	0.80	0.40	0.78	0.41	0.62	0.49
MRT_23	0.81	0.39	0.79	0.41	0.63	0.48
MRT_24	0.91	0.29	0.85	0.35	0.78	0.41

6.1.2 Item Selection Process for the Mental Rotation Test

Rules for item selection were adapted from Krüger and Suchan (2016). Of the 16 different objects included in the stimulus library (Peters & Battista, 2008), only objects 3 to 14 were used as the object pairs 1+2 and 15+16 were known to be identical (but rotated) versions of each other and therefore problematic if used as structure foils for each other (see below). Two separate selection rules were used to create the present item set. Rule 1 was used for items 1-6 and 19-24, while rule 2 was used for items 7-18. If the calculated angular disparity exceeded 360°, it continuous further from 0° again. All angular disparities are given in comparison to the 0° default stimulus, not the reference object, as this agnostic naming scheme enabled picking them from the library via R syntax.

6.1.2.1 Rule 1: Items 1-6 and 19-24

- Reference stimulus: The reference stimulus was rotated around the x axis by 15° * item number + 50° (e.g., 95° for item 3).
- Target 1: The first target was rotated around the x axis by 15° * item number + 175° (e.g., 220° for item 3)
- Target 2: The second target was rotated around the z axis by 15° * item number + 300° (e.g., 345° for item 3). Thereby, target 1 differs from the reference by rotation around the x axis back to the 0° default position and then further around the z axis.
- Distractor 1: The first distractor is a mirrored version of the 0° target object that was then rotated around the x axis by 15° * item number + 50° (e.g., 95° for item 3). This resulted in a mirror foil that is similar to a mirror image of the item's reference object but not identical to it.
- Distractor 2: The second distractor consisted of the next object (e.g., object 7 when the reference stimulus and the targets are rotated versions of object 6) that was rotated around the z axis by 15° * item number + 100° (e.g., 145° for item 3), resulting in a structure foil.

6.1.2.2 Rule 2: Items 7-18

- Reference stimulus: The reference stimulus was rotated around the z axis by 15° * item number + 50° (e.g., 170° for item 8).
- Target 1: The first target was rotated around the z axis by 15° * item number + 175° (e.g., 295° for item 8)
- Target 2: The second target was rotated around the x axis by 15° * item number + 300° (e.g., 420° => 60° for item 3). Thereby, target 1 differs from the reference by rotation around the z axis back to the 0° default position and then further around the x axis.

- Distractor 1: The first distractor is a mirrored version of the 0° target object that was then rotated around the z axis by 15° * item number + 100° (e.g., 220° for item 8). This resulted in a mirror foil that is similar to a mirror image of the item's reference object but not identical to it.
- Distractor 2: The second distractor consisted of the next object (e.g., object 11 when the reference stimulus and the targets are rotated versions of object 10) that was rotated around the x axis by 15° * item number + 100° (e.g., 220° for item 8), resulting in a structure foil.

After test set creation we noticed that the objects 9+10 were also identical but rotated versions of each other. We therefore used object 11 as for the second distractor in the items using object 9 as reference and object 10 as structure foil (items 7+19).

References

Krüger, J. K., & Suchan, B. (2016). You should be the specialist! Weak mental rotation performance in aviation security screeners - reduced performance level in aviation security with no gender effect. *Frontiers in Psychology*, *7*, 333. https://doi.org/10.3389/fpsyg.2016.00333

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260–264. https://doi.org/10.1016/j.bandc.2007.09.003

6.1.3 Detailed Results

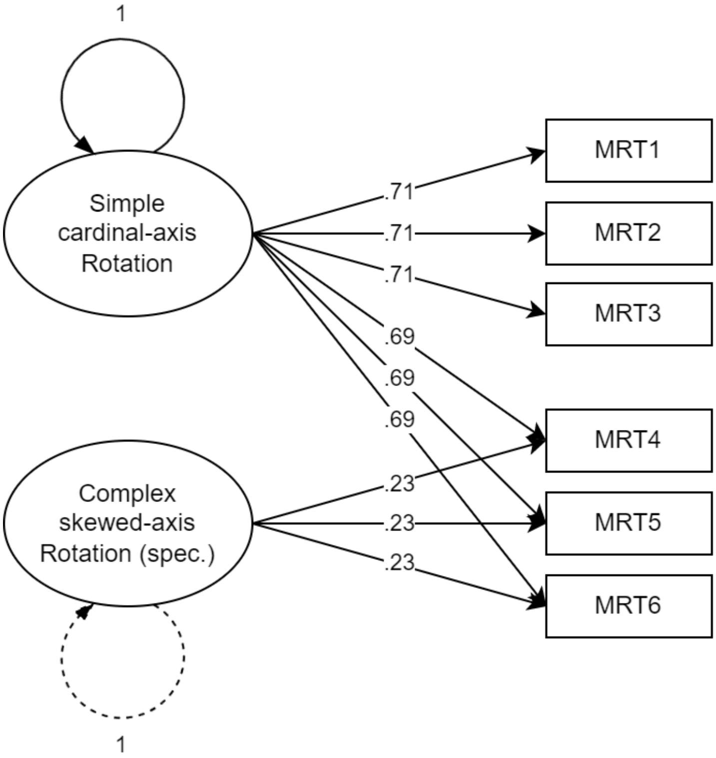
Table A6.2

Correlation Table of Tests and Test Parcels Used in Analysis

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1.	Odd One Out	-	0.47	0.50	0.40	0.17	0.48	0.39	0.44	0.28	0.34	0.31	0.31	0.40	0.33
2.	Matrices		-	0.51	0.29	0.16	0.40	0.37	0.36	0.33	0.30	0.24	0.31	0.28	0.26
3.	Figure classifikation			-	0.53	0.34	0.54	0.50	0.46	0.40	0.45	0.30	0.38	0.37	0.35
4.	Paper folding				-	0.37	0.49	0.42	0.47	0.31	0.40	0.29	0.35	0.34	0.43
5.	Cube construction					-	0.38	0.37	0.33	0.33	0.36	0.16	0.26	0.25	0.28
6.	Rotation total						-	0.89	0.90	0.71	0.75	0.73	0.76	0.74	0.72
7.	Rotation cardinal							-	0.73	0.82	0.83	0.80	0.62	0.59	0.58
8.	Rotation skewed								-	0.56	0.61	0.62	0.84	0.82	0.82
9.	Rotation cardinal parcel1									-	0.53	0.48	0.52	0.42	0.44
10.	Rotation cardinal parcel2										-	0.51	0.53	0.49	0.49
11.	Rotation cardinal parcel3											-	0.49	0.55	0.51
12	Rotation skewed parcel4												-	0.52	0.54
13.	Rotation skewed parcel5													-	0.51
14.	Rotation skewed parcel6														-

Figure A6.1

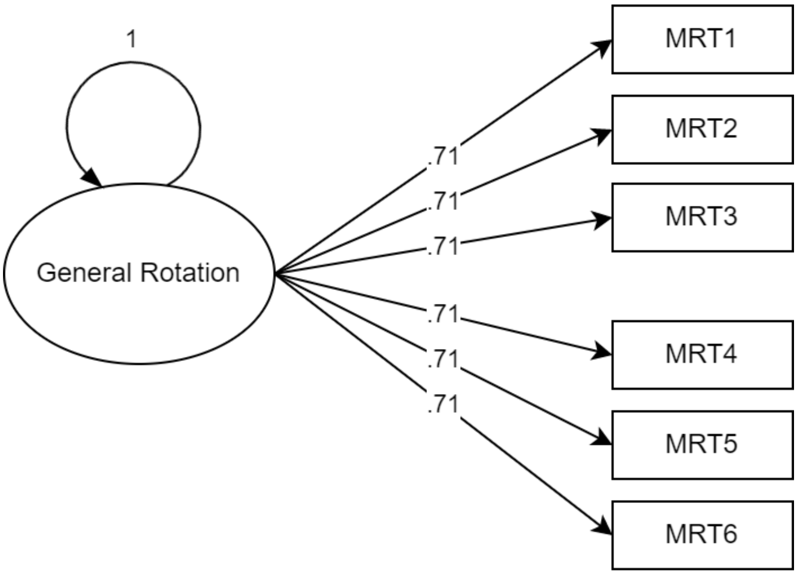
Nested-factor Measurement Model of Mental Rotation Ability



Note. MRT1-6 = Parcels of items from the mental rotation test. Solid lines depict significant loadings and variances while dashed lines depict non-significant loadings and variances.

Figure A6.2

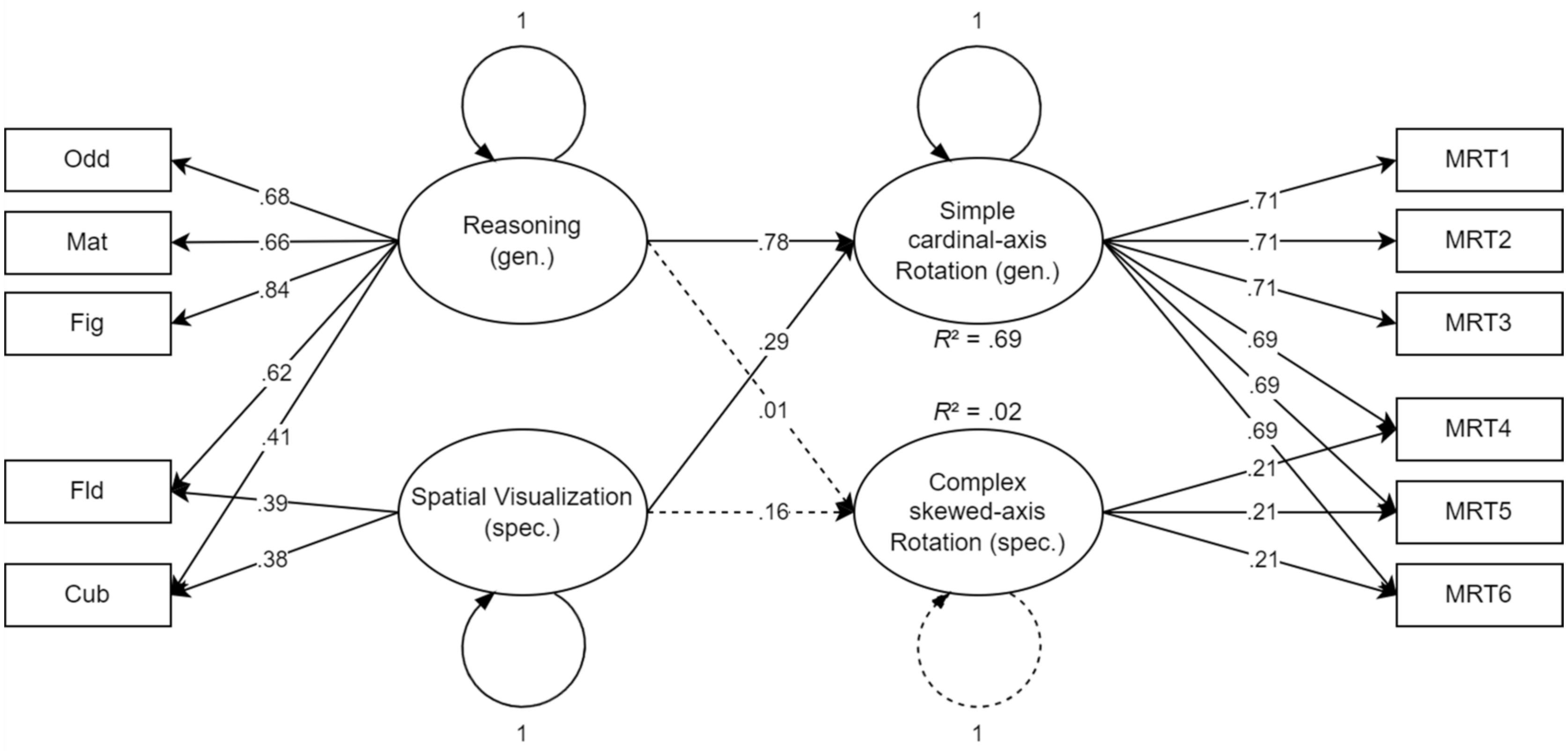
One-factor Measurement Model of Mental Rotation Ability



Note. MRT1-6 = Parcels of items from the mental rotation test. Solid lines depict significant loadings and variances.

Figure A6.3

Structural Model with Nested-factor Model on the Criterion side



6.2 Anhang Studie 2

Table A6.3

Correlation Table of Variables Included in the Analyses

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1.	Vis	-	-0.17	0.25	0.43	0.5	0.35	0.42	0.3	0.4	0.12	0.39	0.29	0.42
2.	GPA		-	-0.16	-0.36	-0.53	-0.29	-0.3	-0.32	-0.32	-0.27	-0.49	-0.31	-0.41
3.	Reasoning			-	0.46	0.45	0.2	0.2	0.17	0.22	0.19	0.06	0.22	0.23
4.	Ck_eng_t1				-	0.69	-	-	-	-	-	-	1	0.69
5.	CK_eng_t2					-	-	-	-	-	-	-	0.69	1
6.	Ck_che_t1						-	0.75	-	-	-	-	1	0.75
7.	Ck_che_t2							-	-	-	-	-	0.75	1
8.	Ck_bio_t1								-	0.65	-	-	1	0.65
9.	Ck_bio_t2									-	-	-	0.65	1
10.	Ck_soc_t1										-	0.48	1	0.48
11.	Ck_soc_t2											-	0.48	1
12.	Ck_tot_t1												-	0.63
13.	Ck_tot_t2													-

Note. All variables are standardized within disciplines. The total content knowledge variable combines the content knowledge variables of all disciplines. Vis = spatial visualization, GPA = high school grade point average, t1 = pretest, t2 = posttest, Ck = content knowledge, eng = civil engineering, che = chemistry, bio = biology, soc = social science, tot = total across all disciplines

Table A6.4

Hierarchic Linear Regression Analysis for Posttest Content Knowledge as Criterion

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9	
	b	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*	*b*	*SE*
CK pre	$.65^{***}$	.03	$.48^{***}$	.06	$.41^{***}$	.06	$.35^{***}$	.06	$.34^{***}$	.06	$.36^{***}$	.06	$.34^{***}$	.06	.34	.06	$.34^{***}$	.06
Domain Eng			0	.08	0	.08	0	.08	0	.08	0	.08	0	.08	0	.08	0	.08
Domain Bio			0	.08	0	.07	0	.07	0	.07	0	.07	0	.07	0	.07	0	.07
Domain Che			0	.08	0	.07	0	.07	0	.07	0	.07	0	.07	0	.07	0	.07
GPA					$-.22^{***}$	.03	$-.38^{***}$	.06	$-.37^{***}$	.06	$-.39^{***}$	.06	$-.33^{***}$	.06	$-.32^{***}$	.06	$-.33^{***}$	.06
Reasoning									$.06^{*}$	.03	.06	.06	.02	.03	.02	.03		
Spat. Vis.													$.22^{***}$	.03	$.29^{***}$	.05	$.22^{***}$	.03
CK pre:Domain Eng			$.20^{*}$	.08	$.19^{*}$	.08	$.21^{*}$	.09	$.19^{*}$	.09	.14	.09	.12	.08	.11	.09	.12	.08
CK pre:Domain Bio			$.18^{*}$	.08	$.20^{**}$	.07	$.27^{**}$	.08	$.28^{**}$	.08	$.26^{**}$	.08	$.23^{**}$	.07	$.24^{**}$	.08	$.23^{**}$	.07
CK pre:Domain Che			$.28^{**}$	.08	$.29^{**}$	.07	$.37^{***}$	.08	$.38^{***}$	.08	$.36^{***}$	.08	$.31^{***}$	.07	$.33^{***}$	.08	$.30^{***}$	.07
GPA:Domain Eng							.09	.09	.08	.09	.10	.09	.04	.08	.02	.08	.04	.08
GPA:Domain Bio							$.25^{**}$	.08	$.24^{**}$	.08	$.27^{**}$	.08	$.24^{**}$	.07	$.22^{**}$	.08	$.24^{**}$	.07
GPA:Domain Che							$.25^{**}$	.08	$.25^{**}$	.08	$.27^{**}$	.08	$.21^{**}$	.07	$.19^{**}$	.07	$.21^{**}$	.07
Reasoning:Domain Eng											$.20^{*}$	.09						
Reasoning:Domain Bio											$.18^{*}$	.08						
Reasoning:Domain Che											.12	.07						
Spat. Vis.:Domain Eng															-.05	.08		
Spat. Vis.:Domain Bio															-.10	.07		
Spat. Vis.:Domain Che															-.12	.07		
R^2	.425		.436		.477		.488		.491		.496		.531		.533		.530	
ΔR^2			$.011^{*}$		$.041^{***}$		$.011^{**}$		$.003^{*}$		.005		$.035^{***}$		.002		-.003	

Note. Social science is used as reference level for the Domain variable. CK pre = pretest content knowledge, Eng = civil engineering, Bio = biology, Che = chemistry, GPA = high school grade point average. Spat. Vis. = spatial visualization ability. * = $p < .05$; ** = $p < .01$;*** = $p < .001$.

Table A6.5

Indirect Effects for the Mediation Model with Posttest Content Knowledge as Criterion

Effects	*b*	β	*SE*	95% CI of *b*	*p*
CivEng					
vis-ck1-ck3	0.35	0.30	0.09	[0.20, 0.54]	0.000
vis-gpa-ck3	0.07	0.06	0.02	[0.04, 0.12]	0.001
vis-gpa-ck1-ck3	0.03	0.03	0.01	[0.02, 0.05]	0.000
vis-ck1-gpa-ck3	0.03	0.05	0.01	[0.02, 0.05]	0.000
total indirect	0.48	0.45	0.09	[0.32, 0.66]	0.000
total	0.81	0.73	0.09	[0.64, 1.00]	0.000
Che					
vis-ck1-ck3	0.43	0.35	0.06	[0.31, 0.57]	0.000
vis-gpa-ck3	0.02	0.02	0.01	[0.00, 0.05]	0.080
vis-gpa-ck1-ck3	0.04	0.05	0.01	[0.03, 0.06]	0.000
vis-ck1-gpa-ck3	0.05	0.08	0.01	[0.02, 0.08]	0.000
total indirect	0.51	0.44	0.07	[0.39, 0.65]	0.000
total	0.83	0.72	0.09	[0.68, 1.01]	0.000
Bio					
vis-ck1-ck3	0.29	0.24	0.06	[0.18, 0.42]	0.000
vis-gpa-ck3	0.02	0.02	0.02	[-0.01, 0.05]	0.195
vis-gpa-ck1-ck3	0.04	0.04	0.01	[0.02, 0.05]	0.000
vis-ck1-gpa-ck3	0.04	0.07	0.01	[0.02, 0.07]	0.000
total indirect	0.36	0.32	0.06	[0.24, 0.49]	0.000
total	0.68	0.60	0.08	[0.54, 0.84]	0.000
SoSci					
vis-ck1-ck3	0.07	0.06	0.04	[-0.01, 0.16]	0.086
vis-gpa-ck3	0.07	0.06	0.02	[0.04, 0.12]	0.000
vis-gpa-ck1-ck3	0.02	0.02	0.01	[0.01, 0.04]	0.000
vis-ck1-gpa-ck3	0.02	0.04	0.01	[0.01, 0.04]	0.003
total indirect	0.19	0.18	0.04	[0.10, 0.28]	0.000
total	0.51	0.46	0.06	[0.39, 0.64]	0.000

Note. Variable chains in the Effects column describe indirect paths from spatial visualization ability to posttest content knowledge. CivEng = civil engineering, Che = chemistry, Bio = biology, SoSci = social science, vis = spatial visualization ability, ck1 = pretest content knowledge, ck3 = posttest content knowledge, gpa = high school grade point average.

6.3 Anhang Studie 3

Table A6.6

Correlation Table of Variables Included in the Analyses

		1.	2.	3.	4.	5.	6.
1.	Vis pre	-	0.77	0.29	0.41	0.32	0.53
2.	Vis post		-	0.36	0.47	0.39	0.62
3.	Ck pre			-	0.48	0.19	0.41
4.	Ck del				-	0.69	0.44
5.	Exam grade					-	0.19
6.	Reasoning						-

Note. Exam grades relate to the proportion of achieved points, therefore ranging from 0 to 1. Pretest content knowledge corresponds to prior knowledge. pre = pretest, post = posttest, del = delayed posttest, vis = spatial visualization ability, ck = content knowledge

Table A6.7

Hierarchic Linear Regression Analysis for Delayed Posttest Content Knowledge

Variable	Model 1			Model 2			Model 3			Model 4		
	b	β	*SE*	*b*	β	*SE*	*b*	β	*SE*	*b*	β	*SE*
Prior knowledge	.74***	.46	.17	.53**	.33	.17	.47**	.29	.17	.53**	.33	.17
Posttest reasoning				.23**	.33	.08	.11	.16	.09			
Posttest spatial vz							.14*	.29	.06	.19***	.38	.05
R^2	.21			.30			.35			.34		
ΔR^2				.09*			.05*			-.01		

Note. Spatial vz = spatial visualization ability. * = *p* < .05; ** = *p* < .01;*** = *p* < .001

Table A6.8

Hierarchic Linear Regression Analysis for Exam Grades

Variable	Model 1			Model 2			Model 3			Model 4		
	b	β	*SE*	*b*	β	*SE*	*b*	β	*SE*	*b*	β	*SE*
Delayed posttest content knowledge	.21***	.70	.03	.23***	.76	.04	.22***	.72	.04	.20***	.67	.04
Prior knowledge				-.06	-.13	.05						
Posttest reasoning							-.01	-.05	.02			
Posttest spatial vz										.01	.07	.02
R^2	.49			.50			.49			.50		
ΔR^2				.01			.00			.01		

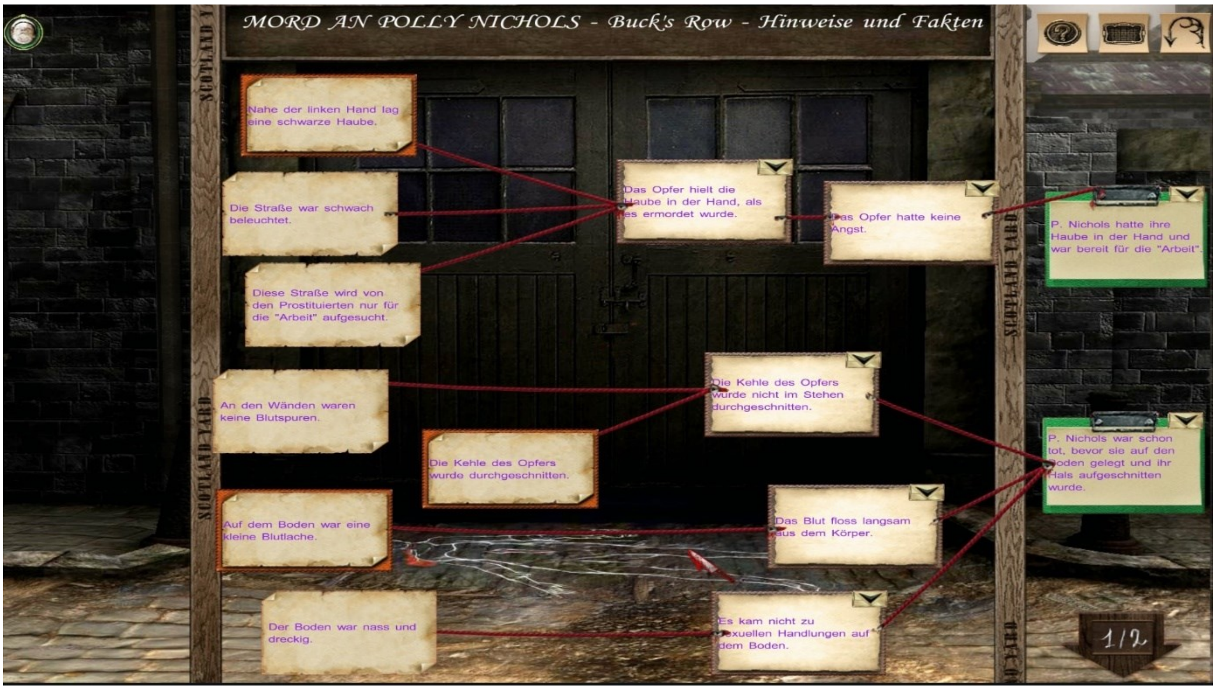
Note. Spatial vz = spatial visualization ability. * = *p* < .05; ** = *p* < .01;*** = *p* < .001

Figure A6.4

Example Screenshot of Sherlock Holmes (First-person Perspective)

Figure A6.5

Example Screenshot of Sherlock Holmes (Deduction Puzzle)



MORD AN POLLY NICHOLS - Buck's Row - Hinweise und Fakten

Nahe der linken Hand lag eine schwarze Haube.

Die Straße war schwach beleuchtet.

Diese Straße wird von den Prostituierten nur für die "Arbeit" aufgesucht.

An den Wänden waren keine Blutspuren.

Die Kehle des Opfers wurde durchgeschnitten.

Auf dem Boden war eine kleine Blutlache.

Der Boden war nass und dreckig.

Das Opfer hielt die Haube in der Hand, als es ermordet wurde.

Das Opfer hatte keine Angst.

Die Kehle des Opfers wurde nicht im Stehen durchgeschnitten.

Das Blut floss langsam aus dem Körper.

Es kam nicht zu sexuellen Handlungen auf dem Boden.

P. Nichols hatte ihre Haube in der Hand und war bereit für die "Arbeit".

P. Nichols war schon tot, bevor sie auf den Boden gelegt und ihr Hals aufgeschnitten wurde.

Figure A6.6

Example Screenshot of Portal 2 (Button Puzzle)

Figure A6.7

Example Screenshot of Portal 2 (Laser Puzzle)