



Rubrics, ein Instrument zur Förderung von Selbstregulation und Selbstbeurteilung

Patricia Bachmann, Robbert Smit

Pädagogische Hochschule St. Gallen
patricia.bachmann@phsg.ch

EINGEREICHT 09 APR 2019

ÜBERARBEITET 24 SEP 2019

ANGENOMMEN 29 SEP 2019

Rubrics (Beurteilungsraster) helfen bei der Selbstregulation und Selbstbeurteilung, da Lernziele anhand klarer Kriterien aufgezeigt werden und Lernende sehen, wo sie bezüglich der Zielerreichung stehen. In einem Forschungsprojekt mit quasi-experimentellem Design der Pädagogischen Hochschulen St. Gallen und Zug (CH) wurde in 44 Klassen (5. und 6. Jahrgangsstufe) u. a. Selbstregulation und Selbstbeurteilung bei den Lernenden im Bereich des mathematischen Argumentierens untersucht. Die Hälfte der Klassen nutzte beim Arbeiten an Argumentationsaufgaben und bei der Selbstbeurteilung einen Rubric. Im Fokus stand die Frage, welche Effekte sich bei den Lernenden durch den Einsatz eines Rubrics auf die Selbstregulation und Selbstbeurteilung zeigen. Zusätzlich wurde geprüft, wie nützlich Lernende den Rubric einschätzen und ob es Unterschiede bei Lernenden mit hohen und niedrigen Mathematikleistungen gibt. Die Daten wurden mittels Fragebögen, einer Mathematikaufgabe zur Selbstbeurteilung und einem Mathematiktest zu mathematischem Argumentieren erhoben. Die Auswertungen zeigen u. a., dass die Lernenden der Interventionsgruppe ihre Fähigkeiten zur Selbstregulation nach Ende der Intervention nicht signifikant höher einschätzen als jene der Kontrollgruppe. Sie gaben jedoch an, sich im Unterricht häufiger selbst zu beurteilen. Bezüglich der Genauigkeit der Selbstbeurteilung schätzten sich die Lernenden der Interventionsgruppe nicht signifikant präziser ein. In Bezug auf die Unterschiede zwischen Lernenden mit hohen und tiefen Mathematikleistungen sind die Ergebnisse ebenfalls nicht signifikant. Im Diskussionsteil werden Hinweise für die Nutzung des Rubrics im Unterricht diskutiert.

SCHLÜSSELWÖRTER: Rubric, Selbstregulation, Selbstbeurteilung,
mathematisches Argumentieren

1. Einleitung

In den "Empfehlungen zur Arbeit in der Grundschule" der Kultusministerkonferenz (KMK, 2015), wird erläutert, dass in der Schule Kompetenzen erworben werden sollen, welche die Basis für die selbständige Kulturaneignung und lebenslanges Lernen bilden. Dabei müssen einerseits fachliche wie auch überfachliche Kompetenzen ausgebildet werden (Fend, 2008; Reusser, 2011). Bedeutsame überfachliche

Kompetenzen sind u.a. Selbstregulation des Lernens und Selbstbeurteilung. Lohbeck, Petermann und Petermann (2015) erläutern, dass Schülerinnen und Schüler ab einer mittleren Altersstufe (ab neun bis zwölf Jahren) grundsätzlich in der Lage sein sollten, ihr Lernverhalten zu beurteilen. Es zeigt sich jedoch, dass es etlichen Lernenden dieser Altersstufe schwerfällt, sich differenziert selbst zu beurteilen oder Aussagen zu ihren Leistungen zu machen (Brouër, 2014; Kostons, van Gog, & Paas, 2011). Es wird zudem angenommen, dass Selbstregulation und Selbstbeurteilen im Unterricht auf der Primarstufe eher selten durchgeführt werden (Brown & Harris, 2012; Dignath, Buettner, & Langfeldt, 2008). Es wird auch moniert, Lehrpersonen möchten beispielsweise die Verantwortung für die Lehr-Lern-Prozesse und deren Beurteilung zu wenig an die Lernenden abgeben (Andrade, 2010).

Vor diesem Hintergrund stellt sich die Frage, ob Rubrics (Beurteilungsraster) verwendet werden können, um Selbstregulation und Selbstbeurteilung im Unterricht zu unterstützen. Rubrics können als Instrument im Rahmen der formativen Beurteilung eingesetzt werden. Sie zeigen den Lernenden u.a. auf, welche Lernziele die Lehrperson setzt, wo die Lernenden diesbezüglich stehen und auf welchem Weg die Lernziele erreicht werden können (Arter & Chappuis, 2006; Ross, 2006). Im Projekt „Lernen mit Rubrics“ zweier pädagogischer Hochschulen aus der Schweiz wurde der Nutzen eines Rubrics bei der Förderung der überfachlichen Kompetenzen Selbstregulation und Selbstbeurteilung im Bereich des mathematischen Argumentierens untersucht. Gemäss Brouër (2014) gibt es nur wenige empirische Studien mit Lernenden der obligatorischen Schule, die sich mit der Selbstbeurteilung bzw. mit deren Nutzung für die Selbstregulation auseinandersetzen. Insbesondere mit Lernenden der Primarstufe und unteren Jahrgänge der Sekundarstufe 1 liegen nur wenige Studien vor (Andrade, 2019). Einige Studien zur Nutzung des Rubrics als Instrument zur Förderung der Selbstbeurteilung im sprachlichen Bereich wurden von Andrade (z.B. Andrade, 2000) durchgeführt. Weitere Untersuchungen in anderen Fachbereichen, beispielsweise in Mathematik, werden gewünscht (Andrade, Du, & Mycek, 2010). Im vorliegenden Beitrag wird der Frage nachgegangen, welche Effekte sich bei den Schülerinnen und Schülern durch den Einsatz von Rubrics in Hinblick auf die Kompetenzen Selbstregulation und Selbstbeurteilung im Bereich des mathematischen Argumentierens zeigen.

2. Theoretischer Hintergrund

2.1 Selbstregulation und Selbstbeurteilung als zentrale überfachliche Kompetenzen

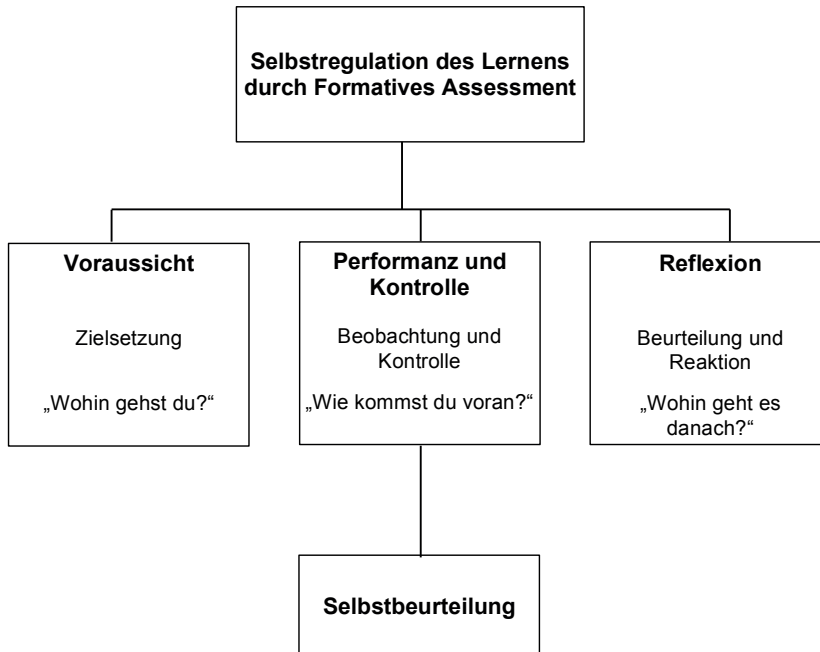
Zentral erscheint die Erkenntnis, dass in der Schule zusätzlich zu den fachlichen Kompetenzen überfachliche Kompetenzen gefördert werden müssen, um die Persönlichkeitsentwicklung der Lernenden und ihre Denk- und Wahrnehmungsfä-

higkeiten positiv zu beeinflussen. Kinder sollen mit ihrer inneren Welt kompetent umgehen, um autonom handeln und Verantwortung übernehmen zu können (Fend, 2008).

Auch im österreichischen Schulsystem werden überfachliche Kompetenzen als Ergänzung zu fachlichen gefördert, damit Lernende im Laufe des Lebens individuelle oder gesellschaftliche Aufgaben oder Probleme meistern (Eder & Hofmann, 2012). Eingeteilt werden die überfachlichen Kompetenzen in fünf Gruppen, wobei eine Gruppe u. a. personale Kompetenzen beinhaltet. Eine dieser personalen Kompetenzen ist die Selbstregulationskompetenz, welche z. B. durch Selbstbeurteilung operationalisiert werden kann (Eder & Hofmann, 2012). Kinder, welche ihre Performanz einschätzen und kontrollieren können, können ihr eigenes Lern- und Arbeitsverhalten analysieren und optimieren (Weiglhofer, 2013).

Laut Zimmerman (2002) beinhaltet Selbstregulation u. a. das Setzen von Zielen, das Aneignen von Strategien, das Überwachen der Leistung usw. Auch laut Andrade (2010) ist Selbstregulation beim Lernen ein Prozess, bei welchem es um persönliche Zielsetzung sowie um Überwachung, Regulierung und Kontrolle von Kognition, Motivation und des Verhaltens geht. Dies sind zwei mögliche Beschreibungen von Selbstregulation. Es existieren bereits verschiedene Definitionen und Modelle (Boekaerts & Corno, 2005). Selbstbeurteilung ist gemäss Andrade (2010) wiederum ein Prozess, bei welchem Lernende über Lernergebnisse nachdenken sowie diese Lernergebnisse in Bezug auf die Lernziele oder die gesetzten Kriterien beurteilen und überarbeiten. Andrade beschreibt vereinfachend, dass es bei Selbstregulation um das Managen des eigenen Lernprozesses geht und dass bei der Selbstbeurteilung Lernende ihre Produkte, wie beispielsweise Lösungen von mathematischen Problemen, beurteilen. Da Lernende bei der Selbstregulation und der Selbstbeurteilung über das eigene Lernen und die Qualität erstellter Produkte nachdenken und sich selbst Feedback geben, können beide Begriffe auch mit Formativem Assessment in Verbindung gebracht werden (Andrade, 2010; Nicol & Macfarlane-Dick, 2006). Formatives Assessment (Formative Beurteilung) kann als Ergänzung zur summativen Beurteilung gesehen werden. Während es bei der summativen Beurteilung um eine abschliessende Bewertung nach einem Lernprozess geht, steht bei der formativen Beurteilung eine Bewertung während des Lernprozesses im Vordergrund, mit dem Ziel, diesen zu optimieren (Maier, 2010). Im Modell von Andrade (2010) (s. Abbildung 1) werden die drei Begriffe Formative Beurteilung, Selbstregulation und Selbstbeurteilung folgendermassen in Verbindung gebracht: Selbstbeurteilung ist Teil der Formativen Beurteilung. Formative Beurteilung ist Teil der Selbstregulation, was bedeutet, dass Selbstbeurteilung als Bestandteil der Formativen Beurteilung und der Selbstregulation gesehen wird.

ABB. 1. Selbstreguliertes Lernen und formatives Assessment (Andrade, 2010, S. 96)



Laut Brown und Harris (2012) verbessern Lernende ihre Selbstregulation durch Selbstbeurteilung, da sie im Sinne des Modells von Andrade z.B. lernen, vorzuschauen, sich Ziele zu setzen und diese beim Lernen zu verfolgen. Indem die Lernenden ihren Lernprozess und ihre Performanz beobachten und selbst beurteilen, können sie die aktuelle Leistung mit ihren gesetzten Lernzielen vergleichen und erfahren, wie sie vorankommen (Ross, 2006). Dadurch können die nächsten Arbeitsschritte einfacher geplant werden. Auch kann die Motivation bei den Lernenden gesteigert werden, da sie das Gefühl bekommen, u.a. die Kontrolle über das eigene Lernen zu haben. Dadurch bekommen Schülerinnen und Schüler u.a. ein realistisches Selbstbild und reduzieren Ohnmachtsgefühle und Angst (Panadero & Jonsson, 2013).

2.2 Schwierigkeiten beim Regulieren des eigenen Lernprozesses und bei der Selbstbeurteilung

Damit Schülerinnen und Schüler Selbstregulation lernen, muss im Unterricht mit allen Lernenden und mit geeigneten Strategien und Hilfsmitteln gezielt trainiert werden (Meusen-Beekman, Joosten-ten Brinke, & Boshuizen, 2016). Auch Selbstbeurteilung ist kein Selbstläufer und muss von den Lehrpersonen thematisiert, gecoached und unterstützt werden (Harris & Brown, 2013; Ross, 2006). Selbstbe-

urteilung muss beispielsweise während des Lernprozesses erfolgen und sollte die Schülerinnen und Schüler kognitiv herausfordern, damit sie lernwirksam ist. So ist das Beurteilen ohne Kriterien, z.B. beim Setzen von Smileys, weniger herausfordernd als die Selbstbeurteilung anhand objektiver Kriterien (Brown & Harris, 2012).

Wie Brown und Harris (2012) jedoch schreiben, wird Selbstbeurteilung im Unterricht eher selten durchgeführt. Schülerinnen und Schüler lernen z.B. zu wenig, ihre eigenen Lernprozesse wahrzunehmen und zu beurteilen. Gründe dafür sind, dass Lehrpersonen ihnen bezüglich der Beurteilung keine Verantwortung übertragen (Andrade, 2010; Brouër, 2014) oder Lehrpersonen nicht wissen, wie sie Strategien und Hilfsmittel anwenden sollen (Dignath, Buetttner, & Langfeldt, 2008).

Weiter muss beachtet werden, dass Lernende mit niedrigen schulischen Leistungen mehr Mühe haben, sich selbst zu beurteilen und ihre Einschätzungen oft weniger genau ausfallen, als solche von Lernenden mit höheren Leistungen und besseren Fähigkeiten im Selbstbeurteilen (Brown & Harris, 2012). Dies kann daran liegen, dass diese Lernenden unrealistische Vorstellungen der eigenen Fähigkeiten haben, weil sie beispielsweise zu selbstbewusst sind, ihre Leistung aufgrund von ungenauem Feedback der Lehrperson falsch einschätzen oder einfach nicht in der Lage sind, ihre Leistung und ihr Lernen zu beurteilen (Panadero, Brown & Strijbos, 2016). Solche Lernenden brauchen mehr Unterstützung durch die Lehrperson und profitieren mehr von differenzierteren und kognitiv anspruchsvolleren Formen der Selbstbeurteilung (Andrade, Du & Mycek, 2010; Brown & Harris, 2012).

2.3 Selbstregulation und Selbstbeurteilung mit Rubrics üben

Da der Begriff Rubric (Beurteilungsraster) nicht eindeutig definiert ist und es verschiedene Arten von Rubrics gibt (Dawson, 2015), soll hier eine kurze Erläuterung erfolgen. Rubrics sind Werkzeuge, welche im Rahmen der formativen Beurteilung eingesetzt werden können (Panadero, Tapia, & Huertas, 2012). Rubrics sind in der Regel Tabellen (s. Abbildung 2), welche zwei Hauptmerkmale aufweisen. Erstens werden in der ersten Spalte verschiedene Qualitätsdimensionen (Kriterien) aufgelistet, welche in Zusammenhang mit einer konkreten Anforderungssituation stehen wie z.B. dem Lösen von Aufgaben im Bereich des mathematischen Argumentierens. Zweitens werden in den Zeilen verschiedene Niveaustufen der Qualitätsdimensionen beschrieben oder mit Zahlen oder einzelnen Wörtern umschrieben (Andrade, 2000). Werden die Niveaus umschrieben, wie dies in Abbildung 2 ersichtlich ist, wird von einem analytischen Rubric gesprochen.

Eine Unterscheidung muss zwischen Rubrics und Kompetenzrastern gemacht werden. In Kompetenzrastern werden übergeordnete Wissensbereiche, Lernziele und Handlungen auf unterschiedlichen Kompetenzstufen (in Zyklen) dargestellt und ermöglichen einen Gesamtüberblick. Beurteilungsraster hingegen sind näher an einzelnen Aufgaben orientiert und begleiten Lernprozesse (Maier, 2015).

Panadero und Jonsson (2013) konnten mithilfe einer systematischen Literatur-Übersicht aufzeigen, dass Rubrics die Selbstregulation des Lernens unterstützen können. Im Sinne des Modells von Andrade (s. Abbildung 1) unterstützen Rubrics, Lernziele zu setzen und den Lernprozess zu planen. Rubrics verdeutlichen den Lernenden auch, was die Lehrperson von ihnen erwartet und schaffen dadurch Transparenz (Arter & Chappuis, 2006). Rubrics helfen auch, den Lernprozess und die Lernergebnisse zu überprüfen. Laut Andrade (2007) dienen Rubrics der Orientierung beim Lernen, unterstützen bei der Selbst- und Peerbeurteilung und helfen beim Überprüfen und Verbessern der Lernergebnisse. Auch Brown & Harris (2012) schreiben, dass Selbstbeurteilung als selbstregulative Fähigkeit mithilfe von Rubrics geübt werden kann. Rubrics können zudem bei der Reflexion des Lernens helfen. Lernende können Feedback einfacher interpretieren und umsetzen (Panadero & Jonsson, 2013). Weil Rubrics zudem die Transparenz in Bezug auf die Lernziele und Erwartungen transparent machen, können sie bei den Lernenden Angst reduzieren (Andrade & Du, 2005; Panadero & Jonsson, 2013).

Rubrics sind jedoch keine Selbstläufer. Sie können nicht einfach ausgehändigt, sondern müssen eingeführt werden (Jonsson & Svingby, 2007). Zudem sollen Rubrics von den Lernenden nicht als enger Raster verstanden werden, um einer Lehrperson die gewünschte Lösung zu liefern. Sie sollen während des Lernens anregen, über Qualität nachzudenken und eine Zielorientierung beim Lernen (anstatt Leistungsorientierung) zu entwickeln (Andrade & Du, 2005).

2.4 Mathematisches Argumentieren

Überfachliche Kompetenzen wie beispielsweise die Selbstregulation und das Selbstbeurteilen werden im Unterricht nicht automatisch gelernt, sondern gezielt beim Auseinandersetzen mit fachlichen Inhalten erworben (Reusser, 2011). Zu solchen fachlichen Kompetenzen gehört fundiertes mathematisches Argumentieren (Organisation für wirtschaftliche Zusammenarbeit und Entwicklung [OECD], 2005). Argumentieren ist eine wichtige Grundlage für die Entwicklung von mathematischen Kompetenzen in der Primarschule, die auf einem mathematischen Verstehen gründen (Nunes et al., 2007). Die Kultusministerkonferenz in Deutschland [KMK] (2004) bezeichnet „Argumentieren“ als zentrale, zu erwerbende mathematische Kompetenz und auch in den Nationalen Bildungsstandards der Schweiz wird das mathematische Argumentieren und Begründen als zentrale Kompetenz umschrieben (Erziehungsdirektorenkonferenz [EDK] (2011). Auch in Österreich beschreibt das Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE, 2010) das Kommunizieren als zentrale Kompetenz für die Grundschule und als Vorstufe für das Argumentieren und Begründen auf der Sekundarstufe (Benischek, Jagric, Jungwirth, Marounek, & Nierdtscheider, 2013). Schülerinnen und Schüler sollen beispielsweise ihr Vorgehen

ABB. 2. Rubric für Schülerinnen und Schüler zu mathematischem Argumentieren

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.

beim Rechnen beschreiben und protokollieren sowie Lösungswege vergleichen und begründen können. Obwohl das Argumentieren im Bereich Mathematik in Bildungsstandards erwähnt wird, ist oft nicht ganz klar, was unter Argumentieren zu verstehen ist und in welchem Verhältnis es zu den Begriffen Begründen und Beweisen steht. Brunner (2014) beschreibt, dass die Begriffe sehr unterschiedlich verwendet werden und in der Literatur kein eindeutiger Konsens bestehe. Eine Möglichkeit ist, Begründen als Oberbegriff bzw. als Kontinuum zu sehen und verschiedene Formen von Argumentieren und Beweisen darunter einzuordnen. So können beispielsweise "alltagsbezogenes Argumentieren, Argumentieren mit mathematischen Mitteln, logisches Argumentieren mit mathematischen Mitteln und formal-deduktives Beweisen" als Formen von Begründen gesehen werden (Brunner, 2014, S. 30).

Im vorliegenden Forschungsprojekt wird Argumentieren im Sinne von Argumentieren mit mathematischen Mitteln gesehen, bei welchem zwar mathematische Mittel in das Argumentieren miteinbezogen werden, jedoch nicht unbedingt logisches Schliessen stattfinden muss. Oft wird nämlich beim Argumentieren im Primarschulbereich nicht von deduktivem Schliessen, sondern eher von Interaktionsprozessen zwischen den Lernenden ausgegangen (Fetzer, 2011). Bereits Kinder im frühen Primarbereich können ihre mathematischen Vorstellungen begründen (Perry, Dockett, & Harley, 2007). Anlässe des Argumentierens können nach Bezold (2009) das Begründen des Vorgehens, von Vermutungen, Ergebnissen, Behauptungen oder Prognosen sowie das Aufzeigen von Regelhaftem sein. Die Kompetenz des Argumentierens wird als Grundlage für das mathematische Beweisen in den späteren Jahren gesehen (Winkelmann & Van den Heuvel-Panhuizen, 2009). Das Argumentieren kann als komplexe Tätigkeit betrachtet werden, da Lernende nicht nur rechnen, sondern auch über den Lösungsweg und das Ergebnis nachdenken, berichten und diese verteidigen müssen. In Gesprächen zwischen Lehrpersonen und Lernenden oder Lernenden untereinander wird über Lernprozesse und Lernergebnisse diskutiert und reflektiert (Bezold, 2009). Rubrics können solche Gespräche unterstützen.

Das mathematische Argumentieren soll im vorliegenden Beitrag jedoch nicht im Zentrum stehen, sondern wird als Gegenstand gesehen, um Selbstregulation und Selbstbeurteilung anzuwenden.

2.5 Fragestellungen

Aufgrund der bisherigen Darstellung soll im vorliegenden Bericht den folgenden Fragestellungen nachgegangen werden:

1. Zeigt sich bei Lernenden durch den Einsatz eines Rubrics ein Effekt auf a) die Selbstregulation und b) die Selbstbeurteilung?

2. Schätzen Lernende der Interventionsgruppe den Rubric als Hilfsmittel für das fachliche und überfachliche Lernen als nützlich ein?

3. Lassen sich Unterschiede bei Lernenden mit niedriger und hoher Mathematikleistung bezüglich der Selbstregulation, der Selbstbeurteilung und der Nützlichkeit des Rubrics erkennen?

3. Methode

3.1 Forschungsdesign

Das Forschungs- und Weiterbildungsprojekt „Lernen mit Rubrics“ mit quasi-experimentellem Design der Pädagogischen Hochschulen St.Gallen und Zug dauerte zwei Jahre von Januar 2015 bis Dezember 2016. Das Projekt wurde in 44 Klassen der 5. und 6. Jahrgangsstufe durchgeführt. Die Intervention fand von August bis November 2015 statt und beinhaltete eine neunwöchige Unterrichtsreihe zu mathematischem Argumentieren. Für das Reflektieren und Beurteilen der eigenen Lernergebnisse und Lernprozesse während der Intervention verwendeten die Klassen der Interventionsgruppe einen Rubric. Dieser wurde in einem Pilotprojekt entwickelt und im Unterricht erprobt (Smit & Birri, 2014) (s. Abbildung 2).

Vor der Intervention besuchten alle Lehrpersonen einen eintägigen Workshop zu mathematischem Argumentieren und erhielten einen Input zu Feedback im Rahmen der formativen Beurteilung. Zusätzlich erhielten die Lehrpersonen der Interventionsgruppe u.a. Informationen zur Implementierung eines Rubrics und dessen Einsatz im Rahmen der formativen Beurteilung im Bereich des mathematischen Argumentierens. Die Kontrollgruppe bekam einen Input mit dem mathematischen Handlungsaspekt Darstellen. Nach der Intervention erfolgte erneut ein halbtägiger Workshop, bei welchem die beiden Gruppen die jeweils ergänzenden Kursinputs erhielten (Arbeiten mit Rubrics oder mathematischer Handlungsaspekt Darstellen) und Erfahrungen ausgetauscht wurden.

Während der Intervention sollten sich die Lehrpersonen an ein vorgegebenes Drehbuch halten und wöchentlich eine Lektion zu mathematischem Argumentieren durchführen. Ausserdem war in beiden Gruppen Peer- und Selbstbeurteilung explizit im Ablauf vorgesehen, damit keine Konfundierung von Rubric und Peer- und Selbstbeurteilung stattfindet.

Vor und nach der Intervention wurden überfachliche Kompetenzen wie die Selbstregulation und die Selbstbeurteilung bei den Lernenden mittels eines Fragebogens erfasst. Ausserdem mussten die Lernenden in Woche 7 der Intervention eine Argumentationsaufgabe lösen und das Ergebnis bzw. den Lösungsweg selbst beurteilen. Die mathematische Argumentationskompetenz der Lernenden wurde zusätzlich mit einem Mathematiktest erfasst.

3.2 Stichprobe

Für das Forschungs- und Weiterbildungsprojekt konnten 44 Lehrpersonen aus den Kantonen St.Gallen und Zug gewonnen werden. Die Rekrutierung erfolgte mithilfe einer Kursauschreibung im Kantonalen Schulblatt und auf der Homepage der Pädagogischen Hochschule St.Gallen, durch Anfragen bei Schulleitungen sowie aufgrund persönlicher Empfehlung. 21 Lehrkräfte waren Teil der Interventionsgruppe, 23 gehörten zur Kontrollgruppe. Die Stichprobe bestand aus 24 Frauen und 20 Männern, welche im Durchschnitt 41 Jahre alt waren und ein durchschnittliches Dienstalter von 11 Jahren aufwiesen. Zehn Lehrpersonen unterrichteten in einer 5. Klasse, 17 Personen lehrten in einer 6. Klasse, während die restlichen 17 Lehrkräfte in Mehrjahrgangsklassen (5. und 6. Klasse) tätig waren. In vier Mehrjahrgangsklassen gehörten auch Lernende der 4. Jahrgangsstufe zur Klasse. Bei der Bildung der zwei Gruppen wurde darauf geachtet, dass die Lehrpersonen bezüglich Geschlecht, Alter, Jahrgangsklasse und Kanton gleichmässig verteilt waren. Insgesamt bestanden die 44 Klassen aus 23 Lernenden der 4. Klasse, 337 Lernenden der 5. Klasse und 402 Lernenden der 6. Klasse, was einer Anzahl von insgesamt 762 Schülerinnen und Schülern entspricht. 52 % waren Jungen, 48 % waren Mädchen.

3.3 Instrumente

3.3.1 Fragebogen

Die Fragebogen für die Lernenden beinhalteten neben Hintergrundvariablen u.a. Items zur Einschätzung der Überzeugungen der Schülerinnen und Schüler (z.B. Selbstregulation) und Items zur Unterrichtsqualität (z.B. Häufigkeit von Selbstbeurteilung und Nützlichkeit des Rubrics). Die Skalen für die Erfassung der Selbstregulation beziehen sich teilweise auf Items von Purdie, Hattie und Douglas (1996) oder wurden auch selbst konstruiert. Die Items zur Erfassung der Selbstbeurteilung wurden in Anlehnung an Brown, Harris und Harnett (2012) sowie Hargreaves, McCallum und Gipps (2000) formuliert. Die Items zur Erfassung der Nützlichkeit des Rubrics wurden selbst entwickelt. Mit Items zur Nützlichkeit des Rubrics wurde beispielsweise erfasst, ob der Rubric zum Verständnis einer Aufgabenlösung und zur Sicherheit beim Aufgabenlösen beiträgt oder ob mithilfe des Rubrics geklärt werden kann, welche Schwierigkeiten aufgetaucht sind (s. Tabelle 1).

Die hier verwendeten Items waren Likert-Skala-Items, welche auf einer sechsstufigen Skala eingeschätzt werden konnten (z.B. „stimmt überhaupt nicht“, „stimmt nicht“, „stimmt eher nicht“, „stimmt eher“, „stimmt“ oder „stimmt völlig“). Items, die nicht zur Skala passten (Cronbachs alpha < 0.60), wurden mit einer Ausnahme entfernt. Die Skala Selbstbeurteilung aus Schülersicht erreichte bei t₁ nur ein Cronbach's Alpha von 0.52 und bei t₂ von 0.62. Da die Skala drei sehr heterogene Items umfasst, wurde die Skala mit konfirmatorischer Faktoranalyse

positiv überprüft (alle Faktorladungen > 0.40) und anschliessend für weitere Analysen als zentrales Konzept berücksichtigt.

TAB. 1. Beschreibung der verwendeten Skalen

Name der Skala	Anzahl Items	Beispielitem	Cronbach's Alpha (τ_1/τ_2)
Selbstregulation Schüler τ_1/τ_2	7	Beim Lösen von Textaufgaben plane ich mein Vorgehen genau.	0.67/ 0.75
Selbstbeurteilung τ_1/τ_2	3	Ich muss eine Selbstbeurteilung meiner Aufgabenlösung machen.	0.52/ 0.62
Nützlichkeit des Rubrics τ_2	6	Der Rubric hilft zu verstehen, auf was es bei der Lösung der Aufgabe ankommt.	- / 0.89

Anmerkungen: N = 762/ bei Nützlichkeit n = 331

3.3.2 Selbstbeurteilung einer Einzelaufgabe

In der drittletzten Woche der Intervention mussten die Lernenden eine Mathematikaufgabe zu mathematischem Argumentieren lösen. Diese Aufgabe ist angelehnt an Aufgaben aus den Schweizer Bildungsstandards (HarmoS-Aufgaben) im Bereich des Handlungsaspekts "Argumentieren und Begründen" (EDK, 2011). Bei diesen Aufgaben geht es darum, dass Schülerinnen und Schüler Aufgaben lösen und "ihren Lösungsweg mithilfe einer Rechnung und Erläuterungen rechtfertigen". Sie können Behauptungen auch "mithilfe von Grössenangaben präzisieren und begründen" (EDK, 2011, S. 40). Anschliessend mussten die Lernenden ihre Lösung und Argumentation durch das Setzen eines Smileys auf den Stufen eins bis vier gesamthaft selbst beurteilen.

Anschliessend wurde die Einzelaufgabe (mit Lösung und Argumentation) von trainierten Raterinnen und Ratern bewertet, welche sich am Rubric resp. an einem Manual orientierten. Im Rater-Training wurde eine Interraterreliabilität von Kappa > 0.70 erreicht. Damit wurde angenommen, dass die Fremdeinschätzung als reliabel angesehen werden kann. Um zu erörtern, wie genau sich die Lernenden selbst einschätzen können, wurde das Ergebnis der Fremd- und Selbstbeurteilung miteinander verglichen und die absolute Differenz berechnet.

3.3.3 Mathematiktest

Bei den Aufgaben zum mathematischen Argumentieren handelt es sich einerseits um validierte freigegebene Aufgaben aus den Schweizer Bildungsstandards (HarmoS-Aufgaben) des Handlungsaspektes "Argumentieren und Begründen" und andererseits um adaptierte Eigenkonstruktionen aus dem Pilotprojekt (Smit & Birri, 2014). Zwei Beispiele finden sich im Anhang. Alle Items wurden in einem Pretest getestet. Die Lernenden lösten einen Test von 10 Items, für welchen sie maximal 45 Min. Zeit bekamen. Die erhaltenen Rohwerte (1-4) wurden anschliessend IRT-skaliert.

3.4 Analyseverfahren

Um Aussagen zu den Forschungsfragen zu generieren, wurden die vorliegenden Daten mithilfe verschiedener Analyseverfahren mit der Software SPSS ausgewertet. Unterschiede zwischen der Kontroll- und Interventionsgruppe bzw. von Lernenden mit niedrigen und hohen Mathematikleistungen bezüglich Selbstregulation, Häufigkeit von Selbstbeurteilung, Genauigkeit der Selbstbeurteilung und Nützlichkeit des Rubrics wurden mittels T-Test für unabhängige Stichproben geprüft. Unterschiedliche Entwicklungen wurden mittels Varianzanalysen mit Messwiederholung geprüft. Beim Vergleich von Lernenden mit niedrigen und hohen Mathematikleistungen handelt es sich um einen Extremgruppenvergleich (Döring & Bortz, 2016). Um die Extremgruppen zu definieren, wurden die Lernenden entsprechend der Mathematikleistung bei t₁ in fünf Stufen eingeteilt (in Anlehnung an die fünf Kompetenzstufen bei PISA). D.h. die Lernenden mit Leistungen innerhalb des 20. Perzentils wurden der Gruppe "niedrige Testleistung" und Kinder oberhalb des 80. Perzentils der Gruppe "hohe Testleistung" zugeordnet. Laut Döring und Bortz können Extremgruppenvergleiche explorativ interessant sein. Jedoch muss beachtet werden, dass es dabei zu einer Überschätzung von Effekten kommen kann.

4. Ergebnisse

Um die Fragestellungen zu beantworten, wurden die Variablen Selbstregulation, Selbstbeurteilung, Nützlichkeit des Rubrics, Mathematikleistung sowie Differenz von Selbst- und Fremdbeurteilung verwendet. Der Tabelle 2 ist u.a. zu entnehmen, dass die Mittelwerte von Selbstregulation und Selbstbeurteilung vom ersten zum zweiten Erhebungszeitpunkt ansteigen. Zudem lassen sich zu beiden Zeitpunkten mittelstarke bis schwache Zusammenhänge zwischen Selbstregulation, Selbstbeurteilung sowie Nützlichkeit des Rubrics erkennen (s. Tabelle 2).

TAB. 2. Deskriptive Statistik und Korrelationen

	Mittelwert	Standardabweichung	1	2	3	4	5	6
1. Selbstregulation Schüler t1	3.70	.76						
2. Selbstregulation Schüler t2	3.91	.81	.47**					
3. Selbstbeurteilung t1	2.83	.95	.31**	.18**				
4. Selbstbeurteilung t2	3.02	1.01	.20**	.34**	.31**			
5. Nützlichkeit des Rubrics t2	3.97	1.12	.24**	.38**	.15**	.21**		

	Mittelwert	Standardabweichung	1	2	3	4	5	6
6. Mathematikleistung t1	-.02	.84	.02	.08*	-.04	.08*	.07	
7. Differenz von Selbst- u. Fremdbeurteilung	1.02	.81	.09*	.01	.05	.02	-.04	-.08*

Anmerkungen: N = 762/ bei 5. Nützlichkeit n = 331; **p < 0.01. Sechsstufige Likert-Skala.

4.1 Effekte des Rubrics auf die Selbstregulation und die Selbstbeurteilung

Bezüglich der Fragestellung 1 a), welcher Effekt sich bei Lernenden durch den Einsatz eines Rubrics auf die Selbstregulation zeigt, konnte mithilfe eines T-Tests aufgezeigt werden, dass sich die Fähigkeit zur Selbstregulation der Kontrollgruppe (KG) ($M = 3.66$, $SD = 0.74$) und der Interventionsgruppe (IG) ($M = 3.37$, $SD = 0.79$) beim ersten Erhebungszeitpunkt nicht signifikant unterscheidet ($t(737) = -1.33$, $p = 0.181$). Beim zweiten Erhebungszeitpunkt ist der Unterschied signifikant (KG: $M = 3.83$, $SD = 0.87$, IG: $M = 3.99$, $SD = 0.75$) ($t(704) = -2.663$, $p = 0.008$). Eine Varianzanalyse mit Messwiederholung ergab jedoch, dass die signifikanten Unterschiede der Entwicklung der Selbstregulation in IG und KG von t1 zu t2 ($F(1) = 46.250$, $p = 0.000$, $\eta^2 = 0.022$) unabhängig von der Gruppenzugehörigkeit bzw. der Intervention mit dem Rubric ($F(1) = 1.993$, $p = 0.159$, $\eta^2 = 0.003$) erfolgt sind (s. Tabelle 3).

TAB. 3. Schülereinschätzungen zur Selbstregulationskompetenz – Mittelwerte und Standardabweichung sowie Ergebnisse Varianzanalyse mit Messwiederholung

	Selbstregulation t1		Selbstregulation t2		Zeit		Zeit*Gruppe	
	M	SD	M	SD	F	Partielles Eta-Quadrat (η^2)	F	Partielles Eta-Quadrat (η^2)
Kontrollgruppe	3.66	0.74	3.83	0.87	46.250**	0.064	1.993°	0.003
Interventionsgruppe	3.73	0.79	3.99	0.75				

Anmerkungen: n = 714; °p = n.s.

In Bezug auf die Fragestellung 1b), welcher Effekt bei Lernenden durch den Einsatz eines Rubrics auf die Häufigkeit von Selbstbeurteilung zu erkennen ist, zeigte das Ergebnis eines T-Tests, dass in der IG aus Sicht der Lernenden häufiger Selbstbeurteilung durchgeführt wurde. Die Mittelwerte der KG ($M = 2.87$, $SD = 0.96$) und der IG ($M = 2.78$, $SD = 0.92$) unterscheiden sich beim ersten Erhebungszeitpunkt nicht signifikant ($t(733) = 0.851$, $p = 0.395$). Beim zweiten Erhebungszeitpunkt ist der Unterschied jedoch signifikant (KG: $M = 2.76$, $SD = 0.90$; IG: $M = 3.29$, $SD = 1.04$) ($t(704) = -7.593$, $p = 0.000$). Eine Varianzanalyse mit Messwiederholung zeigte, dass

diese unterschiedliche Entwicklung in den Gruppen signifikant ist ($F(1) = 54.353$, $p = 0.000$, $\eta^2 = 0.074$) (s. Tabelle 4). Der Rubric scheint somit eine signifikante Wirkung auf den Einsatz von Selbstbeurteilung in der Wahrnehmung der Lernenden zu haben.

TAB. 4. Schülereinschätzungen zur Selbstbeurteilung – Mittelwerte und Standardabweichung sowie Ergebnisse Varianzanalyse mit Messwiederholung

	Selbstbeurteilung t1		Selbstbeurteilung t2		Zeit		Zeit * Gruppe	
	M	SD	M	SD	F	Partielles Eta-Quadrat (η^2)	F	Partielles Eta-Quadrat (η^2)
Kontrollgruppe	2.87	0.96	2.76	0.90	21.898**	0.031	54.353**	0.074
Interventionsgruppe	2.78	0.92	3.29	1.04				

Anmerkungen: $n = 714$; ** $p < 0.01$

Bezüglich Wirkung des Rubrics auf die Selbstbeurteilung (Frage 1b) ergab die Auswertung der Selbstbeurteilung der Einzelaufgabe mittels T-Tests, dass sich die Mittelwerte der Differenz von Selbst- und Fremdbeurteilung der KG ($M = 1.05$, $SD = 0.81$) und der IG ($M = 0.98$, $SD = 0.81$) nicht signifikant unterscheiden ($t(683) = 1.16$, $p = 0.25$). Die Lernenden der IG schätzten sich zwar präziser ein, sprich die Differenz zwischen Selbst- und Fremdeinschätzung ist geringer, jedoch ist das Ergebnis nicht signifikant (s. Tabelle 5). D. h. der Rubric scheint die Fähigkeit zur Selbstbeurteilung nicht zu fördern.

TAB. 5. Unterschiede Differenz zwischen Selbst- und Fremdbeurteilung bei der Einzelaufgabe – Mittelwerte und Standardabweichung sowie Ergebnisse des T-Tests für unabhängige Stichproben

	Selbstbeurteilung		Fremdbeurteilung		Differenz Selbst- und Fremdbeurteilung			
	M	SD	M	SD	M	SD	t	df
Kontrollgruppe	2.60	0.81	1.721	0.79	1.05	0.81	1.16°	683
Interventionsgruppe	2.65	0.76	1.80	0.82	0.98	0.81		

Anmerkungen: $n = 689$; ° $p = n.s.$

4.2 Nützlichkeit des Rubrics

Um die zweite Fragestellung zu beantworten, ob die Lernenden den Rubric als Hilfsmittel für das fachliche und überfachliche Lernen als nützlich einschätzen, wurden nur die Ergebnisse der Lernenden der IG berücksichtigt. Generell fanden

die Lernenden den Rubric nützlich ($M = 3.97$, $SD = 1.12$). Die Mehrheit stimmt folgenden Aussagen eher zu: Der Rubric helfe z. B. zu verstehen, worauf es beim Lösen einer Aufgabe ankommt, wie gut eine Aufgabe gelöst wurde, wie eine Aufgabe zu verbessern ist, wie eine Aufgabe besser zu lösen wäre usw.

4.3 Unterschiede bei Lernenden mit niedriger und hoher Mathematikleistung bezüglich Selbstregulation, Selbstbeurteilung und Nützlichkeit des Rubrics

4.3.1 Selbstregulation

In Bezug auf die dritte Fragestellung, ob es Unterschiede bei den Einschätzungen der Lernenden mit niedrigen und hohen Mathematikleistungen bezüglich der Selbstregulation gibt, ergab ein T-Test, dass sich die Mittelwerte der Gruppe mit hohen Mathematikleistungen ($M = 3.68$, $SD = 0.76$) und der Gruppe mit niedrigen Leistungen ($M = 3.73$, $SD = 0.84$) beim ersten Erhebungszeitpunkt nicht signifikant unterscheiden ($t(286) = 0.299$, $p = 0.765$). Beim zweiten Erhebungszeitpunkt ist der Unterschied zwischen der Gruppe mit hohen Mathematikleistungen ($M = 3.95$, $SD = 0.76$) und der Gruppe mit niedrigen Mathematikleistungen ($M = 3.85$, $SD = 0.85$) ebenfalls nicht signifikant ($t(271) = -1.275$, $p = 0.203$). Eine Varianzanalyse mit Messwiederholung ergab ebenfalls, dass die Unterschiede der Entwicklung der Selbstregulation in beiden Gruppen ($F(1) = 16.325$, $p = 0.000$, $\eta^2 = 0.059$) nicht auf die Gruppenzugehörigkeit ($F(1) = 2.127$, $p = 0.146$, $\eta^2 = 0.008$) zurückzuführen sind (s. Tabelle 6).

TAB. 6. Schülereinschätzungen zur Selbstregulationskompetenz – Mittelwerte und Standardabweichung sowie Ergebnisse Varianzanalyse mit Messwiederholung

	Selbst-regulation t1		Selbst-regulation t2		Zeit		Zeit * Gruppe	
	M	SD	M	SD	F	Partielles Eta-Quadrat (η^2)	F	Partielles Eta-Quadrat (η^2)
Niedrige Mathematikleistung t1	3.73	0.84	3.85	0.85	16.325**	0.059	2.127°	0.008
Hohe Mathematikleistung t1	3.68	0.76	3.95	0.76				

Anmerkungen: $n = 261$; ° $p = n.s.$

4.3.2 Selbstbeurteilung einer Einzelaufgabe

Bei der Betrachtung der gemittelten Differenz von Selbst- und Fremdbeurteilung bei der Einzelaufgabe zeigte das Ergebnis eines T-Tests, dass sich die Gruppen mit hohen Mathematikleistungen ($M = 0.918$, $SD = 0.829$) und jene mit niedrigen Mathematikleistungen ($M = 1.113$, $SD = 0.767$) unterschiedlich genau selbst beurteilen.

Die Lernenden mit hohen Mathematikleistungen schätzten sich jedoch nicht signifikant genauer ein ($t(267) = 1.995$, $p = 0.47$) (s. Tabelle 7).

TAB. 7. Unterschiede Differenz zwischen Selbst- und Fremdbeurteilung bei der Einzelaufgabe – Mittelwerte und Standardabweichung sowie Ergebnisse des T-Tests für unabhängige Stichproben

	Selbst-beurteilung		Fremd-beurteilung		Differenz Selbst- und Fremdbeurteilung			
	M	SD	M	SD	M	SD	t	df
Niedrige Mathematikleistung t1	2.43	0.74	1.37	0.59	1.11	0.77	2.0°	267
Hohe Mathematikleistung t1	2.77	0.77	2.18	0.92	0.92	0.83		

Anmerkungen: $n = 139$; ° $p = n.s.$

4.3.3 Nützlichkeit des Rubrics

Die Mittelwerte und Standardabweichungen der beiden Gruppen zeigen, dass sowohl Lernende mit niedrigen und hohen Mathematikleistungen den Rubric eher als nützlich einschätzen. Die Meinungen der einzelnen Schülerinnen und Schüler weichen jedoch insbesondere bei der schwächeren Gruppe stark voneinander ab (niedrige Leistungen: $M = 3.94$, $SD = 1.26$; hohe Leistungen: $M = 4.10$, $SD = 0.99$). Es zeigt sich, dass Lernende mit hohen Leistungen den Rubric tendenziell als nützlicher einschätzen. Ein Gruppenvergleich mittels T-Tests ergab jedoch, dass sich die zwei Gruppen nicht signifikant voneinander unterscheiden ($t(137) = -0.860$, $p = 0.391$) (s. Tabelle 8).

TAB. 8. Unterschiede Nützlichkeit des Rubrics

	Differenz Selbst- und Fremdbeurteilung			
	M	SD	t	df
Niedrige Mathematikleistung t1	3.94	1.26	-0.86°	137
Hohe Mathematikleistung t1	4.10	0.99		

Anmerkungen: $n = 139$; ° $p = n.s.$

5. Diskussion

5.1 Effekte des Rubrics auf die Selbstregulation und die Selbstbeurteilung

In Bezug auf die erste Fragestellung zeigte sich, dass der Einsatz des Rubrics als Hilfsmittel zwar einen tendenziell positiven, jedoch keinen signifikanten Effekt auf die Selbstregulation hat. Dies steht im Gegensatz zur Literatur-Übersicht von Panadero & Jonsson (2013) untersuchten 21 Studien zur Verwendung von Rubrics für die formative Beurteilung. Sie kamen zum Schluss, dass Rubrics u.a. die Selbstre-

gulation des Lernens unterstützen können, indem Lernende Feedback einfacher einschätzen und nutzen, das Aufgabenlösen besser planen sowie gelöste Aufgaben und Lernfortschritte leichter überprüfen können.

Was die Selbstbeurteilung betrifft, sind die Ergebnisse der Untersuchung unterschiedlich. Schülerinnen und Schüler der Interventionsgruppe geben einerseits an, im Unterricht häufiger Selbstbeurteilung zu betreiben als die Lernenden der Kontrollgruppe. Andererseits schnitten die Kinder der Interventionsgruppe in Bezug auf die Genauigkeit der Selbstbeurteilung einer Einzelaufgabe nicht signifikant besser ab. Unter Umständen war die Projektzeit zum Lernen mit dem Rubric für eine präzisere Selbstbeurteilung zu kurz. Damit der Einsatz von Rubrics Effekte zeigt, muss dieser einige Wochen intensiv eingesetzt werden (Panadero & Jonsson, 2013). Es könnte aber auch sein, dass die gewählte Selbstbeurteilungsaufgabe ungeeignet war, um die Fähigkeit der Selbstbeurteilung wirklich zu erfassen. So entspricht die Selbstbeurteilung der gesamten Aufgabe mit Smileys nicht einer Selbstbeurteilung mit einem analytischen Rubric. Laut Brown und Harris (2012) handelt es sich beim Setzen von Smileys eher um eine Form von Selbstbeurteilung, welche weniger differenziert ist als eine mit einem Rubric. Mehrere Items hätten hier eine zuverlässigere Einschätzung ergeben. Auch könnte angenommen werden, dass die Lernenden das Selbstbeurteilen von Aufgaben zu wenig trainiert haben respektive zu wenig Rückmeldungen der Lehrpersonen zur Passung der Einschätzung bekamen. Laut Goodrich (1996) kann Selbstbeurteilung im Unterricht nur effektiv implementiert werden, wenn Lernende Selbstbeurteilung auch wirklich lernen. Als weiterer Punkt kann angemerkt werden, dass Selbstregulation und Selbstbeurteilung nur funktionieren, wenn die Lernenden den Wert und die Sinnhaftigkeit der beiden Aktivitäten in Bezug auf ihr Lernen und den schulischen Output erkennen (Dignath, Buetter, & Langfelt, 2008). In Bezug auf den Wert und die Sinnhaftigkeit aus Sicht der Lernenden spielen zudem weitere Faktoren eine Rolle z.B. die eigene Selbstwirksamkeit und Motivation, die Erwartungen von Peers oder der Lehrperson usw. (Andrade, 2010).

5.2 Nützlichkeit des Rubrics

Bezüglich der zweiten Fragestellung, zur Nützlichkeit des Rubrics als Hilfsmittel für das fachliche und überfachliche Lernen, wurde aufgezeigt, dass Lernende den Rubric durchaus als nützlich beurteilen. Der Rubric hilft einzuschätzen, worauf es beim Lösen einer Aufgabe ankommt oder wie gut eine Aufgabe gelöst wurde. Der Rubric hilft zu verstehen, wie eine Aufgabe zu verbessern oder eine nächste Aufgabe optimaler zu lösen ist. Brown und Harris (2012) beschreiben fünf Aspekte zur Durchführung und Etablierung von Selbstbeurteilung im Unterricht, welche mit dem Einsatz eines Rubrics abgedeckt werden können: (1) gemeinsame Kriterien-Definition, (2) Verstehen der Kriterien (3) Feedback von Lehrperson und Peers,

(4) Verwendung von weiteren Testergebnissen für die Reflexion und (5) Selbstbeurteilung ohne Angst. Was die vorliegende Studie betrifft, kann angenommen werden, dass die Punkte 2, 3 und 5 aufgrund der Vorgaben in der Unterrichtsreihe umgesetzt worden sind. Ob Punkt 4 durchgeführt wurde, kann nicht beurteilt werden. Punkt 1 wurde aufgrund des vorgegebenen Rubrics mit klaren Kriterien eher indirekt umgesetzt. Die Lehrpersonen haben jeweils mit ihren Schülerinnen und Schülern besprochen, wie die Kriterien des Rubrics zu verstehen sind und wie sie angewendet werden. Damit Rubrics als hilfreiche Werkzeuge im Unterricht eingesetzt werden können, müssen sie sorgfältig und gemeinsam mit den Lernenden eingeführt werden (Panadero & Jonsson, 2013).

5.3 Unterschiede bei Lernenden mit niedrigen und hohen Leistungen im mathematischen Argumentieren

Was die dritte Fragestellung betrifft, lassen sich bezüglich Selbstregulation, Selbstbeurteilung und Nützlichkeit des Rubrics keine signifikanten Unterschiede zwischen Lernenden mit niedrigen und hohen Leistungen im mathematischen Argumentieren feststellen. Die Annahme, dass z.B. schwächere Lernende mehr Schwierigkeiten bei der Selbstbeurteilung haben, da sie sich eher überschätzen (Panadero, Brown, & Strijbos, 2016), konnte nicht bestätigt werden. Jedoch muss in der vorliegenden Untersuchung berücksichtigt werden, dass es sich um einen Extremgruppenvergleich handelt. Laut Döring und Bortz sind Ergebnisse von Extremgruppenvergleichen nicht hypothesenprüfend, sondern nur explorativ zu interpretieren.

5.4 Methodische Einschränkungen

Die vorliegende Auswertung weist einige methodische Limitierungen auf. Erstens beruhen die Ergebnisse hauptsächlich auf den Selbsteinschätzungen der Lernenden. Im Allgemeinen sind Schülereinschätzungen zu Unterrichtsmerkmalen aber durchaus vertrauenswürdig, wenn Hintergrundvariablen der Schülerinnen und Schüler mitberücksichtigt werden (Desimone, Smith, & Frisvold, 2010). Zweitens wurde nur eine Aufgabe eingesetzt, um die Selbstbeurteilungsfähigkeit zu prüfen, wodurch diese Ergebnisse auf einer singulären Grundlage beruhen. Drittens stellt sich die Frage, wie nachhaltig die Implementierung des Rubrics zur Förderung der Selbstregulation und Selbstbeurteilung ist. So beschreiben Panadero und Jonsson (2013), dass die Effekte von Rubrics klarer sind, wenn über einen längeren Zeitraum und während mehreren Perioden mit diesen gearbeitet wird.

6. Ausblick

Um reliablere Aussagen zu Rubrics als Instrument zur Förderung von Selbstregulation und Selbstbeurteilung machen zu können, welche nicht nur auf Selbsteinschätzungen der Lernenden beruhen, müssten weitere Daten beigezogen werden wie beispielsweise die Lehrpersonendaten (Smit, Bachmann, Blum, Birri, & Hess, 2017). Ausserdem könnten zusätzliche Daten erhoben werden, wie beispielsweise Daten aus Interviews (Triangulation) oder von externen Beobachtungen im Unterricht mittels z.B. einer Videoanalyse, wie dies in einer aktuellen Studie der Autorin und des Autors geplant ist. Um die Nachhaltigkeit der Implementierung des Rubrics zu überprüfen, müsste zudem ein dritter Messzeitpunkt berücksichtigt werden (Smit, Hess, Bachmann, Blum & Birri, 2018).

Weiterführend würde interessieren, wie genau sich die Lehrpersonen an die vorgegebene Unterrichtsplanung gehalten haben und wie sicher sie sich bei der Umsetzung gefühlt haben. Wie wirksam Lehrpersonen beispielsweise Selbstbeurteilung im Unterricht einsetzen, hängt davon ab, inwiefern sie neben einem durchgeführten Training in Selbstbeurteilung bereits positive Erfahrungen damit gesammelt haben und ob sie Vorteile für ihren Unterricht erkennen (Panadero, Jonsson, & Strijbos, 2016). Für die Lehrerbildung könnte das bedeuten, dass bereits gemeinsam mit Studierenden Rubrics entwickelt werden, welche ihnen dienen, ihre eigenen Berufs- und Studienkompetenzen zu beurteilen und ihr Lernen zu regulieren.

Literatur

- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–18.
- Andrade, H. (2007). Self-assessment through rubrics. *Educational Leadership*, 65(4), 60–63.
- Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. Andrade & G. Cizek (Ed.), *Handbook of formative assessment* (pp.90–105). New York: Routledge.
- Andrade, H. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, 4(87). doi: 10.3389/feduc.2019.00087
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10, 1–11.
- Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, 17(2), 199–214. doi: <http://dx.doi.org/10.1080/09695941003696172>
- Arter, J., & Chappuis, J. (2006). *Creating & recognizing quality rubrics*. Upper Saddle River, NJ: Pearson Education.

- Benischek, I., Jagric, C., Jungwirth, S., Marounek, R., & Niedertscheider, F. (2013). *Einstiege ins Argumentieren und Begründen in der Sekundarstufe I*. Wien: BIFIE.
- Bezold, A. (2009). *Förderung von Argumentationskompetenzen durch selbstdifferenzierende Lernangebote. Eine Studie im Mathematikunterricht der Grundschule*. Hamburg: Kovac.
- BIFIE (Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens) (Hrsg.). (2010). *Themenheft Mathematik „Kommunizieren“: Volksschule Grundstufe I + II (Bildungsstandards – Information für Lehrer/innen)*. Graz: Leykam.
- Boekaerts, M., & Corno, L. (2005). Self-Regulation in the Classroom: A Perspective on Assessment and Intervention. *Applied Psychology: An International Review*, 54(2), 199–231.
- Brouër, B. (2014). *Selbstbeurteilung in selbstorganisationsoffenen Lernumgebungen. Entwicklung eines Modells der Selbstbeurteilung und evidenzbasierter Impulse für die Förderung der Selbstbeurteilung in der Praxis*. Bad Heilbrunn: Klinkhardt.
- Brown, G., & Harris, L. (2012). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA: Sage.
- Brown, G., Harris, L., & Harnett, J. (2012). Teacher beliefs about feedback within an assessment for learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education*, 28(7), 968–978. doi:10.1016/j.tate.2012.05.003
- Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen. Grundlagen, Befunde und Konzepte*. Berlin, Heidelberg: Springer.
- Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 1–14. doi: <http://www.tandfonline.com/action/showCitFormats?doi=10.1080/02602938.2015.1111294>
- Desimone, L. M., Smith, T. M., & Frisvold, D. E. (2010). Survey measures of classroom instruction. *Educational Policy*, 24(2), 267–329.
- Deutschschweizer Erziehungsdirektorenkonferenz (D-EDK) (2016). *Lehrplan 21. Grundlagen. Von der D-EDK Plenarversammlung am 31.10.2014 zur Einführung in den Kantonen freigegebene Vorlage*. Bereinigte Fassung vom 29.02.2016
- Dignath, Ch., Buettner, G., & Langfeldt H.P. (2008). How can primary school students learn self-regulated learning strategies most effectively? *Educational Research Review*, 3, 101–129. doi: 10.1016/j.edurev.2008.02.003
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation*. (5. Aufl.). Berlin, Heidelberg, New York: Springer. doi: 10.1007/978-3-642-41089-5
- Eder, E., & Hofmann, F. (2012). Überfachliche Kompetenzen in der österreichischen Schule: Bestandsaufnahme, Implikationen, Entwicklungsperspektiven. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 71–109). Graz: Leykam.

- Fend, H. (2008). *Schule gestalten. Systemsteuerung, Schulentwicklung und Unterrichtsqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Fetzer, M. (2011). Wie argumentieren Grundschul Kinder im Mathematikunterricht? Eine argumentationstheoretische Perspektive. *Journal für Mathematik-Didaktik*, 32(1), 27-51. doi: 10.1007/s13138-010-0021-z
- Goodrich, H. (1996). *Student self-assessment: At the intersection of metacognition and authentic assessment*. (Unveröffentlichte Dissertation). Harvard University, Cambridge, MA.
- Hargreaves, E., McCallum, B., & Gipps, C. (2000). Teacher feedback strategies in primary classrooms - new evidence. In A. Susan (Ed.), *Feedback for Learning* (pp. 21–31). London: Routledge.
- Harris, L., & Brown, G. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101–111. doi: <http://dx.doi.org/10.1016/j.tate.2013.07.008>
- Hattie, J., & Timperley H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144.
- Kostons, D., van Gog, T., & Paas, F. (2011). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, 22(2), 121–132. doi: 10.1016/j.learninstruc.2011.08.004
- Kultusministerkonferenz (KMK) (2004). *Bildungsstandards im Fach Mathematik für den Primarbereich. Vereinbarung über Bildungsstandards für den Primarbereich (Jahrgangsstufe 4)*. Beschluss der Kultusministerkonferenz vom 15.10.2004. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_10_15-Bildungsstandards-Mathe-Primar.pdf
- Kultusministerkonferenz (KMK) (2015). *Empfehlungen zur Arbeit in der Grundschule*. Beschluss der Kultusministerkonferenz vom 02.07.1970 i. d. F. vom 11.06.2015. Verfügbar unter https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1970/1970_07_02_Empfehlungen_Grundschule.pdf
- Lohbeck, A., Petermann, F., & Petermann, U. (2015). Selbsteinschätzungen zum Sozial- und Lernverhalten von Grundschulkindern der vierten Jahrgangsstufe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47(1), 1–13. doi: 10.1026/0049-8637/a000118
- Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? *Zeitschrift für Erziehungswissenschaft*, 13, 293–308. doi: 10.1007/s11618-010-0124-9
- Maier, U. (2015). *Leistungsdiagnostik in Schule und Unterricht. Schülerleistungen messen, bewerten und fördern*. Bad Heilbrunn: Klinkhardt.

- Meusen-Beekman, K.D., Joosten-ten Brinke, D., & Boshuizen H.P.A. (2015). Developing young adolescents' self-regulation by means of formative assessment: A theoretical perspective. *Cogent Education*, 2, 1–16. doi: <http://dx.doi.org/10.1080/2331186X.2015.1071233>
- Meusen-Beekman, K.D., Joosten-ten Brinke, D., & Boshuizen H.P.A. (2016). Effects of formative assessments to develop self-regulation among sixth grade students: Results from a randomized controlled intervention. *Studies in Educational Evaluation*, 51, 126–136. doi: <http://dx.doi.org/10.1016/j.stueduc.2016.10.008>
- Nicol, D.J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. doi: <https://doi.org/10.1080/03075070600572090>
- Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, A., & Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology*, 25, 147–166. doi: <https://doi.org/10.1348/026151006X153127>
- Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) (2005). *Definition und Auswahl von Schlüsselkompetenzen. Zusammenfassung*. Verfügbar unter <https://www.oecd.org/pisa/35693281.pdf>
- 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*, 1–8. doi: <https://doi.org/10.1016/j.lindif.2012.04.007>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
- Panadero, E., Brown, G. T.L., & Strijbos, J.-W. (2016). The future of Student Self-Assessment: a Review of Known Unknowns and Potential Directions. *Educational Psychology Review*, 28, 803–830. doi: 10.1007/s10648-015-9350-2
- Panadero, E., Jonsson, A., & Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Ed.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311–326). (The enabling power of assessment; 4th ed.). Boston, MA: Springer.
- Perry, B., Dockett, S., & Harley, E. (2007). Learning stories and children's powerful mathematics. *Early childhood research and practice*, 9(2), 18–35.
- 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), 87–100.
- Reusser, K. (2011). Unterricht und Klassenführung. In L. Criblez, B. Müller & J. Oelkers (Hrsg.), *Die Volksschule – zwischen Innovationsdruck und Reformkritik* (S. 69–83). Zürich: Neue Zürcher Zeitung.

- Ross, J. A. (2006). The Reliability, Validity, and Utility of Self-Assessment. *Practical Assessment, Research & Evaluation*, 11(10), 1–13. doi: <http://hdl.handle.net/1807/30005>
- Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK) (2011). *Grundkompetenzen für die Mathematik. Nationale Bildungsstandards*. Freigegeben von der EDK-Plenarversammlung am 16. Juni 2011.
- Smit, R., & Birri, T. (2014). Assuring the quality of standards-oriented classroom assessment with rubrics for complex competencies. *Studies in Educational Evaluation*, 43, 5–13. doi: <https://doi.org/10.1016/j.stueduc.2014.02.002>
- 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(603). doi: <https://doi.org/10.1007/s11251-017-9416-2>
- Smit, R., Hess, K., Bachmann, P., Blum, V., & Birri, T. (2018). What Happens After the Intervention? Results From Teacher Professional Development in Employing Mathematical Reasoning Tasks and a Supporting Rubric. *Frontiers in Education*, 3(113). doi: 10.3389/educ.2018.00113
- Weiglhofer, H. (2013). *Die Kompetenzlandkarte für Unterrichtsprinzipien und Bildungsanliegen*. BMBWF. Verfügbar unter <https://www.bmb.gv.at/schulen/unterricht/uek/index.html>
- Winkelmann, H., & Van den Heuvel-Panhuizen, M. (2009). Die Erfassung mathematischen Argumentierens. In D. Granzer, O. Köller, & A. Bremerich-Vos (Hrsg.), *Bildungsstandards Deutsch und Mathematik* (S. 157–168). Weinheim: Beltz.
- Zimmerman (2002). Becoming a Self-Regulated Learner: An Overview. *Theory into Practice*, 41(3), 64–70.

Anhang

Beispielaufgaben

(1) Kinder sind etwa 50 cm gross, wenn sie zur Welt kommen. Bis zum Alter von 12 Jahren wachsen Kinder jedes Jahr etwa ... cm.

Kreuze die Antwort an, die am ehesten stimmt und begründe!

- 2 cm
- 4 cm
- 9 cm
- 14 cm

(2) Bilde aus den Ziffern 4, 5, 6, 7, 8, 9 zwei dreistellige Zahlen und addiere sie. Verwende jede Ziffer nur einmal. Beispiel: $465 + 798 = 1263$

Mit welchen Zahlen entsteht die grösste Summe? Begründe!