

Verhaltensbezogenes Engagement im Unterricht

Eine Analyse der Generalisierbarkeit und Zuverlässigkeit von Videobeobachtungen

Bianka Troll¹ , Marcus Pietsch^{1,3} und Michael Besser²

¹Institut für Bildungswissenschaft, Leuphana Universität Lüneburg

²Institut für Mathematik und ihre Didaktik, Leuphana Universität Lüneburg

³Institut für Erziehungswissenschaft, Universität Zürich

Zusammenfassung: Die vorliegende schülerzentrierte Videostudie untersucht die Generalisierbarkeit und Zuverlässigkeit von Videobeobachtungen des verhaltensbezogenen *engagements* von Schülerinnen und Schülern unter Berücksichtigung der aktiven Arbeitsphasen (Partner-, Gruppenarbeit, Ergebnissicherung im Plenum) einer dritten Grundschulklasse ($N = 20$ Schülerinnen und Schüler). Anhand eines adaptierten Münchner Aufmerksamkeitsinventars (Helmke & Renkl, 1992) wurden alle Lernenden in ihrem individuellen, verhaltensbezogenen *engagement* in Ein-Sekunden-Intervallen über den Verlauf des Unterrichts von vier unabhängigen Kodiererinnen während der aktiven Arbeitsphasen beurteilt. Videostudien werden in der pädagogisch-psychologischen Unterrichtsforschung häufig zur Beschreibung und Analyse unterrichtlicher Prozesse durchgeführt, obwohl sie als sehr ressourcenintensiv und aufwändig gelten. Daher wird in dieser Studie ebenso untersucht, bis zu welcher Anzahl an Sekunden das verhaltensbezogene *engagement* der Schülerinnen und Schüler noch reliabel beobachtet werden kann, um eine effiziente und ökonomische Designplanung zu ermöglichen. Die Ergebnisse der G-Studie zeigen erwartungskonform, dass der Großteil der erklärbaren Varianz über alle Arbeitsphasen hinweg auf die individuellen Lernenden (18,85%) zurückzuführen ist. Der Zeitpunkt in Sekunden (2,67%) erklärt hingegen nur einen marginalen Varianzanteil. Es bleibt ein großer Anteil unerklärter Residualvarianz (78,48%). Der Zuverlässigkeitskoeffizient liegt mit $\Phi = .99$ in einem sehr guten Bereich. Die Entscheidungsstudie (D-Studie) ergibt, dass selbst bei 30 beobachteten Sekunden ein zuverlässiges Ergebnis erzielt werden kann ($\Phi = .87$). Die Ergebnisse weisen darauf hin, dass schülerzentrierte Videoanalysen sich besonders zur Erfassung des verhaltensbezogenen *engagements* aller Schülerinnen und Schüler eignen, da über den gesamten Unterrichtsverlauf Informationen über alle Lernenden vorliegen. Ebenso tragen die Ergebnisse zu einer ökonomischen Designplanung von schülerzentrierten Videostudien zum verhaltensbezogenen *engagement* bei.

Schlüsselwörter: Verhaltensbezogenes *engagement*, *classroom engagement*, Generalisierbarkeitstheorie, Videoanalysen

Behavioral engagement in classroom – An analysis of the generalizability and dependability of video observations

Abstract: This student-centered video study examines the generalizability and reliability of video observations of behavioral engagement of students, taking into account the active work phases (partner work, group work, plenary) of a third year primary school class ($N = 20$ students). We used an adapted version of the Munich Attention Inventory (Helmke & Renkl, 1992) to assess the behavioral engagement of all learners during the active work phases in one-second intervals by four independent raters. In educational psychology, video studies are often used to describe and analyze classroom processes – although they are considered as being very resource-intensive and costly. Therefore, in order to enable efficient and economical design planning of studies this study investigates up to how many seconds the behavioral engagement of students can be reliably observed. In line with our expectations, the results of the G-Study show that individual learners variance explanation was 18.85% across all work phases. The time in seconds (2.67%) explains only a marginal part of the variance. There remains a large proportion of unexplained residual variance (78.48%). The reliability coefficient with $\Phi = .99$ is very good due to common standards. The decision study (D-Study) shows that results are reliable even when observing just 30 seconds ($\Phi = .87$). The results indicate that student-centered video analyses offer a valuable way to capture the behavioral engagement of all students, in that information on all learners is available throughout the entire lesson. The results also contribute to an economic design of student-centered video studies on behavioral engagement.

Keywords: Behavioral engagement, classroom engagement, generalizability theory, video analysis

Einleitung

Lernprozesse finden vor allem in den regelmäßigen, wechselseitigen und zunehmend komplexen Interaktionen der

Lernenden mit den Lehrpersonen, den Mitschülerinnen und Mitschülern, anderen Beteiligten am Unterricht sowie mit den physischen Objekten in der Umgebung, die den Lernprozess der Lernenden anregen, statt (Downer,

Rimm-Kaufman & Pianta, 2007; Kunter & Trautwein, 2018; Vieluf, Praetorius, Rakoczy, Kleinknecht & Pietsch, 2020). Um die unterrichtlichen Prozesse und die darauf einwirkenden Faktoren abzubilden, wird in der pädagogisch-psychologischen Unterrichtsforschung häufig auf das Angebot-Nutzungs-Modell der Wirkungsweise von Unterricht zurückgegriffen (Fend, 2019; Helmke, 2009; Kunter & Trautwein, 2018; Lipowsky, 2015; Reusser & Pauli, 2010; Seidel, 2014). Schülerinnen und Schüler sind laut diesem Modell aktive (Ko-)Konstrukteure ihres eigenen Lernerfolgs. Die Effektivität des Unterrichts hängt daher nicht nur davon ab, welche Lerngelegenheiten sich den Schülerinnen und Schülern bieten, sondern auch davon, ob und wie sie diese nutzen (Vieluf et al., 2020). In dem von Vieluf und Kolleginnen und Kollegen (2020) vorgeschlagenen integrierten Angebot-Nutzungs-Modell stellt das unterrichtliche Angebot entsprechend eine Ko-Konstruktion aus Schülerinnen- und Schülerhandeln, Lehrpersonenhandeln und dem fachlichen Inhalt dar. Dabei liegt Unterricht nicht in der alleinigen Verantwortung der Lehrpersonen, sondern ist auf die Partizipation der Schülerinnen und Schüler angewiesen (Scherzinger, Wettstein & Wyler, 2018). Die Nutzung der unterrichtlichen Angebote durch die Schülerinnen und Schüler umfasst einen komplexen, mentalen Prozess aus Kognitionen, emotionalem und motivationalem Erleben (Vieluf et al., 2020). Um unterrichtliche Prozesse aus Perspektive der Lernenden zu beschreiben und zu analysieren wird – insbesondere im anglo-amerikanischen Raum – häufig das Konstrukt des *engagements* herangezogen. *Engagement* gilt als ein konsistenter und starker Prädiktor für schulische Outcomes und wird daher als eine Schlüsselvariable zur Erklärung von schulischem Erfolg bezeichnet (Fredricks, Blumenfeld & Paris, 2004; Gettinger & Walter, 2012; Ponitz, Rimm-Kaufman, Grimm & Curby, 2009; Skinner & Pitzer, 2012). In der Literatur finden sich einerseits Definitionen, die *engagement* eher als ein globales Konstrukt verstehen, das sich sowohl auf schulbezogene als auch unterrichtliche Aspekte bezieht (*school engagement*) (Carter, Reschly, Lovelace, Appleton & Thompson, 2012; Eccles & Wang, 2012). Andererseits werden zum Teil auch nur unterrichtliche Aktivitäten (*classroom engagement*) in die Definition des Konstrukts mit einbezogen (Eccles & Wang, 2012; Lam et al., 2014; Wang, Bergin & Bergin, 2014). Der vorliegende Artikel orientiert sich am letztgenanntem Verständnis, da primär unterrichtliche Prozesse im Fokus der Analysen stehen.

Wird nun das *engagement* der Schülerinnen und Schüler im Unterricht untersucht, so ergeben sich verschiedene Möglichkeiten der Datenerhebung (Fredricks et al., 2004). Einerseits können Lehrpersonen das *engagement* ihrer Schülerinnen und Schüler beurteilen, andererseits können die Lernenden sich selbst retrospektiv in ihrer Aktivität

einschätzen. Eine weitere Möglichkeit der Datengewinnung stellen Videoanalysen dar. Diese Art der Erhebung ermöglicht es, die im Unterricht ablaufenden Prozesse detailliert und objektiv zu erfassen. Dennoch gelten sie als sehr aufwändig und ressourcenintensiv (Begrich, Fauth, Kunter & Klieme, 2017). Im Rahmen dieser Forschungsarbeit wird daher eine auf Videodaten basierende Generalisierbarkeitsstudie durchgeführt, die untersucht, welches Ausmaß (Anzahl an zu beobachtenden Sekunden) an Beobachtungen notwendig ist, um hinreichend zuverlässig Kennwerte über die verhaltensbezogenen Dimensionen des *classroom engagements* der Lernenden zu erhalten (Renkl & Helmke, 1993). Die Ergebnisse der Studie sollen zukünftig eine effiziente Designplanung von Studien zur Beschreibung und Analyse von Unterricht ermöglichen. Im Rahmen dieses Artikels wird zu Beginn das Konstrukt des *classroom engagements* erarbeitet. Im anschließenden Kapitel werden die verhaltensbezogenen Dimensionen ausführlich dargestellt und es wird auf deren methodische Erfassung eingegangen.

Das mehrdimensionale Konstrukt *classroom engagement*

Classroom engagement stellt einen individuellen Prozess dar, der in enger Beziehung zu den im Unterricht stattfindenden Interaktionen steht (Downer et al., 2007; Fredricks & McColskey, 2012). *Engagement* wird als ein Meta-Konstrukt verstanden, das sich sowohl aus verhaltensbezogenen als auch mentalen Dimensionen zusammensetzt. Dabei werden meistens die drei Dimensionen *affective engagement*, *cognitive engagement* und *behavioral engagement* unterschieden (Appleton, Christenson & Furlong, 2008; Fredricks & McColskey, 2012; Lam et al., 2014). Diese drei Dimensionen von *classroom engagement* adressieren ausschließlich die individuellen verhaltensbezogenen und mentalen Aktivitäten der Schülerinnen und Schüler während des Unterrichts. Da die Lernenden ebenso an der Konstruktion des unterrichtlichen Angebots beteiligt sind, schlagen Autorinnen und Autoren eine Erweiterung des dreidimensionalen Konstrukts um die Dimension *agentic engagement* vor (Reeve, 2013; Reeve & Tseng, 2011;). *Behavioral* und *agentic engagement* können dabei als verhaltensbezogene Lernaktivitäten gefasst werden und nehmen eine entscheidende Rolle bei der Gestaltung und dem Verlauf des Unterrichts ein.

Die Dimensionen lassen sich wie folgt charakterisieren: *Affective engagement* umfasst alle unterrichtsbezogenen Emotionen, wie Interesse, Freude, Enthusiasmus, aber auch Angst und Langeweile (Fredricks et al., 2004; Lawson & Lawson, 2013; Linnenbrink-Garcia, Rogat & Koskey, 2011; Wang et al., 2014). *Cognitive engagement* bezieht

mentalen Prozesse, wie Strategien zur Informationsverarbeitung, Strategien zur Selbstregulation und Metakognitionen mit ein (Fredricks et al., 2004; Lam et al., 2014; Lawson & Lawson, 2013; Pekrun & Linnenbrink-Garcia, 2012). Der Dimension *behavioral engagement* wird in der Literatur das beobachtbare Verhalten, wie die gezeigte Anstrengungsbereitschaft (*effort*), das Durchhaltevermögen (*persistence*), das Aufmerksamkeitsverhalten (*attention*), die Partizipation sowie die Beteiligung am Unterrichtsgeschehen zugeschrieben (Downer et al., 2007; Fredricks et al., 2004; Hughes, Luo, Kwok & Loyd, 2008; Lam et al., 2014; Wang et al., 2014). *Agentic engagement* rekuriert vor allem auf ko-konstruktive und interaktive Prozesse des Unterrichts. Hierunter verstehen die Autoren "students' constructive contribution into the flow of the instruction they receive" (Reeve & Tseng, 2011, S.258). *Agentic engagement* umschließt jegliche lernenden-initiierten unterrichtsbezogenen Verhaltensweisen, die zu einer motivierenden und unterstützenden Lernumgebung für den jeweiligen Lernenden bzw. die jeweilige Lernende führen. Dazu zählen beispielsweise verhaltensbezogene Aktivitäten, wie "offer input, express a preference, offer a suggestion or contribution, ask a question, communicate what they are thinking and needing, [...] communicate [the] level of interest, solicit resources or learning opportunities, seek ways to add personal relevance to the lesson, [...], seek clarification, generate options, communicate likes and dislikes, ..." (Reeve & Tseng, 2011, S.258).

Innerhalb eines Individuums stehen die vier Dimensionen in einer dynamischen Wechselbeziehung zueinander (Fredricks et al., 2004). Obwohl die vier Dimensionen als distinkte Faktoren dem Konstrukt *classroom engagement* zugeordnet werden können, korrelieren sie auf einem mittleren bis hohen Niveau ($.30 < r < .57$) (Reeve, 2013). Inhaltlich bedeutet dies, dass motivationales und emotionales Erleben (Skinner, Kindermann & Furrer, 2009) sowie kognitive Lernaktivitäten (Seidel, 2003) sich als beobachtbare Verhaltensweisen der Schülerinnen und Schüler im Unterricht manifestieren – mentale Aktivitäten beeinflussen verhaltensbezogene Aktivitäten bzw. verhaltensbezogene Aktivitäten beeinflussen auch die mentalen Prozesse der Lernenden. Verdeutlicht werden kann diese reziproke Beziehung von verhaltensbezogenen und mentalen Lernaktivitäten der Schülerinnen und Schüler beispielhaft an folgenden Ausführungen: Verbale Äußerungen (als Indikator von *behavioral* und *agentic engagement*) zwingen die Sprecherin bzw. den Sprecher die eigenen Gedanken zu ordnen und für andere Akteurinnen und Akteure im Klassenzimmer verständlich zu formulieren (Lipowsky, Rakoczy, Pauli, Reusser & Klieme, 2007). Komplexe Gespräche können kognitive Konflikte auslösen und dazu beitragen, dass bestehende Konzepte hinterfragt oder erweitert werden (Lipowsky et al., 2007; Wuttke, 2005).

Auch können unterschiedliche kognitive Verarbeitungen von Lernangeboten dazu führen, dass bestimmte Lernende eine Antwort auf z.B. eine von der Lehrkraft gestellte Frage eher geben können und daher häufiger am Unterrichtsgespräch beteiligt sind (Decristan et al., 2019; Lipowsky et al., 2007; Lipowsky, Pauli & Rakoczy, 2008).

Mentalen Lernaktivitäten der Schülerinnen und Schüler (*cognitive* und *affective engagement*) beziehen sich vor allem auf die aktive Verarbeitung relevanter Inhalte im Unterricht, verhaltensbezogenen Lernaktivitäten (*behavioral* und *agentic engagement*) hingegen wird für die Gestaltung und den Verlauf des Unterrichts ein hoher Stellenwert beigemessen.

Verhaltensbezogenes engagement im Unterricht

Studien konnten bereits zeigen, dass das *agentic engagement* der Schülerinnen und Schüler (z.B. Fragen stellen, benötigte Hilfe aufzeigen, Interesse verkünden) die Lehrpersonen zu einem stärker unterstützenden und motivierenden Verhalten gegenüber den Lernenden bewegt (Reeve, 2013). Außerdem eröffnen Schülerinnen und Schüler durch aktive verbale Beteiligung am Unterrichtsgespräch (als ein Indikator von *behavioral engagement*) ein Lernangebot für andere Schülerinnen und Schüler und tragen dadurch ebenso zur Mitgestaltung von Unterricht bei (Vieluf et al., 2020).

Verhaltensbezogenes *engagement* äußert sich in beobachtbaren Verhaltensweisen der Schülerinnen und Schüler während des Unterrichts und steht in einem direkten Zusammenhang mit der akademischen Leistung (Cappella, Kim, Neal & Jackson, 2013; Downer et al., 2007). Das Aufmerksamkeitsverhalten (Ophardt & Thiel, 2015), das Durchhaltevermögen (Beserra, Nussbaum & Oteo, 2019), die aktive Beteiligung am Unterrichtsgespräch (Decristan et al., 2019; Jurik, Gröschner & Seidel, 2013) – als Indikatoren des *behavioral engagements* – und das Äußern von Bedürfnissen (Reeve & Tseng, 2011) – als Indikator des *agentic engagements* – gelten als wichtige Determinanten individueller Leistungsentwicklung. Der Unterricht sollte daher den Lernenden die Möglichkeit bieten, diese Verhaltensweisen zu maximieren, um dadurch die individuelle Leistungsentwicklung zu gewährleisten (Beserra et al., 2019; Ophardt & Thiel, 2015). Dabei wird angenommen, dass das verhaltensbezogene *engagement* der Lernenden innerhalb und über mehrere Schulstunden hinweg variieren kann (Fredricks et al., 2004). Auch zeigten Studien, dass je nach Sozialform das *behavioral engagement* der Schülerinnen und Schüler unterschiedlich stark ausgeprägt sein kann. In Kleingruppenarbeit beobachteten die Wissenschaftlerinnen und Wissenschaftler ein höheres

engagement der Lernenden als in Plenumssituationen oder Einzelarbeitsphasen (Downer et al., 2007). In der Literatur wird in diesem Zusammenhang vielfach auf die *engaged time* Bezug genommen (Gettinger & Walter, 2012). *Engaged time* ist die Zeit, in der die Schülerinnen und Schüler aktiv am Unterrichtsgeschehen teilnehmen. Sie spielt bereits in klassischen Modellen schulischen Lernens (Bloom, 1974; Carroll, 1963) eine entscheidende Rolle und umfasst den Anteil an Unterrichtszeit, in dem die Schülerinnen und Schüler kognitiv und beobachtbar an einer Aufgabe arbeiten, was sich durch entsprechendes passives und aktives Verhalten (Aufmerksamkeitsverhalten, Arbeitsverhalten wie zuhören, schreiben, lesen, Beteiligung an Diskussionen) zeigt. Eine spezifische Form der *engaged time* zeichnet sich in der *time on task* ab, die in der Literatur häufig mit der Dimension *behavioral engagement* gleichgesetzt wird (Bangert-Drowns & Pyke, 2001). *Time on task* umfasst in der Definition von Gettinger und Walter (2012) jedoch ausschließlich die Zeit, in der die Schülerinnen und Schüler eine bestimmte Lernaufgabe bearbeiten. Wenn *time on task* jedoch mit verhaltensbezogenem *engagement* gleichgesetzt wird, würde dies bedeuten, dass die Schülerinnen und Schüler, die z.B. bereits eine Aufgabe fertig bearbeitet haben, weniger involviert sind, als die Schülerinnen und Schüler, die mehr Zeit für die Bearbeitung der Aufgabe benötigen (Bangert-Drowns & Pyke, 2001). *Time on task* spiegelt daher lediglich die Zeit wider, die für gezielte Aufgaben mit klaren Lehr- oder Lernzielen aufgewendet wird (Getttinger & Walter, 2012) und zeigt keine qualitativen Ausprägungen in den Verhaltensweisen der Schülerinnen und Schüler. Im Folgenden wird daher näher auf die Erfassung und Operationalisierung der verhaltensbezogenen Dimensionen von *engagement* eingegangen.

Operationalisierung verhaltensbezogener Dimensionen des *classroom engagements*

Verhaltensbezogenes *engagement* kann auf unterschiedliche Weise erfasst werden. Zum einen können Lehrpersonen Einschätzungen über das *engagement* ihrer Schülerinnen und Schüler abgeben, zum anderen können die Lernenden sich selbst über Fragebögen retrospektiv einschätzen (Fredricks & McColskey, 2012). Diese Art von Erfassung ermöglicht es jedoch nicht, „etwas über die genaue prozessuale Verkettung von Ereignissen zu erfahren, die zur sozialen Konstitution vergangener Unterrichtssituationen beigetragen haben“ (Herrle & Breitenbach, 2016, S.30). Da es sich bei verhaltensbezogenem *engagement* jedoch um beobachtbare Lernaktivitäten der Schülerinnen und Schüler handelt, können ebenso systematische Beobachtungen vorgenommen werden (Appleton & Lawrenz, 2011; Fredricks et al., 2004). Die Verwendung derartiger

Beobachtungsverfahren bringt den Vorteil mit sich, dass das verhaltensbezogene *engagement* im Verlauf des Unterrichts, eingebettet im entsprechenden Kontext der Situation, gemessen werden kann (Fredricks & McColskey, 2012).

In der pädagogisch-psychologischen Unterrichtsforschung sind (Video-)Beobachtungen des Lehrenden- und Lernendenverhaltens bzw. der Interaktion der Akteurinnen und Akteure im Unterricht eine etablierte Methode, um detaillierte Einblicke in das Unterrichtsgeschehen zu erlangen (Cadima, Leal & Burchinal, 2010; Helmke et al., 2008; Herrle & Breitenbach, 2016; Wild, 2003). Aufgrund der hohen Objektivität und des geschulten methodisch-didaktischen Blicks der externen Beobachterinnen und Beobachter (Begrich et al., 2017) bezeichnet Helmke (2009, S.288) diesen Zugang sogar als den „Königsweg zur Beschreibung und Bewertung des Unterrichts“.

Ein im deutschen Raum häufig eingesetztes Beobachtungsinstrument, das zur Erfassung des verhaltensbezogenen *engagements* von Schülerinnen und Schülern herangezogen werden kann, ist das Münchner Aufmerksamkeitsinventar von Helmke und Renkl (1992). Das Aufmerksamkeitsinventar wurde in seiner ursprünglichen Form als ein Instrument zur teilnehmenden Beobachtung entwickelt. Mithilfe des Instruments wird versucht, aus bestimmten Verhaltensweisen, wie der Blickzuwendung und -fixation, der aktiven Beteiligung in Diskussionen oder anderen lernrelevanten Aufgaben, Rückschlüsse auf das Aufmerksamkeitsverhalten von Lernenden (als ein Indikator von *behavioral engagement*) zu ziehen (Helmke, 1988; Hommel, 2011; Ophardt & Thiel, 2015). Das Beobachtungsinventar ist so konzipiert, dass über den gesamten Verlauf des Unterrichts das Verhalten aller Schülerinnen und Schüler erfasst wird. Außerdem wird der Kontext der Situationen mitkodiert (Sozialform, fachliche – nicht-fachliche Situation etc.). Dies ist insofern von Relevanz, da über eine Unterrichtsstunde hinweg und je nach Sozialform die Schülerinnen und Schüler nicht durchgehend ein positives und lernförderliches Verhalten zeigen. Büttner und Schmidt-Atzert (2004, S.31) bezeichnen das Münchner Aufmerksamkeitsinventar als „das differenzierteste und das bestevaluierte Beobachtungssystem, das im deutschen Sprachraum für den schulischen Unterricht entwickelt worden ist“. Die Objektivität und Reliabilität des Instruments wurde in einer Längsschnittstudie in 54 Grundschulklassen geprüft (Helmke & Renkl, 1992). Aufgrund der ressourcenintensiven Datengewinnung über teilnehmende Beobachtung und den damit verbundenen Herausforderungen wurde das Münchner Aufmerksamkeitsinventar von Hommel (2012a) und auch Ranger (2017) als ein Instrument für die Analyse von Videodaten weiterentwickelt.

Im Folgenden wird nun näher auf die Erfassung von verhaltensbezogenem *engagement* mithilfe des Beobach-

Tabelle 1. Operationalisierung des verhaltensbezogenen *engagements* von Lernenden im Verlauf des Unterrichts in Anlehnung an das Münchner Aufmerksamkeitsinventar von Helmke (1988)

Nicht-aufgabenbezogenes Verhalten (off-task / disengaged)	Aufgabenbezogenes Verhalten (engaged)
(0) <i>Aktiv/störend</i> (z.B. scherzen, herumalbern)	(2) <i>Passiv</i> (z.B. zuhören, ohne selbst aktiv zu sein, dem Unterrichtsmittelpunkt zugewandt sein)
(1) <i>Passiv</i> (z.B. längeres Herausschauen aus dem Fenster, ruhiges Beschäftigen mit einem nicht-aufgabenbezogenen Lerngegenstand)	(3) <i>Aktiv fremd-initiiert</i> (z.B. Befolgen von Anweisungen, Diskutieren mit Mitschülerinnen und Mitschülern, Beantwortung von Fragen)
	(4) <i>Aktiv selbst-initiiert</i> (z.B. Initiative in Diskussionen ergreifen, Unterstützung von Mitschülerinnen und Mitschülern, Meldung, Fragen stellen, Input anbieten)

tungsinventars eingegangen. Bei der Erfassung von verhaltensbezogenem *engagement* wird zwischen einem aufgabenbezogenen Verhalten (*engaged*) und nicht-aufgabenbezogenen Verhalten (*off-task* bzw. *disengaged*; Downer et al., 2007; Helmke, 1988; Lam et al., 2014) unterschieden, das jeweils einen passiven als auch aktiven Zustand einnehmen kann (siehe Tabelle 1). Es wird daher nicht nur erfasst, ob ein Schüler bzw. eine Schülerin sich der Lernaufgabe widmet, sondern auch, wie er bzw. sie sich mit der Bearbeitung der Aufgabe auseinandersetzt. Da das verhaltensbezogene *engagement* über die gesamte Unterrichtsstunde hinweg beobachtet wird, ermöglicht dieses Vorgehen, Rückschlüsse auf das Durchhaltevermögen der Schülerinnen und Schüler (als Indikator für *behavioral engagement*). Die Anstrengungsbereitschaft sowie das Aufmerksamkeitsverhalten zeigen sich im Ausmaß der nicht-aufgabenbezogenen und aufgabenbezogenen Verhaltensweisen, wobei die aufgabenbezogenen Aktivitäten in ihrer Qualität ausdifferenziert werden (passives, fremd-initiiertes und selbst-initiiertes aktives Verhalten). Die Dimension *agentic engagement* findet sich vor allem in dem aufgabenbezogenen aktiv selbst-initiierten Verhalten wieder. Verhaltensbezogenes *engagement* wird in diesem Sinne „als Verhalten, das der didaktisch/inhaltlichen Thematik der jeweiligen Lehr-Lernsituation zuordenbar ist“, operationalisiert (Hommel, 2011, S.120).

Fragestellung

Sollen nun Videobeobachtungen zur Untersuchung des verhaltensbezogenen *engagements* durchgeführt werden, so liegen kaum Informationen darüber vor, welches Ausmaß an Beobachtungen notwendig ist, um hinreichend zuverlässige Kennwerte zu erhalten. Diese Informationen sind jedoch von besonderer Relevanz, da es infolge knapper Untersuchungszeiten, begrenzter Belastbarkeit

der Versuchspersonen sowie mangelnder Ressourcen in aller Regel ausgeschlossen ist, prophylaktisch sehr intensive Erhebungen durchzuführen (Renkl & Helmke, 1993). In der aktuellen Literatur finden sich keine den Autorinnen und Autoren bekannten Studien, die sich dieser Analyse widmen. Die vorliegende Studie greift dieses Desideratum auf.

Im Rahmen dieser Studie wird daher untersucht, in welchem Ausmaß (Anzahl an Sekunden) Schülerinnen und Schüler im Verlauf einer Unterrichtsstunde unter Berücksichtigung der Sozialform beobachtet werden müssen, um das verhaltensbezogene *engagement* mithilfe des Münchner Aufmerksamkeitsinventars (Helmke & Renkl, 1992) reliabel messen zu können. Zur Überprüfung der Fragestellung wurde eine quantitativ standardisierte Videoanalyse vorgenommen (Pauli, 2012). Die Daten wurden unter Verwendung der Generalisierbarkeitstheorie (Brennan, 2001) ausgewertet. Die Ergebnisse dieser Studie sollen zukünftig eine effiziente Designplanung von Studien zur Beschreibung von Unterricht ermöglichen.

Methode

Studiendesign

Eine dritte inklusive Klasse einer Gemeinschaftsgrundschule in Nordrhein-Westfalen nahm an dieser Studie teil. Über eine Woche hinweg (immer vormittags, Dienstag bis Freitag) wurde der Unterricht in den Fächern Deutsch, Mathematik und Sachunterricht schülerzentriert videografiert (Paulicke, Schmidt & Ehmke, 2015). Dies bietet den Vorteil, dass für alle Schülerinnen und Schüler der Klasse zu jeder Zeit Informationen zu deren verhaltensbezogenem *engagement* vorliegen. Neben einer dynamischen Lehrkraftkamera wurden zwei GoPro Kameras als Überblicksperspektive neben dem Smartboard sowie

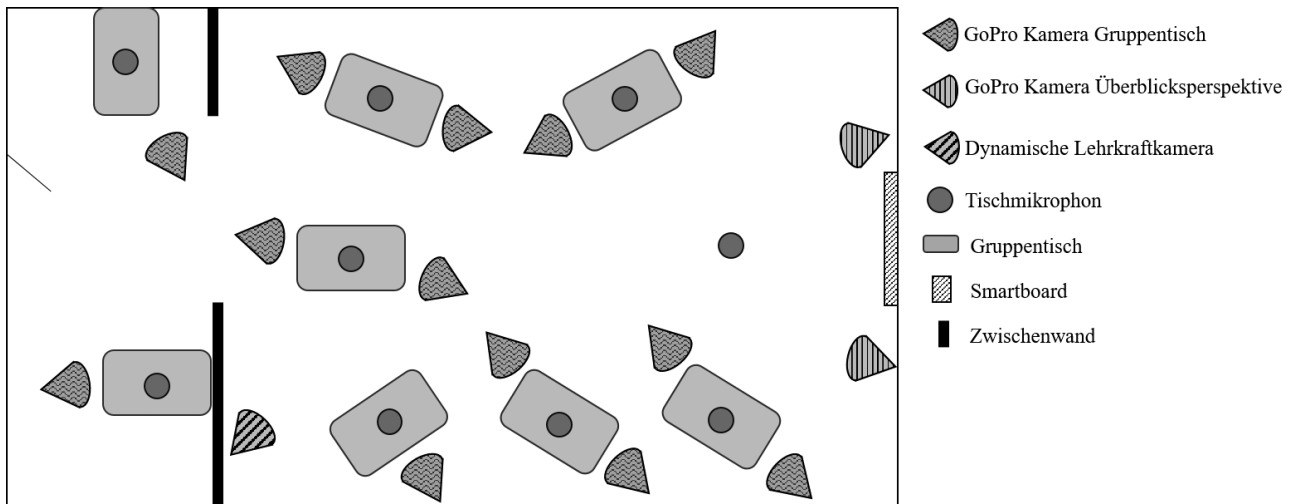


Abbildung 1. Raumaufteilung und Übersicht über die installierten Kameraperspektiven im Klassenzimmer.

zusätzlich an jedem von insgesamt acht Gruppentischen ein bis zwei weitere GoPro Kameras installiert. Die Lehrkraft wurde mit einem Ansteckmikrofon ausgestattet, die Gruppentische waren jeweils mit einem Tischmikrofon versehen (siehe Abbildung 1). Die Bild- und Tonspuren wurden anschließend synchronisiert.

Im Schulkonzept der Gemeinschaftsgrundschule nehmen kooperative Lernformen einen besonderen Raum ein. Aus diesem Grund wurden in den videografierten Unterrichtsstunden häufig kooperative Arbeitsaufträge verteilt, denen erst eine kurze Einführung im Plenum vorge-schaltet war und die dann in Partner- bzw. Gruppenarbeit umgesetzt wurden. Im Anschluss an diese Phase wurden die Gruppenergebnisse im Plenum gesammelt und gemeinsam diskutiert. Die Arbeitsphasen der Deutschstunde, die dieser Arbeit zugrunde liegt, waren wie folgt verteilt: Insgesamt dauerte die Unterrichtsstunde 39 Minuten. Davon benötigte die Lehrkraft knapp sieben Minuten, um eine für die Unterrichtsstunde relevante Geschichte vorzulesen und den Arbeitsauftrag zu verteilen. Circa neun Minuten (545 Sekunden) hatten die Schülerinnen und Schüler in der ersten aktiven Arbeitsphase in Partnerarbeit Zeit, diesen Arbeitsauftrag zu bearbeiten. Im Anschluss folgte eine knapp zwölfminütige Gruppenarbeit (690 Sekunden), in der zwei Partnerteams gemeinsam an einem an die vorherige Aufgabe geknüpften Auftrag arbeiteten. Die Ergebnisse wurden dann in einer circa fünfminütigen Sicherungsphase im Plenum (284 Sekunden) besprochen. Insgesamt sechs Minuten der Unterrichtszeit nahmen Übergänge zwischen den Phasen ein.

Stichprobe

Für die vorliegende Studie wurde die Deutschstunde am Freitagvormittag genutzt, da die Schülerinnen und Schüler zu diesem Zeitpunkt bereits mit dem Aufnahmesystem

vertraut waren. Von den 27 Schülerinnen und Schülern der Klasse waren am Aufnahmetag 24 Schülerinnen und Schüler anwesend. Die Analysen dieser Studie beziehen jedoch lediglich 20 Schülerinnen und Schüler ein, da von vier Lernenden aufgrund von technischer Probleme größtenteils keine Videodaten vorliegen. Von den 20 Schülerinnen und Schülern sind elf weiblich, acht Lernende haben einen sonderpädagogischen Förderbedarf, welcher entweder bereits diagnostiziert oder von der Lehrkraft vermutet wurde. Zwölf Schülerinnen und Schüler wurden im Fach Deutsch von der Lehrkraft als leistungsschwach eingeschätzt (Schulnote 3 bzw. 4).

Instrument

Das verhaltensbezogene *engagement* wurde mittels des niedrig-inferenten Beobachtungssystems „Münchener Aufmerksamkeitsinventar (MAI)“ (Helmke, 1988; Helmke & Renkl, 1992) erfasst (siehe Tabelle 1). Um eine möglichst detailgetreue und realitätsnahe Erfassung des verhaltensbezogenen *engagements* zu ermöglichen, wurde ein Ein-Sekunden-Intervall genutzt (Skinner et al., 2009), mit welchem alle Schülerinnen und Schüler jeweils für die komplette Unterrichtsstunde individuell von externen Raterinnen beobachtet wurden. Dabei wurde sekundenweise das Verhalten der Schülerinnen und Schüler erfasst und entsprechend des Vorgehens nach Helmke und Renkl (1992) zuerst entschieden, ob es sich um ein aufgabenbezogenes Verhalten (*engaged*) oder um ein nicht-aufgabenbezogenes Verhalten (*disengaged*) handelt. In einem nächsten Schritt wurde das Verhalten weiter in seine jeweiligen qualitativen Abstufungen unterteilt (Tabelle 1). So konnte ebenso die Länge des jeweiligen beobachtbaren Verhaltens der Schülerinnen und Schüler kodiert werden.

Die im MAI erhobenen Unterrichtskontexte wurden an den entsprechenden Kontext angepasst. In der vorliegen-

den Studie werden folgende fachliche und direkt unterrichts- und lehrstoffbezogene Arbeitskontexte (Pröscholdt, Stumpf & Schneider, 2011) unterschieden: Plenumsphase, Einzelarbeit, Partnerarbeit, Gruppenarbeit, Ergebnissicherung im Plenum. Die Übergangsphasen zwischen verschiedenen Arbeitsphasen wurden zwar dokumentiert, aber in der Auswertung der Studie nicht weiter berücksichtigt. Eine Arbeitsphase begann immer dann, wenn die Lehrkraft entweder eine mündliche Ansage machte oder die Stoppuhr stellte, die den Zeitrahmen für die Arbeitsphase angab. Eine Arbeitsphase galt als beendet, wenn die Lehrkraft eine mündliche Ansage an alle Schülerinnen und Schüler machte. Für die vorliegende Studie sind lediglich die Arbeitsphasen relevant, in denen die Schülerinnen und Schüler die Möglichkeit hatten, sich aktiv mit dem Lerngegenstand auseinanderzusetzen. Aus diesem Grund beschränken sich die Analysen auf die Partnerarbeitsphase, die Gruppenarbeitsphase und die Sicherungsphase im Plenum.

Kodiererinnenschulung und Beurteilerinnenübereinstimmung

Die Kodierung des verhaltensbezogenen *engagements* wurde von drei Kodiererinnen (sehr gute Studierende des ersten und dritten Mastersemester für das Lehramt an Grundschulen), sowie der Erstautorin des Beitrags vorgenommen. Da für das Projekt Unterricht in den Fächern Mathematik, Deutsch und Sachunterricht videografiert wurde, wurde bei der Auswahl der Beurteilerinnen auf deren Fächerkombination im Studium geachtet – Kodiererin 1 studierte Mathematik und Sachunterricht, Kodiererin 2 Deutsch und Kunst, Kodiererin 3 Deutsch und Sachunterricht. Zum Zeitpunkt der Kodierung waren alle drei Kodiererinnen als studentische Hilfskräfte im Projekt eingestellt.

Vor der Hauptuntersuchung fand eine Kodiererinnenschulung statt (Hess & Denn, 2018; Lotz, 2016). Vor der ersten Kodiererschulung setzen sich die Beurteilerinnen mit dem schülerzentrierten Aufnahmesystem, den vorliegenden Schulstunden und der videografierten Klasse auseinander. Die Kodiererinnenschulung selbst erstreckte sich über zwei Tage mit insgesamt zehn Stunden Schulungszeit. Am ersten Schulungstag (sechs Stunden) wurden das Kategoriensystem und die dahinterliegende Theorie präsentiert und gemeinsam diskutiert. Im Anschluss wurden das methodische Vorgehen erläutert und eine erste gemeinsame Probekodierung durchgeführt. Nach intensiver Diskussion und weiteren Übungen kodierten die Kodiererinnen unabhängig voneinander zuvor festgelegte Sequenzen, die mit der Expertinnenkodierung (Kodierung der Erstautorin) verglichen wurden. Aufgrund einer gerin-

Tabelle 2. Übereinstimmungen zwischen Expertinnenkodierung und trainierten Beurteilerinnen bei der Kodierung des verhaltensbezogenen engagements

	Cohen's Kappa	Cohen's Weighted Kappa ^a	Prozentuale Übereinstimmung
Kodiererin 1	.64	.71	76.21 %
Kodiererin 2	.64	.70	75.21 %
Kodiererin 3	.68	.75	78.56 %

Anmerkungen: 597 Sekunden des verhaltensbezogenen *engagement* drei Lernender aus beiden Leistungsgruppen (leistungsstark, leistungsschwach) als Grundlage für Beurteilerinnenübereinstimmung; Fleiss Kappa $k = .61$; ^aDa es sich um ordinal skalierte Daten handelt, ist es ebenso möglich, für die Übereinstimmung ein Cohen's Weighted Kappa zu berechnen (Cohen, 1968; Döring & Bortz, 2016).

gen Übereinstimmung (Cohen's Kappa $< .50$) wurde ein erneuter Schulungstag organisiert (weitere vier Stunden). Um die kriteriale Validität sicherzustellen, sammelten alle Beurteilerinnen bis zum nächsten Schulungstag Ankerbeispiele aus den Unterrichtsvideos. So konnte überprüft werden, dass ein geteiltes Verständnis über alle Kategorien vorlag. Nach erneuter Übung und umfangreicher Diskussion der Ergebnisse (diskursive Validierung) wurde aufgrund von forschungsökonomischen Gründen das Training mit moderaten bis guten Übereinstimmungswerten (Tabelle 2) abgeschlossen (Döring & Bortz, 2016, S.346).

Für die Hauptuntersuchung wurden die Schülerinnen und Schüler über die gesamte Unterrichtszeit hinweg einzeln von je einer Kodiererin beobachtet. Die schülerzentrierte Aufnahmemethode ermöglichte es, die Schülerinnen und Schüler auch beim Verlassen des Arbeitsplatzes weiter zu beobachten, sodass vollständige Informationen über jede Schülerin bzw. jeden Schüler über die gesamte Stunde hinweg vorliegen.

Datenanalyse

Für die vorliegende Studie erwies sich der methodische Ansatz der Generalisierbarkeitstheorie (Brennan, 2001) als vielversprechend¹. Das verhaltensbezogene *engagement* setzt sich aus dem wahren Wert und einem Messfehler zusammen. Dabei wird zwischen zufälligen und damit nicht erklärbaren sowie nicht zufälligen und damit erklärbaren Messfehlern unterschieden. Der Erwartungswert über alle Beobachtungen hinweg wird in der Generalisierbarkeitstheorie (G-Theorie) als *universe score* bezeichnet und entspricht dem *wahren Wert* der Klassischen Test-Theorie (KTT). Die einzelnen Faktoren, die simultan auf den Messfehler der empirischen Beobachtungen wirken,

¹ Eine detaillierte Erläuterung des statistischen Modells der Generalisierbarkeitstheorie kann im Rahmen dieses Artikels nicht geleistet werden und ist bei Brennan (2001), Casale, Hennemann, Volpe, Briesch und Grosche (2015), Praetorius (2014) oder Renkl und Helmke (1993) ausführlich nachzulesen.

werden in der G-Theorie als *Facetten* bezeichnet. Der *universe score* wird entsprechend über die Werte der festgelegten Bedingungen (Faktorstufen in der KTT) innerhalb einer Facette geschätzt. Theoretisch gibt es eine unendlich große Anzahl an Bedingungen, die innerhalb einer Facette bedeutsam sind. In der Generalisierbarkeitstheorie (G-Theorie) spricht man daher von dem *Universum der zulässigen Bedingungen*. Die Bedingungen aus denen sich die Facetten zusammensetzen, werden aus den Fragestellungen und Forschungsinteressen abgeleitet und stellen somit eine Zufallsauswahl aus dem Universum aller zulässigen Beobachtungen dar (Brennan, 2001; Casale, Henemann, Volpe, Briesch & Grosche, 2015). In der G-Theorie wird geprüft, wie stark der gleichzeitige Einfluss verschiedener Facetten und deren jeweiligen Bedingungen auf den Messfehler eines empirisch beobachteten Werts ist (Brennan, 2001; Casale, Grosche, Volpe & Henemann, 2017).

Methodisch wird in der Generalisierbarkeitstheorie wie folgt vorgegangen: Im Rahmen einer Generalisierbarkeitsstudie (G-Studie) werden die einzelnen Varianzkomponenten und deren Interaktion mittels einer Varianzanalyse geschätzt, um deren Einfluss auf die Messgenauigkeit festzustellen. Im nächsten Schritt wird auf Grundlage der Ergebnisse eine Entscheidungsstudie (decision study, D-Studie) durchgeführt. Hierbei handelt es sich um eine Simulationsstudie, in der analysiert wird, wie sich die Varianzaufklärung verändern würde, wenn die Anzahl zulässiger Bedingungen innerhalb einer bestimmten Facetten variiert (Casale et al., 2017). Für die dieser Studie zugrundeliegende Fragestellung wird demnach die Anzahl an zu beobachtenden Sekunden in der Facette Zeitpunkt variiert, um zu prüfen, bis zu welcher Sekundenanzahl die Beobachtung des verhaltensbezogenen *engagements* noch immer als reliabel und zuverlässig gilt.

Zur Bewertung der Testgüte werden zwei Kennwerte ermittelt. Um die Reliabilität eines Beobachtungsverfahrens beurteilen zu können, kann ein Generalisierbarkeitskoeffizient (ϵp^2) berechnet werden. Er gibt an, wie viel Prozent der relativen Fehlervarianz (σ_{rel}^2) durch das Modell erklärt werden kann (Casale et al., 2017). Der G-Koeffizient, auch als relativer Fehlerkoeffizient bekannt, ermöglicht es, vor allem normorientierte Gruppenvergleiche anzustellen, da er auf der Rangfolge (Relation) der untersuchten Personen basiert und nur die Varianzkomponenten berücksichtigt, die zum relativen Stand der Differenzierungsfacette (hier Schülerinnen und Schüler) beitragen. Der G-Koeffizient kann entsprechend des Reliabilitätskoeffizienten der KTT interpretiert werden und lässt sich aus dem Anteil der universe score-Varianz (σ_p^2) an der Gesamtvarianz, die sich aus der Summe der universe score Varianz und der relativen Fehlervarianz zusammensetzt ($\sigma_p^2 + \sigma_{rel}^2$), berechnen (Casale et al., 2015; Renkl & Helmke, 1993):

$$(1) \epsilon p^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{rel}^2}$$

Bezogen auf die vorliegende Studie erlaubt ein genügend hoher G-Koeffizient einen relativen Vergleich des verhaltensbezogenen *engagements* der individuellen Schülerinnen und Schüler untereinander. Um die dieser Studie zugrundeliegende Fragestellung beantworten zu können, ist jedoch der zweite Kennwert von größerer Bedeutung. Der Zuverlässigkeitsindex Φ (index of dependability oder D-Koeffizient) gibt an, wie viel Prozent der absoluten Fehlervarianz (σ_{abs}^2) durch das Modell erklärt werden kann. Bezüglich dieser Studie bedeutet dies, dass das verhaltensbezogene *engagement* der Schülerinnen und Schüler je nach Unterrichtszeitpunkt anders ausfallen kann, weshalb die Berücksichtigung der absoluten Fehlervarianz für die Berechnung der Testgüte notwendig ist. Dadurch werden intra-individuelle Aussagen über den Verlauf des verhaltensbezogenen *engagements* der Schülerinnen und Schüler innerhalb einer Unterrichtsstunde möglich. Der D-Koeffizient Φ wird entsprechend über das Verhältnis der universe score-Varianz der Person (bzw. der Facette der Differenzierung) (σ_p^2) zur Summe aus der Varianz dieses universe scores und der absoluten Fehlervarianz (σ_{abs}^2) berechnet (Eisend, 2007):

$$(2) \Phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{abs}^2}$$

Damit eignet sich der D-Koeffizient als Kriterium für die Beurteilung der Reliabilität für verlaufsspezifische Beobachtungen (Casale et al., 2017).

Beide Koeffizienten können Werte zwischen 0 und 1 (0% Varianzaufklärung bis 100% Varianzaufklärung durch das Modell) annehmen (Brennan, 2001). Laut der Reliabilitätsdefinition der KTT und Standards der verlaufsdagnostischen Verfahren wird ein Mindestwert von .70 (Casale et al., 2017; Jentsch et al., 2019; Renkl & Helmke, 1993) bzw. .80 (Casale et al., 2015; Volpe & Briesch, 2012) gewünscht.

Da der Unterrichtsstunde drei verschiedene aktive Arbeitsphasen zugrunde liegen und diese verschieden lang andauern, liegt ein unbalanciertes Design vor. Zur Überprüfung der Fragestellung werden daher für alle drei Arbeitsphasen jeweils separate Ein-Facetten Generalisierbarkeitsstudien durchgeführt ($s_{id} \times sec$) (Chiu & Wolfe, 2002; Li & Lautenschlager, 1997). Die Schülerinnen und Schüler (s_{id}) bilden die Differenzierungsfacette.

Die Differenzierungsfacette wird als zufällige Facette erachtet, da die Schülerinnen und Schüler nur eine Auswahl aus allen möglichen Bedingungen eines Universums darstellen und die erzielten Befunde über dieses Universum generalisiert werden. Ebenso handelt es sich bei der

Facette Zeitpunkt in Sekunden um zufällige Facetten, da diese nur eine Auswahl an andauernden Unterrichtssituationen in verschiedenen Arbeitsphasen darstellen.

Die Datenanalyse wird mit dem R-Paket „gtheory“ (Moore, 2016) der Software R durchgeführt. Im ersten Schritt (G-Studie) werden die Varianzkomponenten der einzelnen Facetten für jede der drei Arbeitsphasen geschätzt. Die Varianzkomponenten der jeweiligen G-Studien werden dann in eine gewichtete Varianzkomponentenschätzung überführt (Chiu & Wolfe, 2002). Auf Grundlage der Befunde wird eine D-Studie über die Facette Zeitpunkt in Sekunden berechnet, um zu überprüfen, ob durch eine geringere Anzahl an zu beobachteten Sekunden das verhaltensbezogene *engagement* ebenso zuverlässig erfasst werden kann. Die Testgüte wird jeweils über den G- und D-Koeffizient ermittelt.

Ergebnisse

Deskriptive Befunde

Die deskriptiven Befunde über die drei aktiven Arbeitsphasen für alle Schülerinnen und Schüler hinweg sind in Tabelle 3 gegeben. Im Durchschnitt zeigen die Schülerinnen und Schüler über alle drei Arbeitsphasen hinweg am häufigsten ein passives aufgabenbezogenes Verhalten (*engaged*), was sich vor allem über aufmerksames Verhalten, wie Zuhören oder dem Unterrichtsmittelpunkt zugewandt sein, äußert. Die Schülerinnen und Schüler variieren in ihrem Verhalten bezüglich des aktiven Verhaltens (*engaged* fremd- und selbst-initiiert) sehr stark untereinander. Es zeigt sich durchschnittlich über alle Phasen hinweg, eine große Streuung für die beiden Kategorien, die ein aktives verhaltensbe-

zogenes *engagement* umfassen. Insgesamt wird jedoch deutlich, dass die Schülerinnen und Schüler im Durchschnitt über alle drei aktiven Arbeitsphasen hinweg knapp 75 % der Zeit ein aufgabenbezogenes *engagement* zeigen.

G-Studie

Im Folgenden werden die Schätzungen der Varianzkomponenten des verhaltensbezogenen *engagements* dargestellt (Tabelle 4). In der ersten Arbeitsphase (Partnerarbeit) klären die Schülerinnen und Schüler 20.16 % der Gesamtvarianz auf, während in der Gruppenarbeitsphase (Arbeitsphase 2) lediglich 11.47 % aufgeklärt werden. Die Gesamtvarianz in der Sicherungsphase (Arbeitsphase 3) wird hingegen zu einem größeren Anteil (32.73 %) über die Schülerinnen und Schüler aufgeklärt. Das verhaltensbezogene *engagement* der Schülerinnen und Schüler variiert demnach zwischen den Lernenden in der Sicherungsphase deutlicher als in den anderen Arbeitsphasen.

Der Zeitpunkt in Sekunden klärt sowohl während der Partnerarbeitsphase (1.96 %) als auch während der Gruppenarbeitsphase (2.47 %) nur einen unbedeutenden Anteil an Varianz des verhaltensbezogenen *engagements* auf. Es zeigen sich demnach in den ersten beiden aktiven Arbeitsphasen im zeitlichen Verlauf kaum Unterschiede im verhaltensbezogenen *engagement* der Lernenden. In der Sicherungsphase kann der Zeitpunkt jedoch 6.58 % der Gesamtvarianz aufklären, was für moderate Unterschiede im verhaltensbezogenen *engagement* im Verlauf der Ergebnissicherung spricht. Die bisher durch die Modelle verbleibende unerklärte Varianz ist in allen drei Arbeitsphasen vergleichsweise groß (Partnerarbeitsphase 77.88 %, Sicherungsphase 60.69 %), wobei in der Gruppenarbeitsphase die größte unerklärte Varianz (86.06 %) besteht.

Tabelle 3. Deskriptive Werte in Prozent des verhaltensbezogenen *engagements* über die drei aktiven Arbeitsphasen für alle Schülerinnen und Schüler hinweg ($N = 20$)

	Arbeitsphase 1 Partnerarbeit (545 Sekunden)	Arbeitsphase 2 Gruppenarbeit (690 Sekunden)	Arbeitsphase 3 Ergebnissicherung im Plenum (284 Sekunden)	Gesamt über alle Phasen hinweg
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
<i>Disengaged</i> Aktiv/störend	8.20 (7.15)	17.38 (10.87)	1.90 (4.39)	11.01 (6.09)
<i>Disengaged</i> Passiv	14.55 (12.52)	14.73 (14.54)	5.84 (12.24)	12.86 (10.43)
<i>Engaged</i> Passiv	23.85 (8.08)	32.48 (13.72)	78.24 (21.38)	38.74 (9.25)
<i>Engaged</i> Aktiv fremd-initiiert	28.83 (16.59)	21.41 (18.06)	6.02 (8.47)	20.89 (13.02)
<i>Engaged</i> Aktiv selbst-initiiert	24.57 (23.21)	14.00 (14.00)	8.00 (13.46)	16.50 (14.58)

Tabelle 4. Ergebnisse der Varianzkomponentenschätzung der G-Studie mit s_id x sec-Design des verhaltensbezogenen engagements für jede der drei Arbeitsphasen

Facette	Arbeitsphase 1 Partnerarbeit		Arbeitsphase 2 Gruppenarbeit		Arbeitsphase 3 Ergebnissicherung im Plenum		Gewichtet über alle Phasen hinweg	
	VC	VC (%)	VC	VC (%)	VC	VC (%)	VC	VC (%)
Lernende (s_id)	0.31	20.16	0.19	11.47	0.17	32.73	0.22	18.05
Zeitpunkt in Sekunden (sec)	0.03	1.96	0.04	2.47	0.03	6.58	0.03	2.83
Residuum	1.20	77.88	1.40	86.06	0.31	60.69	0.97	79.12
E_p^2	0.99		0.99		0.99		0.99	
ϕ	0.99		0.99		0.99		0.99	

Anmerkung: VC = Varianzkomponentenschätzung.

Über alle drei Phasen hinweg zeigt sich ein ähnliches Bild. Die gewichtete Schätzung der Varianzkomponente für die Schülerinnen und Schüler klärt 18.05%, der Zeitpunkt in Sekunden 2.83% an der Gesamtvarianz des verhaltensbezogenen engagements auf. 79.12% der Varianz bleiben unerklärt. Über alle Arbeitsphasen hinweg ist die Variation im verhaltensbezogenen engagement demnach zu einem moderaten Anteil auf die unterschiedlichen Schülerinnen und Schüler zurückzuführen, über den zeitlichen Verlauf der Stunde (Zeitpunkt in Sekunden) zeigen sich hingegen kaum Unterschiede im verhaltensbezogenen engagement der Lernenden. Der G-Koeffizient e_p^2 als auch der D-Koeffizient ϕ liegt bei .99, was für eine sehr hohe Generalisierbarkeit und Zuverlässigkeit der Beobachtungen spricht.

D-Studie

Im Rahmen der Entscheidungsstudie wurden die Bedingungen innerhalb der Facette Zeitpunkt in Sekunden variiert, um zu überprüfen, bei welcher Anzahl an zu beobachtenden Sekunden noch immer eine hohe Zuverlässigkeit der Beobachtungen erzielt werden kann. Die Ergebnisse zeigen, dass mit nur 30 Sekunden das verhaltensbezogene engagement der Schülerinnen und Schüler zuverlässig beobachtet werden kann (Abbildung 2). Der D-Koeffizient ϕ liegt bei .87.

Diskussion

Zusammenfassung und Interpretation der Ergebnisse

Videostudien werden in der pädagogisch-psychologischen Unterrichtsforschung vermehrt eingesetzt, da sie eine objektive Perspektive auf den Unterricht ermöglichen, um

Unterrichtsprozesse zu beschreiben und zu analysieren (Helmke, 2009; Herrle & Breitenbach, 2016; Pauli & Reusser, 2006). Unterrichtsvideos ermöglichen es, die Situation zu kontextualisieren (König, 2015) und dadurch eine „natürliche“ Klassenraumsituation annäherungsweise darzustellen (Casale, Strauß, Hennemann & König, 2016). Im Rahmen dieser schülerzentrierten Videostudie wurde untersucht, in welchem Ausmaß (Anzahl an Sekunden) Schülerinnen und Schüler während einer Unterrichtsstunde (Jahrgangsstufe drei) mithilfe von Videoanalysen beobachtet werden müssen, um das verhaltensbezogene engagement (behavioral und agentic engagement) als äußere sichtbare Lernaktivitäten, die den Verlauf und die Gestaltung von Unterricht beeinflussen, reliabel messen zu können. Hierfür wurden allen Schülerinnen und Schüler individuell über alle drei aktiven Arbeitsphasen (Partnerarbeit, Gruppenarbeit, Ergebnissicherung im Plenum) der vorliegenden Deutschstunde mithilfe des für diese Studie adaptierten Münchner Aufmerksamkeitsinventars (Helmke, 1988) in ihrem verhaltensbezogenen engagement von vier unabhängigen Kodierenden bewertet. Zur Beantwortung der Fragestellung wurde der methodische Ansatz der Generalisierbarkeitstheorie (Brennan, 2001) herangezogen.

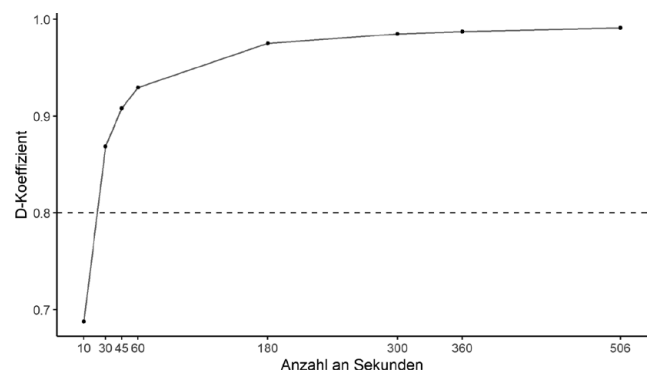


Abbildung 2. Entwicklung des D-Koeffizienten in Abhängigkeit der Anzahl an Zeitpunkten in Sekunden.

gen, um zum einen den Einfluss verschiedener Fehlerquellen auf die Gesamtvarianz simultan zu schätzen. Zum anderen, um durch die Simulation der Bedingungen innerhalb der Facette Zeitpunkt zu überprüfen, bis zu welcher Anzahl an Sekunden eine noch immer ausreichend hohe Reliabilität in der Beobachtung besteht. Die Ergebnisse leisten einen Beitrag zur effizienten und ökonomischen Designplanung von Videostudien zum verhaltensbezogenen *engagement* von Lernenden im Unterricht.

Die Ergebnisse der G-Studie zeigen, dass der Großteil der erklärten Varianz über alle Arbeitsphasen hinweg (Partnerarbeit, Gruppenarbeit, Sicherung im Plenum) auf die Facette der Schülerinnen und Schüler (knapp 20 %) zurückzuführen ist. Die Unterschiede im verhaltensbezogenen *engagement* der Schülerinnen und Schüler können demnach zu einem bedeutsamen Anteil durch die individuellen Reaktionen der Lernenden auf ihre Umgebung erklärt werden. Dieses Ergebnis ist nicht verwunderlich, da Studien bereits gezeigt haben, dass das verhaltensbezogene *engagement* von Lernendem zu Lernendem unterschiedlich ausfällt (Downer et al., 2007) und auch im Verlauf einer Unterrichtsstunde variieren kann (Fredricks et al., 2004). Es handelt sich um einen individuellen Prozess, der je nach Kontext individuelle Verhaltensweisen der Schülerinnen und Schüler auslöst (Downer et al., 2007).

Im zeitlichen Verlauf der Stunde (Zeitpunkt in Sekunden) wurden kaum Unterschiede in dem verhaltensbezogenen *engagement* der Lernenden (knapp 3% Varianzaufklärung) festgestellt. Dies kann dahingehend interpretiert werden, dass über den Verlauf der aktiven Arbeitsphasen der Zeitpunkt nicht ausschlaggebend dafür ist (keinen systematischen Einfluss hat), wie die Lernenden sich verhalten. Die geringe Varianzaufklärung ist ebenfalls ein erwartungskonformer Befund. In den Partner- und Gruppenarbeiten sollten die Schülerinnen und Schüler weitestgehend selbständig und eigenverantwortlich arbeiten (Pauli & Reusser, 2000). Über den Verlauf der aktiven Arbeitsphasen (Partner- und Gruppenarbeit) hinweg gibt es kaum Zeitpunkte, an denen alle Schülerinnen und Schüler das gleiche Verhalten zeigen (sollten). Das Aufmerksamkeitsverhalten der Schülerinnen und Schüler (als ein Indikator des *behavioral engagements*) zeichnet sich durch einen wellenförmigen Verlauf aus – nach Phasen hoher Aufmerksamkeit folgen Phasen niedriger Aufmerksamkeit (Hommel, 2012b). Besonders bei Schülerinnen und Schülern im Grundschulalter ist die Fähigkeit, irrelevante Reize zu unterdrücken, die die Aufmerksamkeit auf sich ziehen können, noch nicht weit ausgebildet (Hasselhorn & Gold, 2017). Welcher Reiz jedoch wann auf welche Schülerin bzw. welchen Schüler trifft, hängt von den individuellen Lernenden ab, was zu einem unterschiedlichen Verhalten in der jeweiligen Situation führt.

Die Lernenden variieren demnach in ihrem Durchhaltevermögen und ihrer Ablenkungstendenz.

In Hinblick auf den hohen Anteil der unerklärten Varianz (knapp 80 %) sind anknüpfende Generalisierbarkeitsstudien sinnvoll, die weitere, in der vorliegenden Studie nicht miteinbezogene, Facetten spezifizieren. Die hohe Residualvarianz kann dahingehend interpretiert und diskutiert werden, dass das Konstrukt des verhaltensbezogenen *engagements* nicht ausschließlich über die individuellen Schülerinnen und Schüler erklärt werden kann, sondern Kontextvariablen einen weiteren Einfluss auf das *engagement* haben. Unterrichtliche Prozesse zeichnen sich vor allem über die Interaktion von Lernenden, Lehrenden und dem fachlichen Inhalt aus (Breidenstein, 2010; Vieluf et al., 2020). Eine mögliche weitere Facette bildet daher das Lehrpersonenhandeln im Unterricht. Besonders in Partner- und Gruppenarbeitsphasen kommt der Lehrkraft eine entscheidende Rolle zu. Nicht alle Schülerinnen und Schüler nutzen in kooperativen Lernphasen die Unterrichtszeit und die zur Verfügung gestellten Lernhilfen gleichermaßen produktiv (Pauli & Reusser, 2000). Die Rolle der Lehrperson während des Unterrichts ist daher zum einen die der „Organisatorin und Managerin“, um einen reibungslosen Ablauf zu ermöglichen und dadurch die aufgabenbezogene Lernzeit (*time on task*) zu maximieren. Zum anderen nehmen Lehrpersonen die Rolle der „adaptiven Lernberaterin“ ein, um situationssensibel die Zusammenarbeit der Schülerinnen und Schüler zu unterstützen und den sachbezogenen Lerndialog zwischen den Gruppenmitgliedern zu intensivieren (Pauli & Reusser, 2000, S. 431–434). In weiteren Generalisierbarkeitsstudien sollte daher auf Interaktionsebene ebenso das Lehrpersonenhandeln erfasst und in die Analysen als weitere Facette einbezogen werden.

Aber nicht nur die Lehrpersonen, auch die Zusammensetzung der Schülerinnen- und Schülergruppen könnte das verhaltensbezogene *engagement* beeinflussen. Merkmale auf Gruppenebene, wie die Gruppenzusammensetzung hinsichtlich Leistung oder Geschlecht, zeigten sich bereits in zahlreichen Studien als prädiktive Variablen für das Interaktionsverhalten von Lernenden (Lipowsky, Herrmann, Ludwig, Eckermann & Heinzel, 2013; Webb, 1989). Dies könnte mit Bezug auf Ergebnisse bei Ranger (2017) sowie Nemeth, Denn, Hirstein und Lipowsky (2019) dahingehend von Bedeutung sein, als dass in Gruppen, in denen überwiegend leistungsstarke Lernende zusammenarbeiten, häufiger und über einen längeren Zeitraum hinweg ein aufgabenbezogenes aktives Verhalten gezeigt wird, während in heterogen zusammengesetzten Gruppen das Verhalten der Schülerinnen und Schüler stärker variiert. Die Zusammensetzung der Gruppe könnte demnach einen zusätzlichen Beitrag zur Erklärung der Unterschiede im verhaltensbezogenen *engagement* leisten.

Einen weiteren Beitrag zur unerklärten Varianz könnten die mentalen Dimensionen des Konstrukts *classroom engagement* leisten. Die vier Dimensionen des *classroom engagements* stehen in einer reziproken Beziehung zueinander und beeinflussen sich gegenseitig (Fredricks et al., 2004; Helmke & Renkl, 1992; Lam et al., 2014). Um mehr Informationen über alle Dimensionen des *classroom engagements* während des Unterrichts zu erhalten, könnte in weiterführenden Videostudien über ein Experience Sampling Verfahren zusätzlich das *cognitive* und *emotional engagement* der Schülerinnen und Schüler im Verlauf des Unterrichts zu mehreren Zeitpunkten individuell erfasst werden. Auf ein bestimmtes Signal hin bearbeiten die Lernenden einen auf Selbsteinschätzung basierenden Fragebogen über ihre derzeitige Aktivität und ihre kognitive und affektive Involviertheit. Dieses Vorgehen ermöglicht es, detaillierte Daten über das in der jeweiligen Situation entsprechende *classroom engagement* zu gewinnen (Fredricks & McColskey, 2012) und könnte ebenso eine weitere Facette in einer anschließenden Studie darstellen.

Im Anschluss an die G-Studie wurde eine D-Studie durchgeführt, in welcher eine Veränderung der Bedingungen in der Facette Zeitpunkt in Sekunden simuliert wurde. Die Ergebnisse aus der D-Studie weisen auf eine hohe Generalisierbarkeit ($E_p^2 = .99$) und eine hohe Zuverlässigkeit ($\phi = .99$) der Befunde hin. Der festgelegte kritische Wert von $.70$ bzw. $.80$ wird weit überschritten. Die Entscheidungsstudie ergab, dass bei lediglich 30 beobachteten Sekunden ein zuverlässiges Ergebnis erzielt werden konnte ($\phi = .87$). Da die G-Studien für die jeweiligen Arbeitsphasen getrennt jedoch gezeigt haben, dass der Zeitpunkt in der Phase der Ergebnissicherung etwas mehr Varianz aufklärt als in den anderen beiden aktiven Arbeitsphasen, wird daraus abgeleitet, dass eine Erfassung des verhaltensbezogenen *engagements* der Schülerinnen und Schüler sinnvollerweise in jeder Arbeitsphase erfolgen sollte. Dies lässt sich dadurch erklären, dass in der Ergebnissicherung zu bestimmten Zeitpunkten (z.B. immer dann, wenn von der Lehrkraft eine Frage gestellt wurde) ein spezifisches Verhalten aller Schülerinnen und Schüler erwartet werden kann (eine Meldung, um die Frage der Lehrkraft zu beantworten). Der Zeitpunkt in Sekunden hat in der Ergebnissicherung demnach einen systematischeren Einfluss auf das verhaltensbezogene *engagement* der Lernenden als in anderen aktiven Arbeitsphasen.

Zusammenfassend kann aus den Ergebnissen dieser Studie folgende praktische Implikation abgeleitet werden: Um das verhaltensbezogene *engagement* im Unterricht reliabel erfassen zu können, genügt es auf Basis dieser Studie, jede Schülerin bzw. jeden Schüler im Unterricht insgesamt 30 Sekunden – die nicht zwangsläufig zusammenhängen müssen, im Idealfall sogar zufällig gewählt wurden – in ihrem Verhalten mithilfe des adaptierten

Münchener Aufmerksamkeitsinventars (Helmke, 1988) zu beobachten. Bei der Auswahl der zu beobachtenden Zeitpunkte im Verlauf des Unterrichts sollte jedoch darauf geachtet werden, dass in allen sich theoretisch substantiell unterscheidenden Arbeitsphasen (im Sinne didaktischer Unterrichtsplanung denke man hier bspw. an die Phasen „Einführung, Erarbeitung, Übung, Sicherung“) eine Beobachtung vorgenommen wird, um den zwar geringen aber dennoch moderaten systematischen Unterschied im verhaltensbezogenen *engagement* in Plenumsphasen und Partner- bzw. Gruppenphase zu berücksichtigen.

Limitationen

Ein Kritikpunkt, der im Kontext der Generalisierbarkeitstheorie adressiert wird, ist die Aussagekraft der Zuverlässigkeit der Ergebnisse über die gewählten Bedingungen der Stichprobe hinaus (Casale et al., 2016). Die G-Theorie wird als eine „Stichprobentheorie“ bezeichnet, deren ausgewählte Bedingungen innerhalb der Facetten immer von dem Erkenntnisinteresse gelenkt werden (Shavelson & Webb, 1991). Die Ergebnisse sind daher auch nur in diesem Kontext bzw. in Bezug auf das jeweilige Universum zulässiger Bedingungen generalisierbar. Durch eine mehrfache Durchführung von G-Studien in unterschiedlichen Kontexten kann die Zuverlässigkeit der Befunde gesteigert werden. Bezüglich des Erkenntnisinteresses der vorliegenden Studie könnten die Bedingungen der Facetten bezüglich der beobachteten Lernenden, des Faches, des Unterrichtstages oder sogar der Klassenstufe oder Schulform variiert werden, um die Ergebnisse zu replizieren bzw. zu widerlegen.

An dieser Stelle sei zudem darauf hingewiesen, dass der hohe Reliabilitätskoeffizient Phi ($\phi = .99$) auch aufgrund der hohen Anzahl an Items (hier Zeitpunkt in Sekunden), die in die Berechnung eingehen, zustande kommen kann (Brennan, 2001; Smith, 1978). Bei der Interpretation der Ergebnisse muss dies berücksichtigt werden.

Eine zusätzliche Limitation dieser Studie zeichnet sich in der lediglich moderaten Beurteilerinnenübereinstimmung ab. Obwohl die Raterinnen intensiv geschult worden sind, konnte lediglich eine moderate bis gute Übereinstimmung erzielt werden (Cohen's Kappa zwischen $k = .64$ und $k = .68$, Cohen's Weighted Kappa zwischen $k = .70$ und $k = .75$). Die Ergebnisse müssen mit Bedacht interpretiert werden, da in einigen Studien bereits gezeigt wurde, dass der Einfluss der Raterinnen und Rater auf die Einschätzung von Verhaltensweisen (Casale et al., 2015) oder von Unterrichtsqualitätsmerkmalen (Praetorius, 2014) nicht unbedeutend ausfallen kann. Ein Grund für die geringe Übereinstimmung könnte womöglich an der nicht gänzlich trennscharfen Unterscheidung eines selbst-initiierten

oder fremd-initiierten aktiven sowie eines passiven aufgabenbezogenen Verhaltens (*engaged*) in kooperativen Arbeitsphasen liegen. Kooperative Arbeitsphasen sind dynamisch und von vielen unterschiedlichen Verhaltensweisen der Schülerinnen und Schüler geprägt. Ein so kleinschrittiges Zeitintervall (Ein-Sekunden-Intervall), wie es dieser Studie zugrunde liegt, birgt womöglich die Gefahr, dass spezifische Verhaltensweisen von manchen Kodierenden erkannt, andere wiederum schnell übersehen werden. Grundsätzlich kann jedoch festgehalten werden, dass die schülerzentrierten Unterrichtsaufnahmen (Paulicke et al., 2015) es ermöglichen, die Schülerinnen und Schüler individuell über die gesamte Unterrichtszeit hinweg zu beobachten. Ähnlich wie das Thin-Slice-Verfahren (Begrich et al., 2017) erscheint auch dieser Ansatz als eine vielversprechende und ökonomische Ergänzung der Unterrichtsforschung, wobei weitere Einsatzmöglichkeiten in anknüpfenden Studien erprobt werden sollten.

Literatur

- Appleton, J. J., Christenson, S. L. & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45 (5), 369–386. <https://doi.org/10.1002/pits.20303>
- Appleton, J. J. & Lawrenz, F. (2011). Student and Teacher Perspectives Across Mathematics and Science Classrooms: The Importance of Engaging Contexts. *School Science and Mathematics*, 111 (4), 143–155. <https://doi.org/10.1111/j.1949-8594.2011.00072.x>
- Bangert-Drowns, R. L. & Pyke, C. (2001). A Taxonomy of Student Engagement with Educational Software: An Exploration of Literate Thinking with Electronic Text. *Journal of Educational Computing Research*, 24 (3), 213–234.
- Begrich, L., Fauth, B., Kunter, M. & Klieme, E. (2017). Wie informativ ist der erste Eindruck? Das Thin-Slices-Verfahren zur videobasierten Erfassung des Unterrichts. *Zeitschrift für Erziehungswissenschaft*, 20 (1), 23–47. <https://doi.org/10.1007/s11618-017-0730-x>
- Beserra, V., Nussbaum, M. & Oteo, M. (2019). On-Task and Off-Task Behavior in the Classroom: A Study on Mathematics Learning With Educational Video Games. *Journal of Educational Computing Research*, 56 (8), 1361–1383. <https://doi.org/10.1177/0735633117744346>
- Bloom, B. S. (1974). Time and learning. *American Psychologist*, 29 (9), 682–688. <https://doi.org/10.1037/h0037632>
- Breidenstein, G. (2010). Überlegungen zu einer Theorie des Unterrichts. *Zeitschrift für Pädagogik*, 56 (6), 869–887.
- Brennan, R. L. (2001). *Generalizability Theory* (Statistics for Social Sciences and Public Policy). New York: Springer New York. <https://doi.org/10.1007/978-1-4757-3456-0>
- Büttner, G. & Schmidt-Atzert, L. (2004). Diagnostische Verfahren zur Erfassung von Konzentration und Aufmerksamkeit. In G. Büttner & L. Schmidt-Atzert (Hrsg.), *Diagnostik von Konzentration und Aufmerksamkeit* (S. 23–62). Göttingen: Hogrefe.
- Cadima, J., Leal, T. & Burchinal, M. (2010). The quality of teacher-student interactions: associations with first graders' academic and behavioral outcomes. *Journal of School Psychology*, 48 (6), 457–482. <https://doi.org/10.1016/j.jsp.2010.09.001>
- Cappella, E., Kim, H. Y., Neal, J. W. & Jackson, D. R. (2013). Classroom Peer Relationships and Behavioral Engagement in Elementary School: The Role of Social Network Equity. *American Journal of Community Psychology*, 52 (3–4), 367–379. <https://doi.org/10.1007/s10464-013-9603-5>
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64 (8), 723–733.
- Carter, C. P., Reschly, A. L., Lovelace, M. D., Appleton, J. J. & Thompson, D. (2012). Measuring student engagement among elementary students: Pilot of the Student Engagement Instrument-Elementary Version. *School Psychology Quarterly*, 27 (2), 61–73. <https://doi.org/10.1037/a0029229>
- Casale, G., Grosche, M., Volpe, R. J. & Hennemann, T. (2017). Zuverlässigkeit von Verhaltensverlaufsdiagnostik über Rater und Messzeitpunkte bei Schülern mit externalisierenden Verhaltensproblemen. *Empirische Sonderpädagogik*, (2), 143–164.
- Casale, G., Hennemann, T., Volpe, R. J., Briesch, A. M. & Grosche, M. (2015). Generalisierbarkeit und Zuverlässigkeit von Direkten Verhaltensbeurteilungen des Lern- und Arbeitsverhaltens in einer inklusiven Grundschulklasse. *Empirische Sonderpädagogik*, (3), 258–268.
- Casale, G., Strauß, S., Hennemann, T. & König, J. (2016). Wie lässt sich Klassenführungsexpertise messen? Überprüfung eines videobasierten Erhebungsinstruments für Lehrkräfte unter Anwendung der Generalisierbarkeitstheorie. *Empirische Sonderpädagogik*, (2), 119–139.
- Chiu, C. W. T. & Wolfe, E. W. (2002). A Method for Analyzing Sparse Data Matrices in the Generalizability Theory Framework. *Applied Psychological Measurement*, 26 (3), 321–338. <https://doi.org/10.1177/0146621602026003006>
- Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70 (4), 213–220. <https://doi.org/10.1037/h0026256>
- Decristan, J., Fauth, B., Heide, E. L., Locher, F. M., Troll, B., Kurucz, C. et al. (2019). Individuelle Beteiligung am Unterrichtsgespräch in Grundschulklassen: Wer ist (nicht) beteiligt und welche Konsequenzen hat das für den Lernerfolg? *Zeitschrift für Pädagogische Psychologie*, 1–16. <https://doi.org/10.1024/1010-0652/a000251>
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (Springer-Lehrbuch, 5. Aufl.). Berlin: Springer. <https://doi.org/10.1007/978-3-642-41089-5>
- Downer, J. T., Rimm-Kaufman, S. E. & Pianta, R. C. (2007). How Do Classroom Conditions and Children's Risk for School Problems Contribute to Children's Behavioral Engagement in Learning? *School Psychology Review*, 36 (3), 413–432.
- Eccles, J. & Wang, M.-T. (2012). So What Is Student Engagement Anyway? In S. L. Christenson, A. L. Reschly & C. Wylie (Hrsg.), *Handbook of Research on Student Engagement* (133–145). Boston, MA: Springer US. <https://doi.org/10.1007/978-1-4614-2018-7>
- Eisend, M. (2007). *Methodische Grundlagen und Anwendungen der Generalisierbarkeitstheorie in der betriebswirtschaftlichen Forschung* (Diskussionsbeiträge des Fachbereichs Wirtschaftswissenschaft der Freien Universität Berlin Betriebswirtschaftliche Reihe, Bd. 4). Berlin: Freie Univ. Fachbereich Wirtschaftswiss.
- Fend, H. (2019). Erklärungen von Unterrichtserträgen im Rahmen des Angebot-Nutzungs-Modells. In U. Steffens & R. Messner (Hrsg.), *Unterrichtsqualität. Konzepte und Bilanzen gelingenden Lehrens und Lernens. Grundlagen der Qualität von Schule 3* (Beiträge zur Schulentwicklung, 1. Aufl., S. 91–104). Münster: Waxmann.
- Fredricks, J. A., Blumenfeld, P. C. & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review*

- of *Educational Research*, 74 (1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A. & McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In S. L. Christenson, A. L. Reschly & C. Wylie (Hrsg.), *Handbook of Research on Student Engagement* (763–782). Boston, MA: Springer US. <https://doi.org/10.1007/978-1-4614-2018-7>
- Gettinger, M. & Walter, M. J. (2012). Classroom Strategies to Enhance Academic Engaged Time. In S. L. Christenson, A. L. Reschly & C. Wylie (Hrsg.), *Handbook of Research on Student Engagement* (653–673). Boston, MA: Springer US. <https://doi.org/10.1007/978-1-4614-2018-7>
- Hasselhorn, M. & Gold, A. (2017). *Pädagogische Psychologie. Erfolgreiches Lernen und Lehren* (Standards Psychologie, 4. Aufl.). Stuttgart: Verlag W. Kohlhammer.
- Helmke, A. (1988). *Das Münchener Aufmerksamkeitsinventar (MAI). Manual für die Beobachtung des Aufmerksamkeitsverhaltens von Grundschulern während des Unterrichts*. Paper 6/1988 (Max-Planck-Institut für psychologische Forschung, Hrsg.). München.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts* (Unterricht verbessern – Schule entwickeln, 3. Aufl.). Seelze-Velber: Klett/Kallmeyer.
- Helmke, A. & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica*, 38, 130–141.
- Helmke, T., Helmke, A., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K. (2008). Die Videostudie des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (Beltz Pädagogik, S.345–363). Weinheim: Beltz.
- Herrle, M. & Breitenbach, S. (2016). Planung, Durchführung und Nachbereitung videogestützter Beobachtungen im Unterricht. In U. Rauin, M. Herrle & T. Engartner (Hrsg.), *Videoanalysen in der Unterrichtsforschung. Methodische Vorgehensweisen und Anwendungsbeispiele* (Grundlagentexte Methoden, S.30–49). Weinheim: Beltz Juventa.
- Hess, M. & Denn, A.-K. (2018). Methodenworkshop. Hoch und niedrig inferente Methoden der Videoanalyse. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 38 (2), 212–222.
- Hommel, M. (2011). Aufmerksamkeitsverlauf – Fremdbeobachtung und Eigeneinschätzung. In U. Faßhauer, B. Fürstenau & E. Wuttke (Hrsg.), *Grundlagenforschung zum Dualen System und Kompetenzentwicklung in der Lehrerbildung* (Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft, S.117–129). Opladen: Verlag Barbara Budrich.
- Hommel, M. (2012a). *Aufmerksamkeitsverhalten und Lernerfolg – eine Videostudie*. Dissertation. Frankfurt am Main.
- Hommel, M. (2012b). Mehr Schüleraufmerksamkeit im Unterricht – Handlungsoptionen für Lehrende. *Berufsbildung*, 66 (138), 34–36.
- Hughes, J. N., Luo, W., Kwok, O.-M. & Loyd, L. K. (2008). Teacher-Student Support, Effortful Engagement, and Achievement: A 3-Year Longitudinal Study. *Journal of Educational Psychology*, 100 (1), 1–14. <https://doi.org/10.1037/0022-0663.100.1.1>
- Jentsch, A., Casale, G., Schlesinger, L., Kaiser, G., König, J. & Blömeke, S. (2019). Variabilität und Generalisierbarkeit von Ratings zur Qualität von Mathematikunterricht zwischen und innerhalb von Unterrichtsstunden. *Unterrichtswissenschaft*, 47 (1). <https://doi.org/10.1007/s42010-019-00061-8>
- Jurik, V., Gröschner, A. & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction*, 23, 33–42. <https://doi.org/10.1016/j.learninstruc.2012.09.002>
- König, J. (2015). Kontextualisierte Erfassung von Lehrerkompetenzen. *Zeitschrift für Pädagogik*, 61 (3), 305–309.
- Kunter, M. & Trautwein, U. (Hrsg.). (2018). *Psychologie des Unterrichts* (StandardWissen Lehramt, Bd. 3895). Paderborn: Ferdinand Schöningh.
- Lam, S.-F., Jimerson, S., Wong, B. P. H., Kikas, E., Shin, H., Veiga, F. H. et al. (2014). Understanding and measuring student engagement in school: the results of an international study from 12 countries. *School Psychology Quarterly*, 29 (2), 213–232. <https://doi.org/10.1037/spq0000057>
- Lawson, M. A. & Lawson, H. A. (2013). New Conceptual Frameworks for Student Engagement Research, Policy, and Practice. *Review of Educational Research*, 83 (3), 432–479. <https://doi.org/10.3102/0034654313480891>
- Li, M.-N. F. & Lautenschlager, G. (1997). Generalizability Theory Applied to Categorical Data. *Educational and Psychological Measurement*, 57 (5), 813–822. <https://doi.org/10.1177/0013164497057005007>
- Linnenbrink-Garcia, L., Rogat, T. K. & Koskey, K. L. K. (2011). Affect and engagement during small group instruction. *Contemporary Educational Psychology*, 36 (1), 13–24. <https://doi.org/10.1016/j.cedpsych.2010.09.001>
- Lipowsky, F. (2015). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (Springer-Lehrbuch, 2. Aufl., S.69–105). Berlin, Heidelberg: Springer Fachmedien.
- Lipowsky, F., Herrmann, M., Ludwig, M., Eckermann, T. & Heinzel, F. (2013). Textrevisionen im Grundschulunterricht – Welchen Einfluss haben die Lernumgebung und die soziale Kohäsion der Gruppe? *Unterrichtswissenschaft*, 41 (1), 38–56.
- Lipowsky, F., Pauli, C. & Rakoczy, K. (2008). Schülerbeteiligung und Unterrichtsqualität. In M. Gläser-Zikuda & J. Seifried (Hrsg.), *Lehrerexpertise. Analyse und Bedeutung unterrichtlichen Handelns* (S.67–90). Münster: Waxmann.
- Lipowsky, F., Rakoczy, K., Pauli, C., Reusser, K. & Klieme, E. (2007). Gleicher Unterricht – gleiche Chancen für alle? Die Verteilung von Schülerbeiträgen im Klassenunterricht. *Unterrichtswissenschaft*, 35 (2), 125–147.
- Lotz, M. (2016). *Kognitive Aktivierung im Leseunterricht der Grundschule*. Dissertation. Springer Fachmedien Wiesbaden GmbH. <https://doi.org/10.1007/978-3-658-10436-8>
- Moore, C. T. (2016). Apply Generalizability Theory with R (Version 0.1.2) [Computer software].
- Nemeth, L., Denn, A.-K., Hirstein, A. & Lipowsky, F. (2019). Interaktionen von SchülerInnen in kooperativen Lernsituationen. In K. Verrière & L. Schäfer (Hrsg.), *Interaktion im Klassenzimmer. Forschungsgeleitete Einblicke in das Geschehen im Unterricht* (S.51–73). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-23173-6_4
- Ophardt, D. & Thiel, F. (2015). Kompetenzen des Klassenmanagements. In S. Reh, K. Berdelmann & J. Dinkelaker (Hrsg.), *Aufmerksamkeit* (S.173–198). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-19381-6_10
- Pauli, C. (2012). Kodierende Beobachtung. In H. de Boer & S. Reh (Hrsg.), *Beobachtung in der Schule – Beobachten lernen* (S.45–63). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Pauli, C. & Reusser, K. (2000). Zur Rolle der Lehrperson beim kooperativen Lernen. *Schweizerische Zeitschrift für Bildungswissenschaften*, 22 (3), 421–442.
- Pauli, C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52 (6), 774–798.
- Paulicke, P., Schmidt, T. & Ehmke, T. (2015). „Hier werden Parallelwelten im Unterricht sichtbar“ – Multiperspektivische Unterrichts-

- videos in der universitären LehrerInnenausbildung. *SEMINAR*, (3), 15–27.
- Pekrun, R. & Linnenbrink-Garcia, L. (2012). Academic Emotions and Student Engagement. In S. L. Christenson, A. L. Reschly & C. Wylie (Hrsg.), *Handbook of Research on Student Engagement* (259–282). Boston, MA: Springer US. <https://doi.org/10.1007/978-1-4614-2018-7>
- Ponitz, C. C., Rimm-Kaufman, S. E., Grimm, K. J. & Curby, T. W. (2009). Kindergarten Classroom Quality, Behavioral Engagement, and Reading Achievement. *School Psychology Review*, 38 (1), 102–120.
- Praetorius, A.-K. (2014). *Messung von Unterrichtsqualität durch Ratings* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 90). Dissertation.
- Pröscholdt, M. V., Stumpf, E. & Schneider, W. (2011). Das Arbeitsverhalten in homogenen Begabtenklassen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43 (2), 55–67. <https://doi.org/10.1026/0049-8637/a000035>
- Ranger, G. (2017). *Kinder in kooperativen Lernphasen kognitiv aktivieren. Eine Videostudie zur Qualität der Peer-Interaktionen im naturwissenschaftlichen Sachunterricht* (Empirische Forschung im Elementar- und Primarbereich, Bd. 1). Dissertation. Bad Heilbrunn: Julius Klinkhardt
- Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, 105 (3), 579–595. <https://doi.org/10.1037/a0032690>
- Reeve, J. & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36 (4), 257–267. <https://doi.org/10.1016/j.cedpsych.2011.05.002>
- Renkl, A. & Helmke, A. (1993). Prinzip, Nutzen und Grenzen der Generalisierungstheorie. *Empirische Pädagogik*, 7 (1), 63–85.
- Reusser, K. & Pauli, C. (2010). Unterrichtsgestaltung und Unterrichtsqualität – Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht: Einleitung und Überblick. In K. Reusser (Hrsg.), *Unterrichtsgestaltung und Unterrichtsqualität. Ergebnisse einer internationalen und schweizerischen Videostudie zum Mathematikunterricht* (S.9–32). Münster: Waxmann.
- Scherzinger, M., Wettstein, A. & Wyler, S. (2018). Aggressive und nicht aggressive Unterrichtsstörungen durch Schülerinnen und Schüler sowie durch Klassen- und Fachlehrpersonen. Eine Videostudie. *Empirische Sonderpädagogik*, 10 (4), 388–407.
- Seidel, T. (2003). *Lehr-Lernskripts im Unterricht. Freiräume und Einschränkungen für kognitive und motivationale Lernprozesse; eine Videostudie im Physikunterricht* (Pädagogische Psychologie und Entwicklungspsychologie, Bd. 35). Zugl.: Kiel, Univ., Diss., 2002. Münster: Waxmann.
- Seidel, T. (2014). Angebots-Nutzungs-Modelle in der Unterrichtspsychologie. Integration von Struktur- und Prozessparadigma. *Zeitschrift für Pädagogik*, 60 (6), 850–866.
- Shavelson, R. J. & Webb, N. M. (1991). *Generalizability theory. A primer* (Measurement methods for the social sciences, vol. 1). Newbury Park, Calif.: Sage.
- Skinner, E. A., Kindermann, T. A. & Furrer, C. J. (2009). A Motivational Perspective on Engagement and Disaffection. *Educational and Psychological Measurement*, 69 (3), 493–525. <https://doi.org/10.1177/0013164408323233>
- Skinner, E. A. & Pitzer, J. R. (2012). Developmental Dynamics of Student Engagement, Coping, and Everyday Resilience. In S. L. Christenson, A. L. Reschly & C. Wylie (Hrsg.), *Handbook of Research on Student Engagement* (S.21–44). Boston, MA: Springer US. <https://doi.org/10.1007/978-1-4614-2018-7>
- Smith, P. L. (1978). Sampling Errors of Variance Components in Small Sample Multifacet Generalizability Studies. *Journal of Educational Statistics*, 3 (4), 319–346. <https://doi.org/10.3102/10769986003004319>
- Vieluf, S., Praetorius, A.-K., Rakoczy, K., Kleinknecht, M. & Pietsch, M. (2020). Angebots-Nutzungsmodelle der Wirkweise des Unterrichts: eine kritische Auseinandersetzung mit ihrer theoretischen Konzeption. *Zeitschrift für Pädagogik*, 66, Beiheft 1/20, 63–80.
- Volpe, R. J. & Briesch, A. M. (2012). Generalizability and Dependability of Single-Item and Multiple-Item Direct Behavior Rating Scales for Engagement and Disruptive Behavior. *School Psychology Review*, 41 (3), 246–261. <https://doi.org/10.1080/02796015.2012.12087506>
- Wang, Z., Bergin, C. & Bergin, D. A. (2014). Measuring engagement in fourth to twelfth grade classrooms: the Classroom Engagement Inventory. *School Psychology Quarterly*, 29 (4), 517–535. <https://doi.org/10.1037/spq0000050>
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13 (1), 21–39. [https://doi.org/10.1016/0883-0355\(89\)90014-1](https://doi.org/10.1016/0883-0355(89)90014-1)
- Wild, K.-P. (2003). Videoanalysen als neue Impulsgeber für eine praxisnahe prozessorientierte empirische Unterrichtsforschung. *Unterrichtswissenschaft*, 31 (2), 98–102.
- Wuttke, E. (2005). *Unterrichtskommunikation und Wissenserwerb. Zum Einfluss von Kommunikation auf den Prozess der Wissensgenerierung* (Konzepte des Lehrens und Lernens, Bd. 11). Zugl.: Mainz, Univ., Habil.-Schr., 2004. Frankfurt am Main: Lang.

Historie

Manuskript eingereicht: 20.03.2020

Manuskript nach Revision angenommen: 31.08.2020

Onlineveröffentlichung: 02.10.2020

Danksagung


Die Autorin und Autoren danken den drei studentischen Hilfskräften vielmals für die gewissenhafte Unterstützung bei der Kodierung der Unterrichtsvideos.

Förderung

Die Erstellung dieses Beitrags wurde durch das ZZL-Netzwerk im Rahmen der gemeinsamen „Qualitätsinitiative Lehrerbildung“ von Bund und Ländern aus Mitteln des Bundesministeriums für Bildung und Forschung gefördert (Förderkennzeichen: 01JA1603; www.leuphana.de/zzl-netzwerk).

ORCID

Bianka Troll

 <https://orcid.org/0000-0001-8229-2454>

Bianka Troll

Leuphana Universität Lüneburg
Institut für Bildungswissenschaft
Universitätsallee 1
21335 Lüneburg
bianka.troll@leuphana.de