

Der Nutzen von Rubrics (Beurteilungsraster) für das selbstregulierte Lernen am Beispiel des mathematischen Argumentierens

Robbert Smit¹, Patricia Bachmann¹, Kurt Hess¹, Thomas Birri¹ & Verena Blum

¹Pädagogische Hochschule St.Gallen, Schweiz

²Pädagogische Hochschule Zug, Schweiz

Zusammenfassung

In der Studie „Lernen mit Rubrics“ (LERU) wurde ein Beurteilungsraster (Rubric) für die formative Beurteilung des mathematischen Argumentierens während 9 Wochen in 22 Klassen der 5. und 6. Jahrgangsstufe eingesetzt. Es interessierte insbesondere die Frage, ob Schülerinnen und Schüler den Rubric hinsichtlich der Fähigkeit der Selbstregulation beim Lösen von Aufgaben zum Argumentieren als nützlich einschätzen und ob die Häufigkeit des Lehrpersonenfeedbacks mit dieser Einschätzung zusammenhängt. Mittels quantitativer Analysen wurde zuerst ermittelt, ob ein Zusammenhang zwischen kognitiven und motivationalen Aspekten der Nützlichkeit und der selbst eingeschätzten Fähigkeit zur Selbstregulation bei unseren Aufgaben zum Argumentieren sowie der Häufigkeit des Lehrpersonenfeedbacks besteht. Anschliessend liess sich mit Hilfe linearer Wachstumsmodelle prüfen, ob eine positiv eingeschätzte Nützlichkeit auch Unterschiede in der individuellen Entwicklung der Selbstregulation beim Lösen von Aufgaben erklärt. Beide Fragen lassen sich positiv beantworten. Schülerinnen und Schüler, die ihre Selbstregulationsfähigkeiten geringer einschätzten, machten zudem grössere Fortschritte im Laufe des Projekts als solche, die ihre Selbstregulationsfähigkeiten höher einschätzten. Die Ergebnisse werden hinsichtlich der praktischen Konsequenzen für Steuerung und Beurteilung des Lernprozesses diskutiert.

Schlüsselwörter: Rubric, Selbstregulation, mathematisches Argumentieren, Feedback

L'utilité de « Rubrics » (grilles d'évaluation) pour l'apprentissage autorégulé à l'aide de l'exemple du raisonnement mathématique

Résumé

Dans l'étude « Apprendre avec Rubrics » (LERU), une grille d'évaluation (Rubric) a été utilisée pour l'évaluation formative du raisonnement mathématique pendant 9 semaines dans 22 classes des 5^{ème} et 6^{ème} années. Une attention particulière a été portée à la question de savoir si les élèves considèrent Rubric comme utile en termes de capacité d'autorégulation lors de la résolution de tâches d'argumentation et si la fréquence des commentaires des enseignants est liée à cette évaluation. Au moyen d'analyses quantitatives, il a d'abord été déterminé s'il existe une corrélation entre les aspects cognitifs et motivationnels de l'utilité perçue et la capacité auto-évaluée à s'autoréguler dans des tâches d'argumentation, et la fréquence des commentaires des enseignants. Par la suite, des modèles de croissance linéaire ont été utilisés pour tester si une utilité évaluée positivement explique également les différences dans le développement individuel de l'autorégulation lors de la résolution de tâches. Les deux questions reçoivent une réponse positive. De plus, les élèves qui ont eu le sentiment que leurs

compétences d'autorégulation étaient faibles ont fait de plus grands progrès au cours du projet que ceux qui les ont évaluées plus positivement. Les résultats sont discutés en ce qui concerne les conséquences pratiques de cet outil pour le contrôle et l'évaluation du processus d'apprentissage.

The benefit of rubrics for self-regulated learning using mathematical reasoning as an example

Abstract

In the study „Learning with Rubrics“ (LERU), a rubric for the formative assessment of mathematical reasoning was used for 9 weeks in 22 classes of the 5th and 6th grade. Of particular interest was the question of whether students perceived the rubric as useful in terms of their ability to self-regulate when solving tasks for mathematical reasoning and whether the frequency of teacher feedback relates to this assessment. Quantitative analyses were first used to determine whether there was a correlation between cognitive and motivational aspects of usefulness and the self-assessed ability to self-regulate in tasks for reasoning as well as the frequency of teacher feedback. Subsequently, linear growth models were used to test whether positively assessed usefulness also explains differences in the individual development of self-regulation when solving tasks for reasoning. Both questions can be answered positively. Students who rated their self-regulation skills lower also made greater progress over the course of the project than those who rated their self-regulation skills higher. Finally, the results are discussed in terms of the practical consequences for monitoring and assessing the learning process.

Keywords: Rubric, self-regulation, mathematical reasoning, feedback

1 Einleitung

Die Entwicklung unterrichtstauglicher Instrumente zur formativen Beurteilung (formative assessment) ist dringend notwendig (Schmidinger et al., 2016; Schmidt, 2020). Schon Klafki (2007, p. 202) empfahl die Entwicklung wissenschaftlich erprobter Beobachtungshilfen für Lehrpersonen und Schülerinnen und Schüler, um spezifische Informationen über das Lernen gewinnen zu können. Damit verbunden ist auch die Absicht, den Lernenden vermehrt Mitverantwortung für die Steuerung ihrer Lernprozesse zu übertragen. Selbstregulierende Schülerinnen und Schüler achten auf ihr Denken, ihre Motivation und ihr Verhalten während des Lernens (Zimmerman, 2002). Es ist zudem eine wichtige Aufgabe, Schülerinnen und Schüler kompetent zu machen im Verstehen und Nutzen von standardorientierten Kriterien zur Selbststeuerung und Optimierung des eigenen Lernens (Wyatt-Smith & Adie, 2019). Rubrics (Beurteilungsraster) können diesbezüglich Lernenden helfen, die Qualität des Lernens an transparenten Erwartungen auszurichten, d.h. eigene Lernergebnisse entlang von inhaltlichen Kriterien und abgestuften Indikatoren der Erfüllung einzuschätzen und folgende Lernprozesse entsprechend selbst zu regulieren. Das Konzept des Rubrics existiert schon seit einiger Zeit, insbesondere und mit einem gewissen Bekanntheitsgrad im angelsächsischen Raum (Keller, 2011). Obschon Rubrics für die formative und die summative Leistungsbeurteilung eingesetzt werden, wird in diesem Beitrag die Funktion für die formative Beurteilung (formative

assessment) fokussiert, weil hier das Potenzial von Rubrics für das Lernen liegt (Andrade, 2019). „Formative Beurteilung ist ein Prozess, welcher dazu dient, kriteriengeleitet Lernhandlungen und -prozesse der Schülerinnen und Schüler zu bewerten, mit dem Ziel, dass anschließend aussagekräftige Informationen für das weitere Lernen der Schülerinnen und Schüler und das Unterrichten der Lehrperson vorliegen“ (Smit, 2009, p. 36).

In Bezug auf den zu lernenden Gegenstand und die Leistung steht im vorliegenden Beitrag exemplarisch das mathematische Argumentieren im Fokus. Mathematisches Argumentieren gehört zu den zentralen, während der obligatorischen Bildung von allen Lernenden zu erwerbenden mathematischen Kompetenzen (Brunner et al., 2022). Trotz der zentralen Bedeutung, auch als Vorstufe des Beweisens in höheren Schulstufen, liegen u.a. nur wenige Erkenntnisse vor, wie das mathematische Argumentieren bei Kindern auf der Primarstufe gefördert werden kann (Brunner, 2022). Bei solch komplexen mathematischen Aufgaben haben Lehrpersonen Mühe, die Entwicklung der Schülerinnen und Schüler zu beurteilen und die Lernenden formativ zu unterstützen (Herbert, 2021). Davidson et al. (2019) konnten aufzeigen, dass ein Rubric im Rahmen der formativen Beurteilung beim mathematischen Argumentieren lernförderlich eingesetzt werden kann.

Im theoretischen Modell von Panadero und Jonsson (2013) zur Wirkung von Rubrics (Beurteilungsrastern) auf die Leistung finden sich sowohl motivationale (wie z.B. positive Affekte) als auch kognitive moderierende Aspekte (wie z.B. Selbst-Beurteilung). Im folgenden Beitrag wird die motivationale und kognitive Nützlichkeit von Rubrics aus Sicht der Schülerinnen und Schüler untersucht. Es wird gefragt nach der Wirkung dieser Nützlichkeit auf die Selbstregulationsfähigkeit beim Lösen von Aufgaben, die u.a. „Anlässe für Entdeckungen von mathematischen Zahl- und Rechenphänomenen“ darstellen und darüber hinaus „Argumentations- bzw. Begründungspotential“ aufweisen (Bezold, 2009, p. 111).

2 Theoretischer Hintergrund

2.1 Zur Funktion von Rubrics als Instrumente des Lehrens und Lernens

Der Begriff Rubric wird in Zusammenhang mit Beurteilung verschieden verwendet, meistens jedoch für die strukturierte Beschreibung erwarteter Kompetenzen, Leistungen oder Fähigkeiten, eingetragen in Rastern. In diesen Rastern (siehe auch Abb. 2) werden inhaltliche Kriterien und die Realisierungsgrade mit Stufen der Performanz auf einer Skala von beispielsweise schwach bis hervorragend aufgelistet (Brookhart, 2013). Die Kriterien beziehen sich jeweils auf einen Kompetenzbereich oder Aufgabentypus und unterscheiden sich damit von Kompetenzrastern, welche mehrere Kompetenzbereiche integrieren und auf grössere Zeiträume ausgerichtet sind (Keller, 2011). Beurteilungsraster hingegen unterstützen den laufenden Lernprozess mit einfachen, klaren Beschreibungen von Lernzielen und unterstützen die Lehrperson beim Beurteilen und Rückmelden während komplexeren Lernsituationen wie Schreibarbeiten, mündlichen Präsentationen oder naturwissenschaftlichen Experimenten (de Leeuw, 2016; Tang et al., 2015).

Lehrpersonen sind besonders gefordert, wenn sie komplexe Leistungen – z.B. schöpferische Schreibarbeiten oder mathematische Problemlöseaufgaben – beurteilen sollen. Unsicherheiten entstehen beispielsweise, weil sich Problemlösestrategien, Lösungswege oder Argumente nicht einfach mit richtig oder falsch deklarieren lassen (Bürgermeister, 2013; Hussmann et al., 2007). Rubrics können komplexe Beurteilungssituationen wesentlich unterstützen: Sie entflechten eine konkrete Anforderungssituation – z.B. die Begründung eines Lösungsweges oder das Verfassen einer Kurzgeschichte – mit einfachen, verständlichen Kriterien (Danielson, 1997). In den Rastern werden (Teil-)Kompetenzen operationalisiert – welche über

verschiedene Aufgaben hinweg erworben werden – mit einzelnen Kriterien und Erfüllungsgraden. Bisher war es vornehmlich in der englischsprachigen Praxis verbreitet, dass Lehrpersonen selber Beurteilungsraster entwickeln. Dies geschieht allerdings mit unterschiedlicher Qualität (Jonsson & Svingby, 2007; Keller, 2011). Öfters werden Häufigkeitsgrade definiert, statt Realisierungsgrade bzw. Niveaus (vgl. Abb. 1) zu beschreiben (Brookhart, 2018). Kriterien für die Qualität von Beurteilungsrastern finden sich z.B. bei Moskal (2000). Lehrpersonen ohne Einführung bzw. Übung zur Nutzung von Rubrics sind oftmals nicht ausreichend in der Lage, solche zu entwickeln, wirkungsvoll einzusetzen und auf Rastern beruhende Beurteilungen und Bewertungen zu interpretieren (Lovorn & Rezaei, 2011; Turley & Gallagher, 2008). Rubrics für Lehrpersonen unterscheiden sich in ihrer Komplexität und Ausführlichkeit von derer für die Schülerinnen und Schüler (Danielson, 1997). So zeichnet sich der Rubric für die Lehrperson zu mathematischem Argumentieren von Herbert (2021) im Vergleich zu unserem durch eine grössere fachdidaktische Breite und Tiefe aus, die allerdings für die Nutzung durch die Schülerinnen und Schüler zu komplex wäre (English et al., 2022).

2.2 Empirische Erkenntnisse zur Wirkung des Einsatzes von Rubrics und dessen wahrgenommene Nützlichkeit für das Lehren und Lernen

Es existiert mittlerweile ein – über Länder- und Kulturgrenzen hinweg – beachtliches Spektrum an Literatur zum unterrichtspraktischen Einsatz von Rubrics (Brookhart, 2013; Danielson, 1997). Hingegen finden sich eher spärlich Publikationen mit empirisch gesicherten Erkenntnissen. Zu Letzteren zählen die Übersichten von Brookhart und Chen (2015) sowie von Panadero und Jonsson (2013), die Studien zur Wirkung von Rubrics verglichen. Brookhart und Chen (2015) hinterfragen die oftmals positiven Ergebnisse von Studien zu Rubrics allerdings. Sie argumentieren, dass Effekte eher durch das zur Verfügung stellen von transparenten Lernzielen, Kriterien und Leistungsanforderungen im Instrument vermittelt wird als durch das Beurteilungsraster als Instrument. Sie empfehlen dennoch dessen Einsatz, weil der Rubric explizit diese vermuteten Wirkungsgrößen fokussiert.

Die meisten dieser Studien beziehen sich auf die Hochschulstufe (English et al., 2022). Etliche beschäftigen sich mit der Wirkung von Rubrics auf die Reliabilität und Validität der Bewertung durch die Lehrperson. Diesbezügliche Ergebnisse fielen für die erwartete Unterstützung der Bewertungsgüte zumeist positiv aus (Cho et al., 2006; Gregori-Giralt & Menéndez-Varela, 2018; Meier et al., 2006). Zur eigentlich bedeutsamen Frage nach der Validität der Folgen aus der Beurteilung (consequential validity) für das Lernen liegen nach Brookhart und Chen (2015) erst wenige Ergebnisse und nur solche auf der Hochschulstufe vor. Da die Folgen der formativen Beurteilung auf den Lernfortschritt abzielt, ist die Beurteilung mittels Rubric dann valide, wenn die Schülerinnen und Schüler einen solchen Fortschritt aufweisen (Stobart, 2006).

Verschiedene Studien zeigen, dass ein Rubric im Rahmen des fachlichen und überfachlichen Lernens bei Lernenden wirksam eingesetzt werden kann. Schafer et al. (2001) berichten z.B. von einer Kurzintervention, in welcher High-School Lehrpersonen in vier Fächern einen Rubric einsetzten. In zweien konnten kleine Leistungseffekte bei Lernenden der Interventionsgruppe – im Vergleich zu einer Kontrollgruppe – nachgewiesen werden.

Da sich Studien zur Wirkung von Rubrics jeweils auf ein bestimmtes Fach, respektive einen Fachinhalt beziehen, sind fachspezifische Resultate im Rahmen der obligatorischen Schulzeit von besonderem Interesse. Aus Studien von Andrade zum Einsatz von Rubrics beim Texteschaffen gehen positive Ergebnisse hinsichtlich erweiterten Schreibkompetenzen hervor (Andrade et al., 2008; Andrade et al., 2009). Spezifisch für das Experimentieren im Biologieunterricht konnten Wollenschläger et al. (2016) die effektive Wirkung eines Rubrics auf die

lehrseitige Feedbackqualität beim experimentellen Handeln der Lernenden nachweisen. Für den Lernerfolg der Schülerinnen und Schüler ist es offenbar entscheidend, dass Feedback beispielsweise mithilfe eines Rubrics über nächste Lernschritte informiert (Wollenschläger et al., 2016). Auch im Feedback-Modell von Hattie und Timperley (2007) fokussiert solches Feedback die dritte und bedeutsame Frage: „Wohin als nächstes?“ (Hattie et al., 2021). Diese Art von Feedback kann dazu beitragen, die nächstgeeigneten Herausforderungen auszuwählen, den Lernprozess besser zu regulieren, verschiedene Strategien und Verfahren zur Bearbeitung der Aufgaben zu entwickeln, das Verständnis zu vertiefen und selbst den eigenen Lernstand zu beurteilen (Hattie & Gan, 2011).

Rubrics unterstützen Unterrichtende und Lernende, ein gemeinsames Verständnis hinsichtlich der Beurteilungskriterien und Bewertungsmaßstäben zu generieren (Kocakulah, 2010). Es ist daher ein Ziel dieses Beitrags zu klären, wie in den unteren Schulstufen die Folgen des Rubric-Gebrauchs hinsichtlich der Beurteilung und Steuerung des Lernens eingeschätzt werden.

Die vorliegenden Ergebnisse zur schülerseitigen Leistungssteigerung durch die Nutzung von Beurteilungsrastern zeichnen ein unklares Bild. Es liegt u. a. an der Vermischung von Messungen zu Rubrics mit solchen zur Feedbackqualität, Unterrichtsform oder Selbstbeurteilung, sodass für die festgestellten Effekte keine Variablen genügend eindeutig verantwortlich gemacht werden können. Deutlichere Ergebnisse zeigen sich eher bezüglich Aspekten des Lernprozesses und überfachlicher Kompetenzen: Lernende schätzten ein, dass Rubrics ihnen helfen, sich gegenseitig (Peer-)Feedback zu geben und Selbstbeurteilungen vorzunehmen (Kocakulah, 2010; Panadero et al., 2012; Saddler & Andrade, 2004). Ross et al. (2002) führten beispielsweise in 5. und 6. Klassen eine 12-wöchige Intervention zur Förderung der Selbstbeurteilung in Mathematik mithilfe eines Rubrics durch. Die Schülerinnen und Schüler der Interventionsgruppe lernten vier Strategien für die Selbstbeurteilung entlang eines Rubrics kennen: (1) gemeinsam Kriterien definieren, (2) zeigen, wie die Kriterien genutzt werden können, (3) Selbstbeurteilung mit Hilfe der Kriterien durchführen und (4) Handlungsschritte auf Grund der Selbstbeurteilung formulieren. Die Wirkung in der Interventionsgruppe äusserte sich durch höhere Leistungen bei mathematischen Problemlöseaufgaben. In der eigenen Studie LERU führte die Intervention mit Rubrics zu positiven indirekten Effekten bei der Selbstwirksamkeit und der Selbstregulation, nicht aber bezüglich Leistung beim mathematischen Argumentieren (Smit et al., 2017).

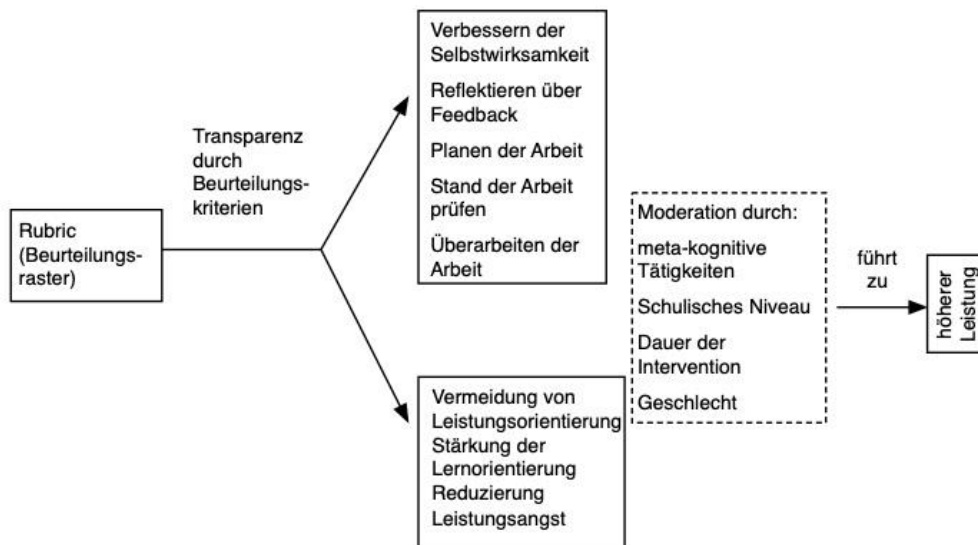


Abb. 1: Rubric und moderierte Wirkung auf Leistungen (Panadero & Jonsson, 2013)

Panadero und Jonsson (2013) (Abb. 1) entwickelten ein theoretisches Modell mit moderierenden Variablen, welches aus einer Literaturanalyse zur Wirkung von Rubrics entstanden ist und bereits bei Ross et al. (2002) angedeutet wurde. Daraus ergibt sich, dass Rubrics Lernenden einen verbindlichen Orientierungsrahmen bieten, durch welchen sie Misserfolgsangst in Mathematik reduzieren und ihre Selbstwirksamkeit stärken können. Insbesondere die Transparenz der Anforderungen vermag die Leistungsangst und verschiedene Ausprägungen von Vermeidungsverhalten (z.B. ausweichen oder blockieren) zu mindern (Panadero et al., 2013). Es wird in der Literaturanalyse deutlich, dass ein Rubric nicht nur in kognitiven Belangen wirkt, sondern auch motivational und emotional (Andrade et al., 2009). Die Begründung dafür ist in der vermittelten Sicherheit zu suchen, welche das Instrument bezüglich Standortbestimmungen, nächster Ziele und Vorgehensweisen auslöst.

Brookhart und Chen (2015) weisen hinsichtlich Effekte moderierender Variablen auf zwei Studien hin, die sich mit der Wirkung von Rubrics auf die Selbstregulation befassen. Die erste zeigt diesbezüglich positive Ergebnisse bei Sekundar-Schülerinnen und Schüler im Geografie-Unterricht (Panadero et al., 2014). In der zweiten Studie zeigten Lehramtsstudierende beim Lernen von Informationstechnologien in einer Interventionsgruppe weniger leistungsvermeidende Aspekte von Selbstregulation, z.B. Konzentrationsschwierigkeiten, bewusste Abweichung von der Zielperspektive, Ängste, Blockaden (Panadero et al., 2013).

2.3 Mathematisches Argumentieren

Gemäss Jahnke und Ufer (2015) ist es ein Ziel des allgemeinbildenden Unterrichts im Fach Mathematik, dass Lernende mathematische Argumentationskompetenzen entwickeln. Viholainen (2011) sieht das Argumentieren auf der Primarstufe als einen Prozess, bei dem Argumente ausgetauscht werden, um zur besten Schlussfolgerung zu gelangen. Diesen diskursiven Charakter des Argumentierens erläutern wir im nächsten Kapitel genauer. In der relativ breiten Definition von Kilpatrick et al. (2001, p. 129) ist Argumentieren der Leim, der alles zusammenhält beim Mathematik treiben, nämlich Fakten, Prozeduren, Konzepte, und Lösungsmethoden. Es geht darum, dass alle diese Komponenten in ihrem Zusammenspiel Sinn machen.

Das Konzept des mathematischen Argumentierens ist jedoch vage, und es gibt keine gemeinsame Definition innerhalb der Forschungsgemeinschaft (Jeannotte & Kieran, 2017). Einigkeit besteht jedoch darüber, dass im allgemeinbildenden Mathematikunterricht kein streng deduktives Beweisen stattfinden muss, sondern, dass laut Jahnke und Ufer „Aussagen auf Gründe zurückgeführt werden sollen“ (S. 333). In diesem Sinne wird im Folgenden auf den Begriff «Beweisen» nicht weiter eingegangen, da dieser für die Primarstufe keine Bedeutung aufweist. In der Literatur finden sich sowohl die Begriffe „Argumentation« wie auch «Begründung bzw. Reasoning“, ohne dass zwischen den beiden Tätigkeiten unterschieden wird (Whitenack & Yackel, 2002), da sie keine klar definierten Grenzen haben (Hanna, 2014). Im deutschsprachigen Raum gibt es auch Forschende, die Argumentieren und Begründen als zwei unterschiedliche Tätigkeiten betrachten und jeweils den einen oder andern Begriff als übergeordnet sehen (Bersch, 2023; Brunner, 2014). Während im angelsächsischen Raum der Begriff «mathematical reasoning» häufiger ist, findet sich im deutschsprachigen Raum zumeist der Begriff «mathematisches Argumentieren». Es lässt sich darauf hinweisen, dass sich Argumentieren und Begründen als individuelle Aktivitäten beide auf denselben Prozess in der Mathematik beziehen (Conner et al., 2014). In der vorliegenden Studie wurde analog des HarmoS-Kompetenzmodells der schweizerischen Bildungsstandards für Mathematik (Linneweber-Lammerskitten et al., 2010) nicht zwischen Argumentieren und Begründen unterschieden.

Jeannotte und Kieran (2017) haben ein Modell des mathematischen Argumentierens für den Einsatz in Schulen vorgeschlagen, das aus zwei Hauptaspekten besteht: dem strukturellen und dem prozeduralen. Der prozedurale Aspekt unterscheidet zwischen verschiedenen, aber miteinander verbundenen Kategorien von Denkprozessen, z.B. Verallgemeinern, Begründen und Exemplifizieren (Stylianides, 2008). Diese Unterscheidung spielt aber in der HarmoS-Kompetenzbeschreibung keine Rolle. Im Folgenden fokussieren wir den bspw. in Lithner (2000) und Toulmin (2003) beschriebenen strukturellen Aspekt, der sich auf bestimmte Schritte oder Bausteine bezieht, die den Prozess des Argumentierens leiten. In diesem Prozess werden Argumente und damit Gründe für oder gegen Aussagen, Hypothesen, Zusammenhänge oder Meinungen gesucht, konstruiert, verwendet, zusammengestellt, verknüpft, rezipiert oder beurteilt (Bersch, 2023). Lithner (2000) beschreibt das Argumentieren als einen vierstufigen Prozess, der aus 1) einer problematischen Situation, 2) einer strategischen Entscheidung, 3) ihrer Anwendung und 4) einer Schlussfolgerung besteht. Einen ähnlichen Prozess lässt sich auch im „Argumentationsmodell für Forscheraufgaben“ von Bezold (2009, S. 131) erkennen, welches sie in Anlehnung an Toulmin (2003) in Bausteine als Voraussetzung für das mathematische Argumentieren und Bausteine für das eigentliche Argumentieren unterteilt. Voraussetzende Bausteine beinhalten das „Entwickeln von Lösungsideen“ sowie die „Planung von Lösungswegen oder komplexer Handlungsabläufe“ (Bezold, 2012). Zum eigentlichen Argumentieren gehört gemäss Bezold (2009) etwa das Erklären von Vorgehensweisen, das Behaupten und Prüfen sowie das Vorhersagen und Verallgemeinern. Wenn ein Kind diese Bausteine zusammenhängt, führt es im Sinne von Toulmin eine Argumentation durch, mit der es erklärt, wie Bausteine logisch miteinander verbunden sind und die dann zu einer abschliessenden Begründung oder Schlussfolgerung ähnlich wie bei Lithner führen.

Mathematisches Argumentieren und Begründen gilt als bedeutsam für einen Mathematikunterricht, der das Verstehen ins Zentrum rückt bzw. Lernanlässe zur individuellen und sozialen Annäherung an mathematische Konzepte anbietet (Thompson & Schultz-Ferrel, 2008). In der Grundschule und Sekundarstufe 1 erfolgt das mathematische Argumentieren beim Erforschen von Mustern und Strukturen oder beim Beschreiben von Beziehungen oder Zusammenhängen. Ende der 6. Primarschulklasse sollten Schülerinnen und Schüler gemäss den Schweizer Bildungsstandards einfache Aussagen durch Nachprüfen an einem konkreten

Beispiel, durch Nutzen vorhandener Daten oder durch naheliegende Argumente begründen oder falsifizieren können (Linneweber-Lammerskitten et al., 2010; Schweizerische Konferenz der Erziehungsdirektoren (EDK), 2011). Dazu gehören etwa die folgenden von uns verwendeten Aufgaben:

1. Bilde aus den Ziffern 4, 5, 6, 7, 8, 9 zwei dreistellige Zahlen und addiere sie. Verwende jede Ziffer nur einmal.
Beispiel: $645 + 798 = 1443$. Mit welchen Zahlen entsteht die grösste Summe?
2. Babys sind etwa 50 cm gross, wenn sie auf die Welt kommen. Bis zum Alter von 12 Jahren wachsen Kinder im Durchschnitt etwa jedes Jahr cm
Kreuze die Antwort an, die am ehesten stimmt und begründe.
A 2 cm
B 4 cm
C 9 cm
D 14 cm

In Anlehnung an Bezold (2010) sehen wir für die Primarstufe das Beschreiben und Begründen als eigentliches Argumentieren an. Nach Ufer und Kramer (2012, p. 92) sind für das Argumentieren auch inhaltsbezogene Kompetenzen notwendig, wie etwa reichhaltiges und vernetztes Begriffswissen und die sichere und flexible Beherrschung mathematischer Tätigkeiten wie das Umformen von Termen oder die Erstellung einer Wertetabelle“. Solche Vorstufen des Argumentationsprozesses wie etwa das Verstehen des Sachverhaltes, das Anwenden von Strategien und das Durchführen von Operationen stellen notwendige Schritte (Bausteine) dar, um Schlussfolgerungen ziehen und damit den Gedankengang des Argumentierens abschliessen zu können (Lithner, 2000). Hier finden sich enge Bezüge zum Problemlösen (Herbert & Williams, 2023). Bezold (2012, p. 75) schreibt dann auch: „In dem Fall, dass zunächst Lösungen für eine Argumentationsgrundlage entwickelt werden müssen, dienen diese Aktivitäten des Problemlösens als unabdingbare Voraussetzungen für das Argumentieren.“

2.4 Rubric zum mathematischen Argumentieren

Der implementierte Rubric (Abb. 2) orientiert sich an den Schweizer Bildungsstandards in Mathematik (Schweizerische Konferenz der Erziehungsdirektoren (EDK), 2011, p. 40), welche u. a. enthalten, dass Schülerinnen und Schüler bis am Ende der Volksschulzeit (Ende 6. Klasse) „ihren Lösungsweg mithilfe einer Rechnung und Erläuterungen darlegen“ oder „ihren Lösungsweg zu funktionalen Zusammenhängen mithilfe von Wertetabellen, Berechnungen und Erklärungen rechtfertigen“ können. Theoretisch wurde auf das „Argumentationsmodell für Forscheraufgaben“ von Bezold (2009) Bezug genommen.

Entsprechend der obigen Erläuterung und der Definition von Kilpatrick et al. (2001) finden sich in unserem Rubric (Abb. 2) nebst dem eigentlich Beschreiben und Begründen auch die Dimensionen Operationen und Strategien. Argumentationen lassen sich ferner auch mit nicht sprachlichen Mitteln vollziehen (Bezold, 2010). Dazu gehören etwa geometrische Muster (z.B. Punkte- oder Kästchenmuster), freie Zeichnungen oder Tabellen (Boonen et al., 2016). Deshalb haben wir eine vierte Dimension im Rubric hinzugefügt für Darstellungen oder Beispiele. Inhaltlich bezieht unser Rubric auf Aufgaben zu drei der vier Inhaltsbereiche der Kompetenzen für den Zyklus 5-8: Zahl und Variable, funktionale Beziehungen sowie Grössen und Masse (Schweizerische Konferenz der Erziehungsdirektoren (EDK), 2011).

Bereich	Reflexionsfragen	Niveau 1	Niveau 2	Niveau 3	Niveau 4
Passendes und verständliches Vorgehen	Wie bist du vorgegangen? Findest du noch andere Wege?	Das Vorgehen fehlt oder ist nicht richtig.	Das Vorgehen zeigt Ansätze, welche zu einer richtigen Lösung führen könnten. Das Vorgehen ist nicht vollständig dargestellt.	Das Vorgehen kann zu einer richtigen Lösung führen. Das Vorgehen ist fast vollständig aufgeschrieben. Es sind noch weitere, aber nicht sinnvolle Vorgehensweisen notiert.	Das Vorgehen ist richtig und vollständig aufgeschrieben. Mögliche eigene sinnvolle zusätzliche Überlegungen oder Fragen sind notiert. Es sind noch weitere korrekte Vorgehensweisen aufgeführt.
Korrektes Rechnen	Kann dein Ergebnis stimmen?	Viele Rechenfehler, welche zu einem unbrauchbaren Ergebnis führen.	Wenige, aber bedeutende Rechenfehler.	Wenige, aber eher unwichtige Rechenfehler.	Richtige Rechnungen
Verständliche und ausführliche Begründung (Beschreibung)	Welche Besonderheiten hast du entdeckt? Welche Lösungen hast du gefunden? Warum ist das so? Was ist (nicht) gemeinsam?	Begründung (Beschreibung) fehlt, ist nicht verständlich oder sie hat nichts mit der Aufgabe zu tun.	Eine sinnvolle, verständliche Begründung (Beschreibung) steht teilweise da. Die Begründung passt teilweise zum Vorgehen und zur Aufgabe.	Eine sinnvolle, verständliche Begründung (Beschreibung) steht fast vollständig da. Die Begründung passt recht gut zur Aufgabe und zum Vorgehen.	Eine mathematisch gute Begründung (Beschreibung) steht da. Die Begründung passt gut zur Aufgabe und zum Vorgehen.
Bilder, Beispiele	Welche Bilder oder Beispiele könnten helfen, deine Begründung zu verstehen?	Hilfreiche und mögliche Bilder fehlen.	Bilder oder Beispiele stehen, sind aber nicht verständlich.	Bilder oder Beispiele ergänzen die Begründung.	Bilder oder Beispiele helfen die Begründung zu verstehen.

Abb. 2: Rubric zum mathematischen „Argumentieren und Begründen“

Lernanlässe zum mathematischen Argumentieren setzen gute Aufgaben voraus, die je nach Bedingungen und Möglichkeiten der Lernenden unterschiedlich komplexe und abstrakte Lösungswege zulassen sowie den Austausch unter Lernenden einfordern (Stein et al., 1996). Dies erfolgt vorzugsweise in kooperativen Lernanlässen, die der Lehrperson Gelegenheit bieten, mehr über die Denkprozesse der Kinder zu erfahren und darauf Einfluss zu nehmen (Ginsburg, 2009; Sfard, 2001). Die unterrichtliche Umsetzung des mathematischen Argumentierens erfordert von Lehrpersonen anspruchsvolle fachdidaktische Kompetenzen wie das Leiten von Diskursen und das Fördern von Ideen der Lernenden (Blanton & Kaput, 2005). Lehrpersonen berichten zudem über mangelnde zeitliche Ressourcen, um längere mathematische Sachkontexte mit Gruppen von Schülerinnen und Schüler zu erforschen und die Ergebnisse dann schriftlich festzuhalten und zu diskutieren (Keiser & Lambdin, 1996). Mathematisches Argumentieren ist begleitet von Steckenbleiben, Wege ändern und Bedarf somit der Fähigkeit Ausdauer zu zeigen (Barnes, 2019). Emotionen wie Frustration oder Verwirrung sind häufige Begleiterscheinungen. Kinder mit einem fremdsprachigen Hintergrund bekunden zudem größere Schwierigkeiten mit problemhaltigen Aufgaben, wie sie z.B. beim mathematischen Argumentieren vorliegen, als diejenigen Kinder mit Deutscher Muttersprache (Kempert et al.,

2011; Stanat & Christensen, 2006). Als Unterstützung in selbstregulierten Arbeitsphasen können den Lernprozess stützende Werkzeuge eingesetzt werden. Dazu können ein formativ genutztes Feedback und Rubrics gehören (Herbert et al., 2022; Turner, 2014).

2.5 Rubric als Werkzeug der Selbstregulation beim mathematischen Argumentieren

Laut Brunner (2014) gehören auch metakognitive, motivationale und emotionale Aspekte zur Argumentationskompetenz. Dies wollen wir im Folgenden erläutern. Die in unserer Studie fokussierte Kompetenz des mathematischen Argumentierens hat einen ausgeprägt diskursiven Charakter. Dabei kann es sich um Lehrer-Schüler Diskurse in der Klasse oder um Schüler-Schüler Diskurse in der Kleingruppe handeln (Adie et al., 2018). In Übungsphasen kommt es auch zu Eins-zu-Eins Situationen zwischen Lehrperson und Schüler/-in. Mathematisches Wissen entwickelt sich beim gemeinsamen Bestätigen oder Falsifizieren von Behauptungen im Klassenzimmer (Ball & Bass, 2000). Im Diskurs-Modell von Shilo und Kramarski (2019) brauchen Lernende beim Bearbeiten von Argumentationsaufgaben kognitives Wissen und metakognitive Fähigkeiten. Kognitives Wissen, so die Autoren, besteht aus deklarativem (Begriffe, Konzepte), prozeduralem (Operationen, Strategien) und Erklärungs-Wissen (Generalisieren, Verifizieren, Vergleichen, ...). Die Kompetenz argumentieren und begründen zu können, so Kilpatrick et al. (2001), zeigt sich darin logische Schlussfolgerungen ziehen zu können (z.B. durch Vergleichen) auf der Grundlage von Beziehungen zwischen Konzepten, Begriffen und Situationen, in denen Strategien verwendet werden. Shilo und Kramarski (2019) ergänzen nun in ihrem Diskurs-Modell das kognitive Wissen mit metakognitiven Fähigkeiten: planen, überwachen und reflektieren. Gebraucht werden solche metakognitiven Fähigkeiten in Zusammenhang mit der Fähigkeit zur Selbstregulation (Boekaerts, 1999). Selbstgesteuertes Lernen ist ein zyklischer Prozess, bei dem der Schüler eine Aufgabe plant, seine Leistung überwacht und dann über das Ergebnis nachdenkt (Zimmerman, 2002). Diskurse, in welchen von Schülerinnen und Schüler erwartet wird Erklärungen und Begründungen für ihre Antworten zu geben, fördern die Entwicklung von Selbstregulation (Pape & Smith, 2002). Damit ein Diskurs initiiert wird, braucht es Aufgaben, welche den Lernenden Freiräume ermöglichen verschiedene Vorgehensweisen und unterschiedliche Darstellungen einzusetzen. In der Folge sollte sich ein diskursiver Austausch ergeben, in welchem Schülerinnen und Schüler ihre Lösungsansätze oder Antworten erklären und begründen (Stein et al., 1996). Der Diskurs funktioniert aber nur, wenn das kognitive Wissen durch metakognitive Fähigkeiten (etwa Reflexion) nutzbar gemacht werden kann (Pape & Smith, 2002; Shilo & Kramarski, 2019). Barnes (2019) weist darauf hin, dass während des Lösens von Argumentationsaufgaben Selbstregulation der Lernenden nötig ist, um fokussiert zu bleiben auf das eigentliche Argumentieren. Grund dafür könnte die fehlende emotionale Selbststeuerung sein. Die Freude an gefundenen Lösungen kann zum Bestimmen weiterer Lösungen führen anstatt über das Begründen derselben nachzudenken (Barnes, 2019). Hier wäre es dann auch die Aufgabe der Lehrperson, um notfalls durch Feedback im Bereich Selbststeuerung (Hattie & Gan, 2011) auf das eigentliche Ziel der Argumentation zurückzuführen.

Auch im Prozess-Modell des selbstregulierten Lernens von Butler und Winne (1995) spielen metakognitive Fähigkeiten eine bedeutende Rolle. Sie helfen die Zielerreichung, das Lösen einer Aufgabe, selbst zu überwachen. Das Modell von Butler und Winne (1995) gibt darüber hinaus eine externe Unterstützung der Selbstregulation vor. Eine solche kann beispielsweise durch formatives Feedback der Lehrperson an Lernende, beispielsweise auf der Grundlage eines Rubrics erfolgen (Herbert et al., 2022; Turner, 2014). Der Rubric aus dem

Projekt LERU (Abb. 2) unterstützt die Lernenden insbesondere beim Steuern und Reflektieren von Prozessschritten respektive hinsichtlich des Erwerbs prozeduralen Wissens beim mathematischen Argumentieren. Dazu enthält der Raster verschiedene Hinweise in Form von Reflexionsfragen (Forscherfragen bei Bezold (2009)).

Im Projekt „Lernen mit Rubrics“ (LERU) wurde untersucht, welche Effekte eine Weiterbildung für Lehrpersonen bezüglich der Nutzung von Rubrics im Mathematikunterricht zeigt. Inhaltlich stand die eher komplexe Kompetenz des mathematischen Argumentierens und Begründens auf der Primarstufe im Fokus, weil entsprechende Aufgaben laut Prenzel et al. (2013) noch eher wenig eingesetzt werden. In den Lehrmitteln finden sich zudem nur wenig Aufgaben zum Argumentieren (Brunner et al., 2019). Es ergab sich in der Interventionsgruppe ein positiver Effekt von Rubrics auf die Selbstregulation bei Argumentationsaufgaben von Schülerinnen und Schülern der 5. und 6. Primarschulklasse, wobei die kognitiven und motivationalen Aspekte nicht gesondert untersucht wurden (Smit et al., 2017). Die Wirkung lässt sich mit häufigeren Peer- und Selbstbeurteilungen sowie mit vermehrtem formativem Feedback der Lehrperson an Lernende erklären. Ein (indirekter) Effekt des Rubrics auf die Leistung liess sich jedoch nicht nachweisen und wird im Folgenden deshalb auch nicht geprüft.

2.6 Forschungsfragen

Die hier zu klärenden Forschungsfragen beziehen sich auf das theoretische Modell von Panadero und Jonsson (2013) (Abb. 1). Im Mathematik-Unterricht übernehmen Rubrics insofern eine zentrale Rolle, als sie den Lehrenden helfen, die Beurteilungskriterien (und Ziele) zu Beginn einer Lektionsreihe und während des Arbeitens transparent zu machen. Damit haben die Lehrpersonen auch eine Grundlage, auf der sie während des Unterrichts gezielt Rückmeldungen erteilen können. Wird der Rubric von den Schülerinnen und Schülern als nützlich für die Selbststeuerung empfunden, sind theoretisch positive Effekte für die Lernenden, z.B. erhöhte Selbstwirksamkeit, weniger Unsicherheit, bessere Selbstregulation und letztlich auch bessere Leistungen zu erwarten. Die implizierte Wirkung lässt sich mit der gesicherten Umsetzung metakognitiver Strategien und der Dauer der Instrumenten-Nutzung steigern. Die Nützlichkeits einschätzungen eines Rubrics lassen sich kognitiven und motivationalen Aspekten zuordnen. Zu den kognitiven Aspekten gehören etwa „Planen der Arbeit“, „Stand der Arbeit prüfen“ oder „Reflektieren über erhaltene Feedbacks“ seitens der Lehrperson. Den motivationalen Aspekten lassen sich „Reduzierung der Leistungsangst“, „Vermeidung von Leistungs-, Stärkung der Lernorientierung“ oder „Selbstwirksamkeit“ zuordnen. Solche Aspekte spielen insbesondere bei den eher offenen Argumentationsaufgaben – mit einem stärkeren Explorations- und Unsicherheitscharakter – eine bedeutende Rolle.

Die bisherige Darstellung und das Modell von Panadero & Jonsson (2013) führt zu folgenden Forschungsfragen:

1. Hängen die Einschätzungen der Lernenden bezüglich der Nützlichkeits des Rubrics für das Lernen (kognitive und motivationale Aspekte), mit der eigenen Fähigkeit zur Selbstregulation und der wahrgenommenen Häufigkeit des Lehrpersonenfeedbacks bezüglich der Selbstregulation zusammen?
2. Lassen sich Entwicklungen während des Projekts in der selbst eingeschätzten Fähigkeit zur Selbstregulation auf die Einschätzung der Nützlichkeits zurückführen?

Folgende Hypothesen zur Nützlichkeit von Rubrics im Unterricht werden formuliert und geprüft:

H1: Es gibt am Ende der Intervention einen positiven Zusammenhang zwischen den Einschätzungen von Schülerinnen und Schülern zur Nützlichkeit des Rubrics, ihrer Einschätzung der eigenen Fähigkeiten zur Selbstregulation beim Lösen von Argumentationsaufgaben und der Wahrnehmung der Häufigkeit des Lehrerfeedbacks zur Selbstregulation.

H2: Die Schülereinschätzungen zur kognitiven und motivationalen Nützlichkeit haben je eine individuelle, positive Wirkung auf die Entwicklung der selbst eingeschätzten Selbstregulation während der Interventionsdauer t1-t2.

H2a: Die Schülerinnen und Schüler unterscheiden sich individuell in ihrer Entwicklung der Selbstregulation.

H2b: Die kognitive und motivationale Nützlichkeit steht je in einem Zusammenhang mit der Entwicklung der Selbstregulation.

3 Methoden

3.1 Forschungsdesign und Stichprobe

Die dem Beitrag zu Grunde liegende Studie war Teil des Projektes „LERU – Lernen mit Rubrics“. Das zweijährige Projekt mit Beginn 2015 wurde von den pädagogischen Hochschulen (PH) St. Gallen und Zug durchgeführt. Das quasi-experimentelle, längsschnittliche Forschungsdesign bestand aus zwei prä/post Erhebungszeitpunkten. Als Erhebungsinstrumente wurden Fragebogen (für Lehrpersonen und Schüler/-innen) sowie Leistungstests für Schüler/-innen eingesetzt. Alle Teilnehmenden haben sich freiwillig auf Grund einer Ausschreibung oder persönlicher Kontakte gemeldet. Die Gelegenheitsstichprobe umfasste 44 Lehrpersonen der 5. und 6. Klassenstufe sowie deren 762 Schülerinnen und Schüler. In die folgenden Analysen zur Nützlichkeit des Rubrics werden ausschliesslich Daten der Interventionsgruppe einbezogen, da nur diese mit dem Rubric arbeitete: 22 Klassen, 338 Schülerinnen und Schüler.

Die Implementierung des Treatments zum mathematischen Argumentieren – als offizieller Weiterbildungsinhalt – erfolgte in einem eintägigen Workshop. Die Interventionsgruppe wurde zusätzlich in den von uns entwickelten Rubric eingeführt, während sich die Kontrollgruppe mit einem anderen mathematikdidaktischen Inhalt befasste. In der folgenden 9-wöchigen Umsetzungsphase arbeiteten die Interventions- und die Kontrollgruppe jeweils eine Lektion pro Woche an Aufgaben zum mathematischen Argumentieren. Dies entlang eines vorgegebenen Unterrichtsskripts, dessen Aufbau dem methodischen Vorgehen von Andrade et al. (2008) entspricht. Unser abgeleitetes didaktisches Konzept für die Interventionsgruppe bestand aus sechs Schritten, wobei jeder Schritt etwa eine Unterrichtsstunde in Anspruch nahm: 1) Die Schüler begannen mit der Bearbeitung einer typischen Aufgabe. Nach einer gewissen Zeit wurden einige vorbereitete Schülerbeispiele von möglichen Argumentationen angeboten, um die Grundlagen des mathematischen Argumentierens einzuführen; 2) die Schüler diskutierten die Stärken und Schwächen der Beispiele und erstellten eine Liste der Eigenschaften einer guten Begründung, wobei sie von der Lehrkraft unterstützt wurden; 3) die Schüler erhielten den Rubric; 4) die Schüler benutzten den Rubric, um ihre Lösungen laufend einzuschätzen; 5) die Schüler trainierten ihre Kompetenz, indem sie an Aufgaben zum Argumentieren arbeiteten, wobei sie öfters auch ein Lehrerfeedback auf der Grundlage des Rubrics erhielten; und 6) die Schüler lernten, die Argumentationen selbst und in Gruppen zu bewerten. Dasjenige Skript der Kontrollgruppe war ähnlich, einfach ohne Rubric. Die Unterrichtspläne waren

sozio-konstruktivistisch ausgerichtet: Kooperative Gruppenarbeiten oder Lernen in Peers (z. B. Placemat oder Lerntempduett) zur Förderung des Schülerdiskurses war Teil fast jeder Unterrichtsstunde (Knudsen et al., 2014). Im ersten Teil des Skripts wurden die Lehrkräfte gebeten, die Qualität verschiedener Argumentationsbeispiele mit der Klasse zu diskutieren, um die Ziele zu verdeutlichen. Weiter nutzten die Schülerinnen und Schüler Peer- und Selbstbewertungen für das Verstehen der Kriterien. In den letzten Wochen sollten die Lehrpersonen jedem Kind Rückmeldungen über den Grad der Kompetenz im mathematischen Argumentieren geben. Die Lehrkräfte erhielten eine Reihe von Aufgaben zum mathematischen Argumentieren, einschließlich möglicher Lösungen. Diese Aufgaben umfassten inhaltlich Zahlenreihen, Mengen und Einheiten, das Dezimalsystem, Grundrechenarten, Proportionen und Schätzungen. Einige dieser Aufgaben waren rein mathematisch, während andere einen Bezug zur Realität hatten und kontextbezogen waren (inner- und aussermathematische Aufgaben). Am Ende der Umsetzungsphase wurde die Weiterbildung mit den Lehrpersonen beider Gruppen evaluiert und die Teilnehmenden mit und ohne Einführung in die Rubrics auf einen vergleichbaren Wissensstand gebracht.

3.2 Instrumente und Items

Die Schülerinnen und Schüler erhielten einen Fragebogen mit Items zu Überzeugungen und Unterrichtswahrnehmungen. Die Formulierung der selbst entwickelten Items zur Nützlichkeit basieren auf dem theoretischen Modell von Panadero und Jonsson (2013, p. 139) sowie auf der Studie von Andrade et al. (2009). Ferner flossen auch Ergebnisse eines Interviews mit Schülerinnen und Schüler aus einer Pilot-Studie ein (Smit & Birri, 2014).

Die Aspekte zur Nützlichkeit bezogen sich auf den Selbstregulationsprozess. Drei Items fokussierten kognitive Aspekte: „Ziele klären“, „Lernstände prüfen“ und „nächste Schritte bestimmen“. Drei weitere bezogen sich auf motivationale Aspekte: „motiviert zur Weiterarbeit“, „gibt Sicherheit“ und „zeigt Probleme auf“. Probleme aufzeigen lässt sich sowohl kognitiv wie auch affektiv, volitional deuten. Aus der eindimensionalen Skala mit sechs Items ergab sich ein Cronbach alpha von .88.

Die Schülerinnen und Schüler-Items zur Selbstregulation bei der Bearbeitung mathematischer Aufgaben wurden ausgehend von Zimmerman (2002) und der Studie von Purdie et al. (1996) adaptiert und teilweise neu konstruiert (vgl. Anhang). Es wurde nach „Textaufgaben“ im Sinne von „word problems“ gefragt, weil die Schülerinnen und Schüler zum Zeitpunkt t1 noch kaum vertraut waren mit der Bezeichnung „Argumentationsaufgaben“. Zum Zeitpunkt t2 war ihnen aber der Typus Argumentationsaufgaben mit einer verlangten Begründung vertraut. Es ist anzunehmen, dass die Schülerinnen und Schüler einen Bezug zu den bearbeiteten Aufgaben herstellten. Zwei der sieben Items mussten bei der konfirmatorischen Faktorenanalyse auf Grund ungenügender Faktorladung ausgeschlossen werden.

Schlussendlich wurde noch eine selbst konstruierte Skala für die Wahrnehmung des Feedbacks zur Selbstregulation in Anlehnung an Hattie und Timperley (2007) in die Befragung einbezogen. Lernende sollen angehalten werden, ihren Lernprozess zu beobachten, beispielsweise durch Zielüberwachung, Klären von Lösungsstrategien oder Einschätzen der Lernprodukte.

In der Unterrichtsforschung hat es sich etabliert, auch spezifische Kontextvariablen zu erfassen (Helmke, 2015). Als Kontextvariablen für selbstregulative Strategien wurden „Geschlecht“, „Alter“, „Schulstufe“ und „über Bücher in der Familie sprechen“ (Indikator für den sozioökonomischen Status - SoS) (Zimmerman & Pons, 1986) einbezogen.

3.3 Analyse

Alle Strukturgleichungsmodelle (SEM) wurden mit Mplus 8 und einer „full information maximum likelihood“ Schätzung (FIML) gerechnet, d.h. fehlende Werte wurden in die Berechnung der Schätzwerte miteinbezogen. Die Evaluation der Modellanpassung erfolgte mit komparativen und absoluten Fit-Indices. Werte des komparativen Fit-Index CFI > .95 zeigen eine sehr gute Passung (Hu & Bentler, 1999). Dasselbe gilt für diejenigen des globalen Fit-Index Root Mean Square Error of Approximation (RMSEA) < .05. Bei den Analysen mit Daten der Schülerinnen und Schüler wurde die Clusterstruktur (Designeffekt, Adjustierung der Standardfehler) in Klassen berücksichtigt, ohne dass Mehrebenenanalysen gerechnet wurden (McNeish, Stapleton, & Silverman, 2017). Dies erfolgte in MPlus mit dem Befehl `type = complex`. Modellvergleiche lassen sich mit dem Chi-Quadrat Test durchführen: Unterschiede mit einem $p < .05$, weisen auf einen signifikanten Modellunterschied hin. Im vorliegenden Fall (MLR-Schätzer) wurde zudem die Sartorra-Bentler-Korrektur mitberücksichtigt.

Lineare Wachstumsmodelle erlauben, individuelle Veränderungen über die Zeit zu prüfen. Im Modell gibt es zwei latente Variablen – intercept und slope – welche einen mittleren Wert zum Zeitpunkt t1 und die Wachstumsrate zwischen den Zeitpunkten t1 und t2 repräsentieren. Da es schwierig ist, für kleinere Stichproben ein SEM mit allen manifesten Items zu berechnen, wurden zu den Wachstums-Modellen Item-Päckchen (parcels) über die Selbstregulations-Skala hinweg gebildet (Little, 2013, p. 20 ff). Die Variable Nützlichkeit wurde jeweils als latente Variable modelliert. Für kleine und nicht normalverteilte Stichproben ist zudem die Bayes'sche-Analyse eine gut geeignete Alternative zu ML-Schätzverfahren (Muthén, 2010). Die Überprüfung des Modell-Fits kann in einer flexiblen Weise mittels Posterior Predictive Checking erfolgen. Ein exzellentes Modell sollte einen Posterior Predictive P-(PPP)-Wert von rund 0.5 aufweisen (Muthén & Asparouhov, 2012).

4 Ergebnisse

4.1 Der Nutzen des Rubrics aus Sicht der Schülerinnen und Schüler – Deskriptive Angaben

Schülerinnen und Schüler bejahen tendenziell die Nützlichkeit des Rubrics und dies über alle Items hinweg (Tab. 1). Die Items 1 bis 3, welche auf kognitive Aspekte – „der Rubric hilft verstehen, worauf es ankommt“, „er hilft Lösungen einzuschätzen“, „er hilft Aufgaben zu verbessern“ – ausgerichtet sind, finden am Ende des Projektes eine etwas höhere Zustimmung als die Items 4 bis 6 auf der motivationalen Ebene („der Rubric hilft Probleme aufzuzeigen“, „er motiviert“, „er gibt Sicherheit“). Die bei allen Items relativ hohen Standardabweichungen weisen darauf, dass die Schülerinnen und Schüler den Nutzen unterschiedlich beurteilen. Dies gilt insbesondere für das am schwächsten bewertete Item „der Rubric motiviert“. Alle Items korrelieren hoch miteinander, was auf eine eher generelle Einschätzung des Nutzens und weniger auf eine spezifische Einschätzung von Teilaspekten schließen lässt.

Tabelle 1: Mittelwerte, Standardabweichungen und Inter-Item-Korrelationen der Schülerinnen und Schüler-Items zur Nützlichkeit des Rubrics (t2)

Der Rubric...	<i>M</i>	<i>SD</i>	1	2	3	4	5
1. hilft verstehen, worauf es ankommt	4.12	1.38					
2. hilft Lösungen einzuschätzen	4.37	1.27	.71**				
3. hilft Aufgaben zu verbessern	4.01	1.36	.65**	.58**			
4. hilft Probleme aufzuzeigen	3.82	1.37	.51**	.48**	.58**		
5. motiviert	3.57	1.52	.53**	.42**	.57**	.56**	
6. gibt Sicherheit	3.89	1.46	.62**	.53**	.59**	.57**	.66**

Anmerkung: $N = 338$; ** $p < .01$. Likert Skala 1–6; Pole: stimmt absolut nicht/stimmt genau

4.2 Nützlichkeit des Rubrics und Fähigkeiten zur Selbstregulation

4.2.1 Zusammenhang zwischen den eingeschätzten Nützlichkeitsaspekten, Fähigkeiten zur Selbstregulation und Feedback der Lehrperson

Im Folgenden wird die erste Hypothese geklärt, ob die von den Schülerinnen und Schüler eingeschätzte Nützlichkeit des Rubrics mit der eingeschätzten Fähigkeit zur Selbstregulation bei Argumentationsaufgaben korrespondiert. Damit verbunden ist die Prüfung, ob sich die eingeschätzte kognitive und motivationale Nützlichkeit zum Zeitpunkt t2 in einen Zusammenhang mit den selbst eingeschätzten Fähigkeiten zur Selbstregulation und Feedbackhäufigkeit setzen lassen. Zum gerechneten Strukturgleichungsmodell (SEM; Abb. 3) wurden alle drei Kontextvariablen explorativ geprüft, aber nur die Kontextvariable „über Bücher in der Familie sprechen“ (SoS; vgl. Kap. 3.2) half zusätzlich Varianz aufzuklären. Schülerinnen und Schüler mit einem tieferen sozioökonomischen Status (SoS) schätzten den kognitiven Aspekt des Rubrics etwas nützlicher ein. Die Variable wurde aber im weiteren Vorgehen nicht weiter berücksichtigt, da der Effekt eher gering ist.

Das Modell weist gute Fit-Werte auf ($\chi^2(84) = 133.38$, $p = .00$, CFI = .97, TLI = .96, RMSEA = .04, SRMR = .04). Eine Berücksichtigung der Kovarianzen zweier Residuen verbessert das 2-dimensionale Modell nochmals leicht. Solche Kovarianzen können entstehen, wenn sich die Items ähneln (Byrne, 2012). Das finale Modell hat folgende Fit-Werte ($\chi^2(82) = 109.49$, $p = .02$, CFI = .98, TLI = .98, RMSEA = .03, SRMR = .04).

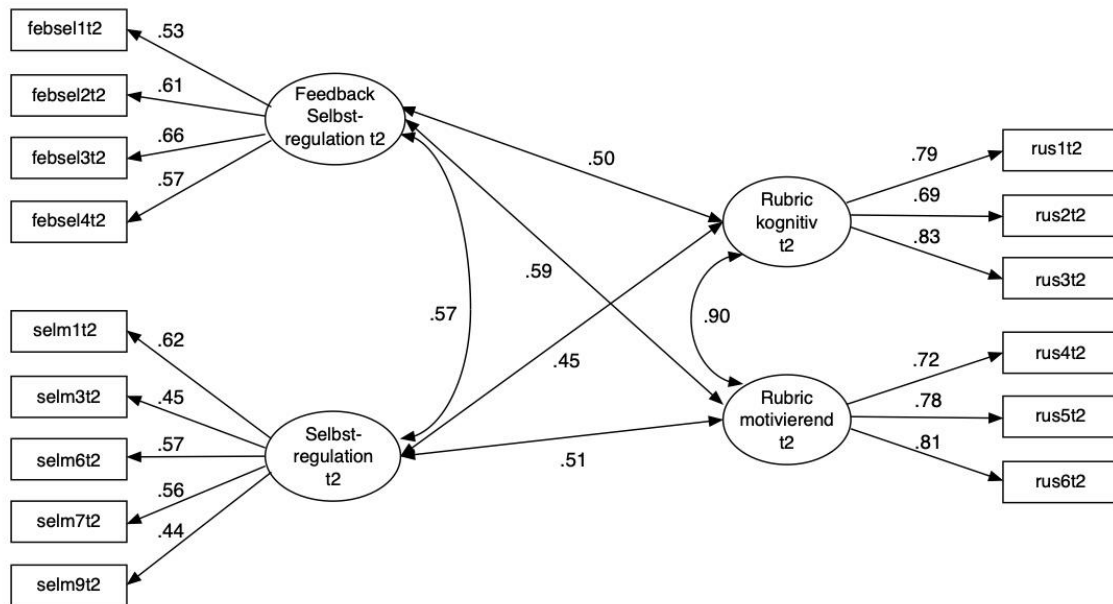


Abb. 3.: Strukturgleichungsmodell für den Zusammenhang zwischen „Wahrnehmung der Fähigkeiten zur Selbstregulation“, „Wahrnehmung des Lehrpersonenfeedbacks zur Selbstregulation“ und „Nutzung des Rubrics“ basierend auf Selbsteinschätzungen der Schülerinnen und Schüler. Alle standardisierten Pfadwerte sind signifikant, $N = 338$, $p < .01$.

Es zeigt sich wie erwartet ein positiver Zusammenhang zwischen den drei Faktoren „Nützlichkeit des Rubrics“, „Fähigkeiten zur Selbstregulation“ bei Argumentationsaufgaben und Wahrnehmung des Feedbacks für die Selbstregulation seitens der Lehrperson. Je nützlicher die Schülerinnen und Schüler den Rubric einschätzten, desto eher schrieben sie sich zu, selbstverantwortlich und zielgerichtet vorzugehen. Die Betakoeffizienten sind für beide Faktoren etwa gleich gross ($\beta = .45/.51$). Desgleichen geht die Nützlichkeit auch einher mit der Wahrnehmung eines häufigeren Feedbacks zur Selbstregulation ($\beta = .50/.59$). Schülerinnen und Schüler, welche sich eine grössere Fähigkeit zur Selbstregulation zuschreiben, schätzen auch die Häufigkeit des erhaltenen Feedbacks höher ein ($\beta = .57$).

4.2.2 Wirksamkeit der Nützlichkeitsaspekte auf die Entwicklung der Fähigkeit zur Selbstregulation

Zur Prüfung der 2. Hypothese wurden drei lineare Wachstumsmodelle mit dem Bayes-Schätzer gerechnet (Grimm et al., 2017, p. 68 ff). Alle drei Modelle wurden auf konvergierende Schätzwerte geprüft und ein PPP-Wert (siehe 3.3), der möglichst um .50 liegen sollte, wurde berechnet. Das erste Basis-Modell klärt die Hypothese 2a, den individuellen Zuwachs selbst eingeschätzter Fähigkeit zur Selbstregulation bei Aufgaben über die gesamte Projektlaufzeit ohne Berücksichtigung von weiteren Prädiktoren. Über die gesamte Interventionszeit konnte eine signifikant zunehmend positivere Einschätzung der Fähigkeit zur Selbstregulation nachgewiesen werden (slope, Tab. 3). Generell zeigt sich, dass die Lernenden ihre Fähigkeiten signifikant unterschiedlich entwickelten (slope Varianz). Zudem ist die Zunahme bei Schülerinnen und Schüler mit geringer eingeschätzter Fähigkeit stärker als bei solchen mit höherer ($s * i$). Dieses erste Wachstumsmodell erklärt jeweils 87% (t1) und 89% (t2) der Varianz der selbst eingeschätzten Fähigkeiten zur Selbstregulation bei Aufgaben.

Ausgehend von diesem Basis-Modell wurden zwei weitere Modelle zur Klärung der Hypothese 2b aufgestellt. Die beiden zusätzlichen Modelle erweitern das Basis-Modell jeweils um zwei Prädiktoren (siehe Abb. 4): Nützlichkeit des Rubrics im kognitiven resp. motivationalen Bereich und „Bücher in Familie“ (SoS). Dabei geht es um die Frage, ob die beiden Prädiktoren zusätzlich interindividuelle Unterschiede bezüglich der eingeschätzten Selbstregulation erklären. Die Effekte der beiden Nützlichkeits-Aspekte zeigen ähnlich hohe, signifikante Werte. Die Hypothese 2b kann damit also bestätigt werden. Schülerinnen und Schüler, die jeweils kognitive oder motivationale Aspekte der Rubric-Nützlichkeit höher einschätzten, steigerten ihre Fähigkeiten zur Selbstregulation stärker ($\beta = .17$ resp. $.21$). Bei der explorativ untersuchten Kontextvariablen „sozioökonomischer Status“ (SoS) sieht es ein wenig anders aus: Schülerinnen und Schüler mit einem hohen SoS starteten bei einer höheren Fähigkeitseinschätzung (intercept). Sie steigerten allerdings ihre Selbstregulation weniger stark als solche mit tieferem SoS ($\beta = -.18$ resp. $-.21$). Wie bereits in Kap. 4.2.1 erwähnt, schätzten Schülerinnen und Schüler mit einem hohen SoS die kognitiven Aspekte des Rubrics als weniger nützlich ein als Lernende mit einem tiefen SoS. Auf die Einschätzung motivationaler Nützlichkeits-Aspekte hat der SoS keine signifikante Wirkung. Die beiden Prädiktoren erklären jeweils 14% resp. 18% der Varianz eingeschätzter Fähigkeit zur Selbstregulation bei Aufgaben zum Zeitpunkt t1 und 7% resp. 9% der Varianz in der Wachstumsrate.

Tabelle 2: Lineares Wachstums-Modell für Selbstregulation t1-t2 (Basis-Modell)

	Schätzwert	Posterior S.D.
<i>Fester Effekt</i>		
Intercept (Ausgangslage)	5.20**	.55
Slope (lineare Wachstumsrate)	.30**	.16
Slope mit Intercept (s * i)	-.42**	.12
<i>Varianz (nicht standardisiert)</i>		
Intercept	.51**	.08
slope	.52**	.16

Anmerkung: Standardisierte Bayessche Schätzwerte, Posterior Predictive P-Wert: .48, ** $p < .01$

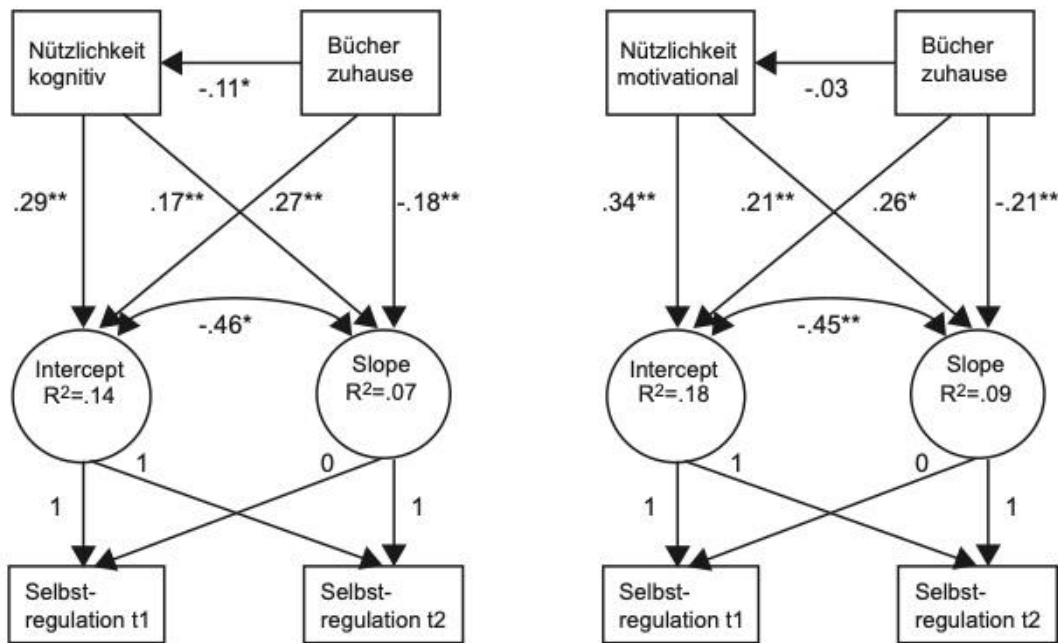


Abb. 4: Lineares Wachstums-Modell für die Selbstregulation zwischen den Zeitpunkten t1 und t2 unter Einschluss der Prädiktoren kognitive und motivationale Nützlichkei[t] sowie Anzahl Bücher zuhause

Anmerkung: Standardisierte Bayessche Schätzwerte, Ergebnisse für die motivationale Nützlichkei[t] in Klammern, Posterior Predictive P-Wert: .56 (.48), ** $p < .01$, * $p < .05$.

5 Diskussion

Zusammenfassung

Mit der Studie wurde beabsichtigt, einen Rubric, ein im deutschsprachigen Gebiet weniger bekanntes Instrument zur formativen Beurteilung (formative assessment), exemplarisch für das mathematische Argumentieren, vorzustellen. Es sollten kognitive und motivationale Aspekte der Nützlichkei[t] eines Rubrics empirisch mittels Einschätzungen der Schülerinnen und Schüler nachgewiesen und Zusammenhänge zur Entwicklung der Selbstregulation beim mathematischen Argumentieren und zur Häufigkeit des Feedbacks geprüft werden. Sämtliche Hypothesen konnten bestätigt werden: 1. Die von Lernenden eingeschätzte Rubric-Nützlichkei[t] beim Argumentieren steht in einem positiven Zusammenhang zur eingeschätzten Fähigkeit zur Selbstregulation und zur Häufigkeit des Lehrpersonenfeedbacks. Dies gilt für beide theoretisch erwartete Aspekte, die kognitive und die motivationale Nützlichkei[t]. 2. Generell besteht eine signifikante Varianz in den Wachstumsraten der selbst eingeschätzten Fähigkeit zur Selbstregulation beim mathematischen Argumentieren. Das heisst, dass die Entwicklung der Selbstregulation interindividuell unterschiedlich verlief. Lernende mit tiefen Selbsteinschätzungen der Selbstregulationsfähigkeit haben sich zudem stärker gesteigert als solche mit hohen Einschätzungen. Die Varianz lässt sich – zumindest teilweise – durch die eingeschätzte Nützlichkei[t] des Rubrics erklären, da sich sowohl die kognitiven als auch die motivationalen Aspekte signifikant auf die Entwicklung der Selbstregulations-Einschätzungen beim Argumentieren auswirken. Bei Lernenden, welche die Nützlichkei[t] höher bewerteten, nahm die Selbstregulation bis Ende der Intervention stärker zu als bei solchen mit tieferen Einschätzungen. Es

ist zudem eine Tendenz auszumachen, dass Lernende mit einem tieferen Bildungshintergrund stärker von Rubric profitierten, da diese die Nützlichkeit höher einschätzten und damit indirekt auch die eingeschätzte Fähigkeit der Selbstregulation zunahm. Dieses Ergebnis passt zu einer Studie von Holmes und Hwang (2016), in der gezeigt werden konnte, dass die Förderung kognitiver Strategien bei komplexeren, authentischen Mathematikaufgaben in kooperativen Settings zu einer Abnahme des SoS-Effekts auf die Mathematikleistungen führt.

Folgerungen und Empfehlungen

Bei den motivationalen Aspekten sind die Werte generell eher etwas tiefer als bei den kognitiven Aspekten. Dies könnte damit zusammenhängen, dass Lernende mit elaborierteren Fähigkeiten zum mathematischen Argumentieren auch ohne Rubric motiviert und fähig sind, sich selber zielorientiert zu steuern (Barnes, 2019). Lehrpersonen, sind somit gut beraten, wenn sie in ihrem Unterricht zum Argumentieren vermehrt Peer- und Selbstbeurteilungen entlang eines Rubrics einfordern, damit insbesondere Lernende mit tieferen Fähigkeiten in der Selbstregulation die Kriterien des Rubrics gut verstehen und nutzen lernen (siehe auch Smit et al., 2017). Dies könnte bei diesen zu einer höheren Motivation bei Argumentationsaufgaben führen. Wie Brunner (2014) postuliert sind eben auch metakognitive, motivationale und emotionale Aspekte Teil der Argumentationskompetenz.

Es ist zudem bedeutsam, einige auch generelle unterrichtspraktische Überlegungen dahingehend anzustellen, wie die Einschätzung zur Nützlichkeit für alle Schülerinnen und Schüler noch etwas gesteigert werden könnte. Einleitend gilt es festzuhalten: Rubrics sind nicht eine Methode „per se“ (Brookhart und Chen (2015)). Sie bedürfen der Einbettung in den Unterricht bzw. in eine Unterrichtskultur, in welcher Prozesse zur formativen Beurteilung selbstverständlich ge- und belebt werden. Dazu gehört insbesondere Feedback durch die Lehrperson auf die Frage „Wohin geht es als nächstes“ gemäss dem Modell von Hattie und Timperley (2007). Olsson und D'Arcy (2022) zeigten, dass Schülerinnen und Schüler, die ein auf ihre Denkprozesse gerichtetes Feedback erhielten, ihre Lösungsversuche begründeten, während die andere Gruppe mit Feedback zur Lösung häufig die Lösungsvorschläge der Lehrperson repetierte. Selbstgesteuerte, dialogische Lernsituationen ermöglichen den Schülerinnen und Schüler beim Bearbeiten von Argumentationsaufgaben wertvolle Erfahrungen, weil sie dabei lernen, Feedback der Lehrperson und auch der Peers zu verstehen und zu nutzen (Adie et al., 2018). In solchen Lernanlässen zum Argumentieren trägt die in Rubrics bereit gestellte Zieltransparenz wesentlich zu Sicherheiten bezüglich Selbstregulierung und Wirksamkeit von Feedback durch die Lehrperson bei, wie es von Wyatt-Smith und Adie (2019) als Teil einer professionellen Praxis im Mathematikunterricht gefordert wird.

Einschränkungen und Ausblick

Es ist eine Einschränkung der Studie, dass die Daten nur aus Selbsteinschätzungen der Schülerinnen und Schüler bestehen. Allerdings gilt es zu beachten, dass gemäss neuerer Theorien zu Feedback die Rolle der Feedbackabnehmenden vermehrt Beachtung geschenkt werden sollte (Yan & Carless, 2021). Die entscheidende Rolle der Schülerinnen und Schüler bei der Verarbeitung der Rückmeldung zum Argumentieren muss anerkannt werden, da die Rückmeldung nur dann wirksam sein kann, wenn sich die Schülerinnen und Schüler aktiv damit befassen (Brooks et al., 2021). Insofern macht es keinen Sinn, die Selbsteinschätzungen als subjektiv herabzustufen. Weiter kann man nur teilweise von einer Wirkung zwischen Nützlichkeit und Selbstregulierung sprechen, da es weder möglich ist, einen Längsschnitt zur Rubric-Nützlichkeit noch eine Kontrollgruppe in die Analyse einzubeziehen. Ausblickend weisen wir darauf hin, dass ein kurzfristiger Zusammenhang der Nützlichkeitseinschätzung mit der

Mathematikleistung nicht gezeigt werden konnte, dass wir aber in einer neueren Studie eine Wirkung von wahrgenommenem Feedback auf die Leistung im mathematischen Argumentieren nachweisen konnten (Smit et al., 2022).

Die im theoretischen Modell postulierte indirekte Wirkung des Rubrics auf die mathematischen Kompetenzen muss noch präziser und vermutlich auch in einem längeren Zeitraum geklärt werden. Eine einmalige Verwendung eines Rubrics in einem Projekt reicht eher nicht, um das Instrument als festen Bestandteil des Lehrens und Lernens im Unterricht zu verankern (siehe auch Smit et al., 2019). In der vorliegenden Studie waren nur wenige Lehrpersonen aus dem gleichen Schulteam. Interessant für die Lehrerbildung und Forschung könnte aber eine schulhausinterne Diskussion des Verständnisses des mathematischen Argumentierens anhand des formativen Gebrauchs des Rubrics sein. Herbert et al. (2022) empfehlen dies unter Bezug von Schülerlösungen zu tun, damit die Lehrpersonen so lernen das eigene Feedback für diese komplexe Kompetenz zielgerichteter für die individuellen Schülerinnen und Schüler zu formulieren.

Anmerkungen:

Das Projekt wurde vom Schweizerischen Nationalfonds (Nr. 149386) unterstützt.

Literatur

- Adie, L., Kleij, F., & Cumming, J. (2018). The development and application of coding frameworks to explore dialogic feedback interactions and self-regulated learning. *British Educational Research Journal*, 44(4), 704-723. <https://doi.org/10.1002/berj.3463>
- Andrade, H. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, 4, 87.
- Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 3-13.
- Andrade, H., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-Referenced Self-Assessment and Self-Efficacy for Writing. *The Journal of Educational Research*, 102(4), 287-302. <https://doi.org/10.3200/JOER.102.4.287-302>
- Ball, D. L., & Bass, H. (2000). Making believe: The collective construction of public mathematical knowledge in the elementary classroom. In D. C. Philipps (Ed.), *Constructivism in education: Opinions and second opinions on controversial issues. Yearbook of the National Society for the Study of Education* (pp. 193-224). University of Chicago Press.
- Barnes, A. (2019). Perseverance in mathematical reasoning: the role of children's conative focus in the productive interplay between cognition and affect. *Research in Mathematics Education*, 21(3), 271-294. <https://doi.org/10.1080/14794802.2019.1590229>
- Bersch, S. (2023). *Mathematisches Argumentieren im Analysisunterricht: Explorative Studien zu Herausforderungen und Lösungsansätzen aus der Perspektive von Lehrkräften*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-40969-2_2
- Bezold, A. (2009). *Förderung von Argumentationskompetenzen durch selbstdifferenzierende Lernangebote. Eine Studie im Mathematikunterricht der Grundschule*. Kovac.
- Bezold, A. (2010). *Mathematisches Argumentieren in der Grundschule fördern. Was Lehrkräfte dazu beitragen können*. Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN). http://www.sinus-an-grundschulen.de/fileadmin/uploads/Material_aus_SGS/Handreichung_Mathe_Bezold.pdf
- Bezold, A. (2012). Förderung von Argumentationskompetenzen auf der Grundlage von Forscheraufgaben: Eine empirische Studie im Mathematikunterricht der Grundschule. *mathematica didactica*, 35, 73-103.
- Blanton, M. L., & Kaput, J. J. (2005). Characterizing a classroom practice that promotes algebraic reasoning. *Journal for Research in Mathematics Education*, 36(5), 412-446.
- Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational research and evaluation*, 31, 445-457.
- Boonen, A., Reed, H., Schoonenboom, J., & Jolles, J. (2016). It's not a math lesson-we're learning to draw! Teachers' use of visual representations in instructing word problem solving in sixth grade of elementary school. *Frontline Learning Research*, 4(5), 34-61.
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Ascd.
- Brookhart, S. M. (2018). Appropriate Criteria: Key to Effective Rubrics [Review]. *Frontiers in Education*, 3(22). <https://doi.org/10.3389/feduc.2018.00022>
- Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343-368. <https://doi.org/10.1080/00131911.2014.929565>
- Brooks, C., Burton, R., van der Kleij, F., Ablaza, C., Carroll, A., Hattie, J., & Neill, S. (2021). Teachers activating learners: The effects of a student-centred feedback approach on writing achievement. *Teaching and Teacher Education*, 105, 103387. <https://doi.org/https://doi.org/10.1016/j.tate.2021.103387>
- Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen*. Springer.
- Brunner, E., Jullier, R., & Lampart, J. (2019). Aufgabenangebot zum mathematischen Begründen in je zwei aktuellen Mathematikbüchern. *Schweizer Zeitschrift für Bildungswissenschaften*, 41(3), 647-664.
- Brunner, E., Lampart, J., & Jullier, R. (2022). Schriftliches mathematisches Argumentieren in zwei unterschiedlichen Inhaltsbereichen in den Jahrgangsstufen 4–6. *Journal für Mathematik-Didaktik*. <https://doi.org/10.1007/s13138-022-00197-8>
- Bürgermeister, A. (2013). *Leistungsbeurteilung im Mathematikunterricht: Bedingungen und Effekte von Beurteilungspraxis und Beurteilungsgenauigkeit*. Waxmann Verlag.

- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, 65(3), 245-281.
- Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891.
- Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014). Identifying kinds of reasoning in collective argumentation. *Mathematical Thinking and Learning*, 16(3), 181-200.
- Danielson, C. (1997). *A Collection of Performance Tasks and Rubrics. Middle School Mathematics*. Eye on Education. <http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED421541>
- Davidson, A., Herbert, S., & Bragg, L. A. (2019). Supporting Elementary Teachers' Planning and Assessing of Mathematical Reasoning. *International Journal of Science and Mathematics Education*, 17(6), 1151-1171. <https://doi.org/10.1007/s10763-018-9904-0>
- de Leeuw, J. (2016). Rubrics and Exemplars in Writing Assessment. In S. Scott, E. D. Scott, & F. C. Webber (Eds.), *Leadership of Assessment, Inclusion, and Learning* (pp. 89-110). Springer International Publishing. https://doi.org/10.1007/978-3-319-23347-5_4
- English, N., Robertson, P., Gillis, S., & Graham, L. (2022). Rubrics and formative assessment in K-12 education: A scoping review of literature. *International Journal of Educational Research*, 113, 101964. <https://doi.org/https://doi.org/10.1016/j.ijer.2022.101964>
- Ginsburg, H. P. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development*, 52, 109-128.
- Gregori-Giralt, E., & Menéndez-Varela, J.-L. (2018). The reliability and sources of error of using rubrics-based assessment for student projects. *Assessment & Evaluation in Higher Education*, 43(3), 488-499. <https://doi.org/10.1080/02602938.2017.1360838>
- Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Press.
- Hanna, G. (2014). Mathematical proof, argumentation, and reasoning. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education*. (pp. 404-408). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_102
- Hattie, J., Crivelli, J., Van Gompel, K., West-Smith, P., & Wike, K. (2021). Feedback That Leads to Improvement in Student Essays: Testing the Hypothesis that "Where to Next" Feedback is Most Powerful [Original Research]. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.645758>
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In *Handbook of research on learning and instruction* (pp. 263-285). Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Helmke, A. (2015). *Unterrichtsqualität und Lehrerprofessionalität* (6th ed.). Kallmeyer.
- Herbert, S. (2021). Overcoming challenges in assessing mathematical reasoning. *Australian Journal of Teacher Education*, 46(8).
- Herbert, S., Vale, C., White, P., & Bragg, L. A. (2022). Engagement with a formative assessment rubric: A case of mathematical reasoning. *International Journal of Educational Research*, 111, 101899. <https://doi.org/https://doi.org/10.1016/j.ijer.2021.101899>
- Herbert, S., & Williams, G. (2023). Eliciting mathematical reasoning during early primary problem solving. *Mathematics Education Research Journal*, 35(1), 77-103. <https://doi.org/10.1007/s13394-021-00376-9>
- Holmes, V.-L., & Hwang, Y. (2016). Exploring the effects of project-based learning in secondary mathematics education. *The Journal of Educational Research*, 109(5), 449-463. <https://doi.org/10.1080/00220671.2014.979911>
- Hu, L.-T., & 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.
- Hussmann, S., Leuders, T., & Prediger, S. (2007). Schülerleistungen verstehen - Diagnose im Alltag. *Praxis der Mathematik in der Schule*, 49, 1-8.
- Jahnke, H. N., & Ufer, S. (2015). Argumentieren und Beweisen. In R. Bruder, L. Hefendehl-Hebeker, B. Schmidt-Thieme, & H.-G. Weigand (Eds.), *Handbuch der Mathematikdidaktik* (pp. 331-355). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35119-8_12
- Jeannotte, D., & Kieran, C. (2017). A conceptual model of mathematical reasoning for school mathematics. *Educational Studies in Mathematics*, 96(1), 1-16.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144. <http://www.scopus.com/inward/record.url?eid=2-s2.0-36049027816&partnerID=40&md5=d404fb17aaf9e5ee5c2ba743069b20fb>

- Keiser, J. M., & Lambdin, D. V. (1996). The clock is ticking: Time constraint issues in mathematics teaching reform. *The Journal of Educational Research*, 90(1), 23-31. <https://doi.org/10.1080/00220671.1996.9944440>
- Keller, S. (2011). Beurteilungsraster und Kompetenzmodelle. In W. Sacher & F. Winter (Eds.), *Diagnose und Beurteilung von Schülerleistungen* (pp. 143-160). Schneider Verlag Hohengehren.
- Kempert, S., Saalbach, H., & Hardy, I. (2011). Cognitive benefits and costs of bilingualism in elementary school students: The case of mathematical word problems. *Journal of Educational Psychology*, 103(3), 547.
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National academic press.
- Klafki, W. (2007). *Neue Studien zur Bildungstheorie und Didaktik* (6. ed.). Beltz.
- Knudsen, J., Lara-Meloy, T., Stevens Stallworth, H., & Wise Rutstein, D. (2014). Advice for mathematical argumentation. *Mathematics Teaching in the Middle School*, 19(8), 494-500. <https://doi.org/10.5951/mathteacmiddscho.19.8.0494>
- Kocakülal, M. S. (2010). Development and Application of a Rubric for Evaluating Students' Performance on Newton's Laws of Motion [journal article]. *Journal of Science Education Technology Teacher*, 19(2), 146-164. <https://doi.org/10.1007/s10956-009-9188-9>
- Linneweber-Lammerskitten, H., Wälti, B., Smit, R., Marc, V., Pochon, L.-O., Frapolli, A., Cadorin, L., Moser Opitz, E., Hirt, U., & Keller, R. (2010). *Basisstandards für die Mathematik. Unterlagen für den Anhörungsprozess*. 25.10.2010.
- Lithner, J. (2000). Mathematical Reasoning in Task Solving. *Educational Studies in Mathematics*, 41(2), 165-190. <http://www.jstor.org/stable/3483188>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.
- Lovorn, M. G., & Rezaei, A. R. (2011). Assessing the Assessment: Rubrics Training for Pre-Service and New In-Service Teachers. *Practical Assessment, Research & Evaluation*, 16(16), 1-18.
- Meier, S. L., Rich, B. S., & Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: Factors related to discrepancies in scoring. *Assessment in Education*, 13(01), 69-95.
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research & Evaluation*, 7(3).
- Muthén, B. O. (2010). *Bayesian analysis in Mplus: A brief introduction*. <https://www.statmodel.com/download/IntroBayesVersion%203.pdf>
- Muthén, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological methods*, 17(3), 313-335. <https://doi.org/10.1037/a0026802>
- Olsson, J., & D'Arcy, D. (2022). Students' reasoning and feedback from a teacher. *Nordisk matematikdidaktikk, NOMAD: [Nordic Studies in Mathematics Education]*, 27(1), 27-49.
- Panadero, E., Alonso-Tapia, J., & Huertas, J.-A. (2014). Rubrics vs. self-assessment scripts: effects on first year university students' self-regulation and performance. *Infancia y Aprendizaje*, 37(1), 149-183. <https://doi.org/10.1080/02103702.2014.881655>
- Panadero, E., Alonso-Tapia, J., & Reche, E. (2013). Rubrics vs. self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers. *Studies in Educational Evaluation*, 39(3), 125-132. <https://doi.org/http://dx.doi.org/10.1016/j.stueduc.2013.04.001>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. <https://doi.org/http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806-813. <https://doi.org/http://dx.doi.org/10.1016/j.lindif.2012.04.007>
- Pape, S. J., & Smith, C. (2002). Self-Regulating Mathematics Skills. *Theory into Practice*, 41(2), 93-101. https://doi.org/10.1207/s15430421tip4102_5
- Prenzel, M., Sälzer, C., Klieme, E., & Köller, O. (2013). *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Waxmann Verlag.
- Purdie, N., Hattie, J., & Douglas, G. (1996). Student conceptions of learning and their use of self-regulated learning strategies: A cross-cultural comparison. *Journal of Educational Psychology*, 88(1), 87.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5-6 mathematics effects on problem-solving achievement. *Educational Assessment*, 8(1), 43-58.
- Saddler, B., & Andrade, H. (2004). The writing rubric. *Educational Leadership*, 62(2), 48-52.
- Schafer, W. D., Swanson, G., Bené, N., & Newberry, G. (2001). Effects of Teacher Knowledge of Rubrics on Student Achievement in Four Content Areas. *Applied Measurement in Education*, 14(2), 151-170.

- Schmidinger, E., Hofmann, F., & Stern, T. (2016). Leistungsbeurteilung unter Berücksichtigung ihrer formativen Funktion. In F. Eder, K. Krainer, A. Seel, C. Schreiner, & C. Spiel (Eds.), *Nationaler Bildungsbericht Österreich 2015, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (pp. 59-94). Bundesinstitut BIFIE.
- Schmidt, C. (2020). Empirische Befunde zum formativen Assessment. In *Formatives Assessment in der Grundschule: Konzept, Einschätzungen der Lehrkräfte und Zusammenhänge* (pp. 63-81). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-26921-0_4
- Schweizerische Konferenz der Erziehungsdirektoren (EDK). (2011). Grundkompetenzen für die Mathematik. Nationale Bildungsstandards. http://edudoc.ch/record/96784/files/grundkomp_math_d.pdf
- Sfard, A. (2001). There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning. *Educational Studies in Mathematics*, 46(1/3), 13-57. <http://www.jstor.org/stable/3483239>
- Shilo, A., & Kramarski, B. (2019). Mathematical-metacognitive discourse: how can it be developed among teachers and their students? Empirical evidence from a videotaped lesson and two case studies [journal article]. *ZDM*, 51, 625-640. <https://doi.org/10.1007/s11858-018-01016-6>
- Smit, R. (2009). *Die formative Beurteilung und ihr Nutzen für die Entwicklung von Lernkompetenz*. Schneider Verlag Hohengehren.
- Smit, R., Bachmann, P., Blum, V., Birri, T., & Hess, K. (2017). Effects of a rubric for mathematical reasoning on teaching and learning in primary school. *Instructional Science*, 45(5), 603-622. <https://doi.org/10.1007/s11251-017-9416-2>
- Smit, R., & Birri, T. (2014). Assuring the quality of standards-oriented classroom assessment with rubrics for complex competencies. *Studies in Educational Evaluation*, 43(December), 5-13. <https://doi.org/10.1016/j.stueduc.2014.02.002>
- Smit, R., Dober, H., Hess, K., Bachmann, P., & Birri, T. (2022). Supporting primary students' mathematical reasoning practice: the effects of formative feedback and the mediating role of self-efficacy. *Research in Mathematics Education*, 1-24. <https://doi.org/10.1080/14794802.2022.2062780>
- Smit, R., Hess, K., Bachmann, P., Blum, V., & Birri, T. (2019). What happens after the intervention? Results from teacher professional development in employing mathematical reasoning tasks and a supporting rubric. *Frontiers in Education*, 3. <https://doi.org/10.3389/educ.2018.00113>
- Stanat, P., & Christensen, G. (2006). *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003*. OECD.
- Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, 33(2), 455-488.
- Stobart, G. (2006). The validity of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 133-146). Sage.
- Stylianides, G. J. (2008). An analytic framework of reasoning-and-proving. *For the Learning of Mathematics*, 28(1), 9-16.
- Tang, X., Coffey, J., & Levin, D. M. (2015). Reconsidering the use of scoring rubrics in biology instruction. *The American Biology Teacher*, 77(9), 669-675.
- Thompson, D. R., & Schultz-Ferrel, K. (2008). *Introduction to reasoning and proof. Grades 6-8*. Heinemann.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press.
- Turley, E. D., & Gallagher, C. W. (2008). On the "Uses" of Rubrics: Reframing the Great Rubric Debate. *English Journal*, 97(4), 87-92. <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ788590>
- Turner, S. L. (2014). Creating an assessment-centered classroom: Five essential assessment strategies to support middle grades student learning and achievement. *Middle School Journal*, 45(5), 3-16.
- Ufer, S., & Kramer, J. (2012). Die Kompetenz mathematisch Argumentieren. In W. Blum, C. Drüke-Noe, R. Hartung, & O. Köller (Eds.), *Bildungsstandards Mathematik: konkret*. Westermann Schroedel Diesterweg Schöningh Winklers.
- Viholainen, A. (2011). The view of mathematics and argumentation. Proceedings of the 7th Congress of the European Society for Research in Mathematics Education, 9th-13th February, University of Rzeszów, Poland.
- Whitenack, J., & Yackel, E. (2002). Making mathematical arguments in the primary grades: the importance of explaining and justifying ideas. *Teaching Children Mathematics*, 8(9), 524-528.
- Wollenschläger, M., Hattie, J., Machts, N., Möller, J., & Harms, U. (2016). What makes rubrics effective in teacher-feedback? Transparency of learning goals is not enough. *Contemporary Educational Psychology*, 44-45, 1-11. <https://doi.org/http://dx.doi.org/10.1016/j.cedpsych.2015.11.003>

- Wyatt-Smith, C., & Adie, L. (2019). The development of students' evaluative expertise: enabling conditions for integrating criteria into pedagogic practice. *Journal of Curriculum Studies*, 1-21. <https://doi.org/10.1080/00220272.2019.1624831>
- Yan, Z., & Carless, D. (2021). Self-assessment is about more than self: the enabling role of feedback literacy. *Assessment & Evaluation in Higher Education*, 1-13. <https://doi.org/10.1080/02602938.2021.2001431>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70.
- Zimmerman, B. J., & Pons, M. M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, 23(4), 614-628.

Anhang:

Skala Selbstregulation bei Textaufgaben (t2)

Item	Formulierung:	M	SD	r _{itc}
selm1	Beim Lösen von Textaufgaben plane ich mein Vorgehen genau.	4.05	1.13	.55
selm3	Wenn ich eine Textaufgabe lösen muss, versuche ich zuerst herauszufinden, welches die wichtigsten Dinge sind.	4.52	1.19	.44
selm4	Es fällt mir leicht einzuschätzen, ob ich eine Textaufgabe richtig gelöst habe.	3.84	1.17	.32
selm6	Wenn ich eine schwierige Textaufgabe lösen muss, habe ich viele Tricks um sie anzupacken.	3.87	1.25	.47
selm7	Wenn ich schwierige Textaufgaben lösen muss, überlege ich, wie ich bei ähnlichen Aufgaben vorgegangen bin.	4.23	1.23	.43
selm8	Bei Textaufgaben mache ich Zeichnungen, um besser zu verstehen, wie ich rechnen muss.	3.44	1.42	.49
selm9	Bei Textaufgaben mache ich Zeichnungen, um zu zeigen, wie ich gerechnet habe.	3.36	1.41	.54

Anmerkungen. Die Antwortkategorien lauten: 1 (völlig unzutreffend), 2 (weitgehend unzutreffend), 3 (eher unzutreffend), 4 (eher zutreffend), 5 (weitgehend zutreffend), 6 (völlig zutreffend); N = 762; t2: $\alpha = 0.75$

Skala Lehrpersonenfeedback zur Selbstregulation (t1)

Item	Formulierung:	M	SD	r _{itc}
srlev1	Meine Lehrperson sagt mir, wie ich die Lösung selbst überprüfen kann.	3.72	1.24	.42
srlev2	Meine Lehrperson ermutigt mich, selbst zu überlegen, ob ich auf dem richtigen Weg bin.	3.59	1.34	.47
srlev3	Meine Lehrperson sagt mir, ich solle überlegen, ob ich die Aufgabe gut bearbeitet habe.	3.20	1.23	.47
srlev4	Meine Lehrperson fragt mich beim Arbeiten, ob ich glaube, das Ziel schon erreicht zu haben.	2.50	1.31	.43

Anmerkungen. Die Antwortkategorien lauten: 1 (nie) 2 (selten) 3 (manchmal) 4 (öfters) 5 (viel) 6 (sehr viel); N = 762; t2: $\alpha = 0.67$