

# Textverständnis oder mathematisches Verständnis: Was macht Aufgaben der AHS-Zentralmatura Mathematik schwierig?

ANDREAS VOHNS, TAMARA OBEREDER, JANINE EGGER, STEFAN SCHEIBER  
(UNIVERSITÄT KLAGENFURT)

In öffentlichen Reaktionen auf die als unbefriedigend empfundenen Ergebnisse der zentralen schriftlichen Reifeprüfung in Mathematik im Jahr 2018 wurde verschiedentlich gemutmaßt, dass Schwierigkeiten der Geprüften mit den Aufgaben weniger in deren mathematischer Bewältigung als in der sprachlichen Verarbeitung der Aufgabentexte liegen könnten. Im folgenden Beitrag wird diskutiert, inwieweit eine Trennung sprachlichen und mathematischen Verständnisses von Aufgabentexten theoretisch und empirisch sinnvoll möglich ist, und es wird über ausgewählte Ergebnisse eines Analyseprojekts zu sprachlichen und konzeptuellen Hürden von Aufgabenstellungen der AHS-Zentralmatura berichtet.

## 1. Einführung

Die seit dem Schuljahr 2014/5 flächendeckend an den Allgemeinbildenden Höheren Schulen (AHS) durchgeführten zentralen schriftlichen Reifeprüfungen sind, soweit wir dies überblicken können, in der Breite recht beständig von den betroffenen Lehrpersonen, Prüflingen und deren Eltern ebenso wie von Hochschullehrenden aus dem MINT-Bereich verhalten bis skeptisch aufgenommen worden. Auf den Tag genau am 26.06.2018 wurde die AHS-Zentralmatura Mathematik dann auch von ministerieller Seite öffentlich auf den Prüfstand gestellt. Der Grund: „An den BHS ist es das bisher schlechteste Ergebnis, an den AHS hat es rund doppelt so viele Fünfer gegeben wie im Vorjahr.“ (Rieger & APA, 2018). Während die Deutsch-Matura Jahr für Jahr verlässlich sozialverträgliche Bestehensquoten jenseits der 90 % vor bzw. z. T. sogar jenseits der 99 % nach Kompensationsprüfungen liefert, erzeugen die deutlich größeren Schwankungen und höheren absoluten Durchfallsquoten in Mathematik, insbesondere im AHS-Bereich (2015: 9.7 %, 2016: 23.2 %, 2017: 11.8 %, 2018: 22.5 %; jeweils vor Kompensationprüfung), medial ein großes Echo und machen dem Anschein nach auch dem Ministerium gewisse Sorgen. Man kommt jedenfalls zu dem Schluss, bei der Mathematik-Matura müsse nun endlich „eine ausgiebige Reflexion über die bisherigen Prüfungen“ (Rieger & APA, 2018) stattfinden. Inhaltlich werde man sich dabei etwa auf den Katalog der Grundkompetenzen konzentrieren. „Sind alle davon noch notwendig? Oder fehlen welche, die eigentlich viel sinnvoller wären?“, fragt Minister Faßmann (a. a. O.). Außerdem wird die Textlastigkeit der Mathematik-Aufgaben kritisch betrachtet. „Es wurde empirisch erhoben, dass lange Texte eine Schwierigkeit darstellen“, sagt Faßmann (a. a. O.). Tatsächlich hatten schon vor Verkündung der Ergebnisse Vertreterinnen und Vertreter von Eltern und Lernenden in dieselbe Kerbe geschlagen: „Die Mathematura darf keine zweite Deutschklausur sein“ (Egyed & Mittelstaedt, 2018), sagt Bundeschulsprecher Harald Zierfuß am 28.05.2018 im Interview dem Standard. Ähnlich sah das der oberste Elternvertreter Gernot Schreyer: „Es ist falsch, Mathematik zu verwenden, um Deutschkompetenzen abzufragen“ (a. a. O.). Ganz ähnlich hörte es sich an, wenn Stadtschulratspräsident a. D. Kurt Scholz am 15.11.2018 die Ergebnisse seiner aufgrund der ministeriellen Sorgen einberufenen „Zuhörtour“ in den Bundesländern präsentierte. Die Mathematikzentralmatura solle 2019 „nicht leichter, aber verständlicher werden“ (APA, 2018). Die Schulpartner waren sich in ihren Forderungen laut Scholz erstaunlich einig. Eine Hauptforderung mit Blick auf die AHS-Aufgaben: die Fragestellungen müssten verständlicher werden. Viele Schüler seien laut Angaben der Schulpartner mathematisch in der Lage, die Aufgaben zu lösen, würden jedoch an zu komplizierten Einleitungstexten scheitern.

Inwieweit eine solche Einschätzung sich mit den konstant hohen Bestehensquoten bei der Deutschmatura verträgt, sei dahingestellt. Schon dadurch, dass die besondere Aufmerksamkeit, die man 2019 erklärmaßen auf die Textverständlichkeit der Aufgaben gelegt haben will, mit einigen anderen Maßnahmen verbunden wurde (u. a. Fallenlassen bestimmter Aufgabentypen, kein Absammeln des ersten Teils mehr nach der Hälfte der Prüfungszeit, „zwei Wege zum Genügend“), lässt sich aus den höheren Bestehensquoten 2019 kaum retrospektiv argumentieren, dass die Maßnahmen zur Erhöhung der Textverständlich-

keit für diese Entwicklung ausschlaggebend waren – es lässt sich nicht klären, ob die Prüfung nun bloß verständlicher oder doch auch leichter war.

Tatsächlich ist die Forschungslage zur Bedeutung verschiedener Faktoren des Sprachverständnisses bzw. der sprachlichen Zugänglichkeit von Aufgabentexten für die erfolgreiche Bearbeitung von mathematischen Aufgabenstellungen in der deutschsprachigen Mathematikdidaktik (vgl. Abschnitt 2.2) komplexer als „je kürzer desto besser“, in Teilen ambivalent/widersprüchlich und zumindest mit Blick auf die hier in Rede stehende Altersgruppe auch sehr schmal; für Österreich liegen unseres Wissens bislang überhaupt keine gesicherten (empirischen) Forschungsbefunde vor. Schließlich ist zu berücksichtigen, dass die Frage der sprachlichen Komplexität von Aufgabentexten nicht unabhängig von der Frage ist, welche mathematischen Anforderungen mit den Aufgaben verbunden sind, oder anders: welches *mathematische Verständnis* eine Prüfung erheben soll, dürfte einen Einfluss darauf haben, wie Prüfungsaufgaben *sprachlich* gestaltet werden – insbesondere dort, wo Allgemeinbildungsvorstellungen verfolgt werden, die sich erkennbar an Konzeptionen von “mathematical literacy” anlehnen, also *intentional* und keineswegs unbeabsichtigt eine Lese- und Interpretationsfähigkeit mathemathikhaltiger Texte als erklärtes Bildungsziel setzen, was für die AHS-Zentralmatura und den Lehrplan, auf dem diese basiert, wohl recht eindeutig bejaht werden kann.

Der hier vorliegende Beitrag berichtet über ausgewählte Ergebnisse eines im November 2018 initiierten Analyseprojekts „Textverständnis bei der zentralen schriftlichen Reifeprüfung aus Mathematik AHS (& BHS) – sprachliche Schwierigkeiten versus mathematische Fähigkeitskomponenten“ des Instituts für Didaktik der Mathematik an der Universität Klagenfurt, dem seitens des Bildungsministeriums im Rahmen einer Kooperationsvereinbarung Zugang zu bestimmten Daten gewährt wurde und welches anhand der vorliegenden Daten zu den AHS-Maturajahrgängen 2015 bis 2018 erste Erkenntnisse zu möglichen Zusammenhängen zwischen mathematischen und sprachlichen Anforderungen der Prüfungsaufgaben des für das Bestehen oder Nicht-Bestehen besonders relevanten 1. Teils der AHS-Zentralmatura erbringen sollte<sup>1</sup>. Das Projekt versteht sich als ein erster Beitrag zur näheren wissenschaftlichen Aufklärung, es geht uns also vor allem darum, die medial populäre These der Überlagerung (zulässiger) mathematischer durch (unzulässige) sprachliche Anforderung einer ersten kritischen Prüfung zu unterziehen, welche sowohl konzeptionell-theoretische als auch explorativ-empirische Argumente vorbringt. Alle im Rahmen des Artikels getroffenen Aussagen spiegeln dabei ausschließlich die Einschätzungen der Autorinnen und Autoren des Beitrags wider und sind nicht als offizielle oder inoffizielle Positionierungen des Ministeriums oder dessen Mitarbeiterinnen und Mitarbeiter zu verstehen.

## **2. Mathematisches und sprachliches Verständnis – Desiderate fachdidaktischer und fachbezogen-sprachwissenschaftlicher Forschung**

### **2.1. Mathematisches Verstehen/Verständnis und mathematische Bildungsziele**

Im Allgemeinen wird die Forderung, die Lernenden sollten im Mathematikunterricht auch etwas verstehen, als eher wenig strittig gelten, jedenfalls scheint deren Umkehrung (die Lernenden sollen nichts verstehen) schwerlich argumentierbar. An der Konkretisierung dieser Forderung werden sich dann allerdings die Geister scheiden, denn es ist alles andere als eindeutig, wie man einen üblicherweise als inneren kognitiv aufgefassten Prozess wie das Verstehen in beobachtbare und somit prüfbare Fähigkeiten und Fertigkeiten der Lernenden zu operationalisieren hat. Verstehen ist nur evaluierbar auf Basis von *Verständnisprodukten*, die ganz unterschiedlicher Natur sein können. Modelle des Verstehens unterscheiden daher meist verschiedene Stufen oder Ebenen von Verstehen (die nicht immer hierarchisch gedacht sein müssen).

<sup>1</sup> Angesichts der deutlich sprachintensiveren Gestaltung des zweiten Teils der Prüfung mag die Entscheidung überraschen, den ersten Teil zu fokussieren. Das Prüfungskonzept der AHS-Zentralmatura hat jedenfalls im hier betrachteten Zeitraum das Bestehen der Prüfung allerdings weitgehend vom ersten Teil abhängig gemacht. Es hätte sich angeboten, in der Analyse zumindest noch die Aufgaben aus dem zweiten Teil der Prüfung zu berücksichtigen, die die Kompensationspunkte für den ersten Teil liefern. Bei diesen Beispielen ergeben sich aber dann hinsichtlich des für die Aufgabenbearbeitung tatsächlich benötigten Teils des Aufgabentextes deutliche Abgrenzungsprobleme, so dass auch darauf letztlich verzichtet wurde.

Eine dichotome Unterteilung in *syntaktisches* und *semantisches Verstehen* (verwendet etwa bei Prediger und Wittmann, 2009) geht auf Arbeiten von Skemp (1976) bzw. Mellin-Olsen zurück, bei denen die entsprechenden Ebenen „instrumental understanding“ (syntaktisches Verstehen) und „relational understanding“ (semantisches Verstehen) heißen. Semantisches Verstehen bzw. „relational understanding“ bezeichnet dabei das dem Alltagssprachgebrauch vermutlich näher stehende „knowing both what to do and why“ (Skemp, 1976, S. 20), während das syntaktische Verstehen bzw. „instrumental understanding“ sich auf „rules without reason“ (a. a. O.) beschränkt. Blickt man auf Aufgabenstellungen in Prüfungen, so ist die – seit Skemp und vermutlich noch weitaus länger – Kritik nahezu omnipräsent, dass diese aus verschiedenen Gründen zu stark syntaktisches Verständnis fokussieren und dies im Zweifelsfall dann auch dazu führt, dass im Unterricht zu wenig Aufmerksamkeit auf semantisches Verstehen gelegt wird. So heißt es etwa im in Anschluss an die TIMS-Studie erstellten Gutachten zur Vorbereitung des SINUS-Programms (einem deutschen Pendant zum IMST) der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung:

Im Allgemeinen harmonieren die üblichen Leistungsüberprüfungen durch Klassen- (Schul-) Arbeiten gut mit dem erarbeiteten Stoff. Die Aufgaben überprüfen das neuerworbene, geübte und routinisierte Wissen. Versucht man, Lernen und Leisten im Unterricht stärker zu trennen, um verständnisvolles Lernen, in dem Fehler ein eigenes Recht haben, zu intensivieren, wird zu prüfen sein, wie das Spektrum der Prüfungsaufgaben zur Unterstützung des gewünschten Lernens verbreitert werden kann. Wahrscheinlich müssen neben den Routineaufgaben, die nicht geringgeschätzt werden dürfen, systematisch Aufgaben berücksichtigt werden, die das Verständnis von Konzepten und Verfahren durch die Kombination von neuem und zurückliegendem Stoff und die Flexibilität des Verständnisses durch die Anwendung des Gelernten auf neue Situationen erfassen (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, 1997, S. 23).

Diese Empfehlung rührt ganz wesentlich aus der aus TIMSS gezogene(n) (und nach PISA wiederholt gestellten) Diagnose her, deutsche (ebenso wie österreichische) Schülerinnen und Schüler hätten ihre relativen Schwächen „bei realitätsorientierten Aufgaben, die einen eigenständigen Mathematisierungsprozess verlangten, und bei begriffsorientierten Aufgaben, die das selbstständige Herstellen von außer- und innermathematischen Zusammenhängen voraussetzten“ (Neubrand et al., 2004, S. 32), relativen Stärken seien „freilich auf insgesamt niedrigem Niveau, [...] hingegen bei Routineaufgaben und bei reinen Berechnungsaufgaben“ (a. a. O.) zu erkennen. Diese Diagnosen haben dann in Deutschland im Rahmen des Programms SINUS und in Österreich im Rahmen des Programms IMST auch jeweils zu Schwerpunkten im Bereich der Aufgaben- und Prüfungskultur geführt, die dem Zitat entsprechende Veränderungen unterstützen sollten. Es ist aus unserer Sicht durchaus zulässig davon auszugehen, dass dem Unterfangen „AHS-Zentralmatura Mathematik“ seitens der fachdidaktisch Befürwortenden die Philosophie zu Grunde liegt, sich auch in der Prüfungssituation verstärkt auf „semantisches Verständnis“ einzulassen – sehr deutlich etwa bei Peschek (2011), wenn dieser an der traditionellen, dezentralen Matura mit Wagenschein kritisiert, Maturavorbereitung sei vielfach schlicht „Dressur des Unverstandenen“ (a. a. O.).

In Maier (1995) und Maier und Steinbring (1998) haben die Autoren basierend u. a. auf Arbeiten zu Wissensstrukturen von Chi (1987) ein zweidimensionales Modell zur Beschreibung von mathematischen Verständnisprodukten entwickelt, welches diese hinsichtlich einer modalen und einer mentalen Dimension näher aufgliedert. Von Interesse scheint hierbei insbesondere die mentale Dimension, bei der die folgenden Komponenten mathematischer Verständnisprodukte unterschieden werden:

- *konzeptuelle Komponente*: erklären, erläutern oder definieren mathematischer Begriffe;
- *prozedurale Komponente*: Beschreibung und Ausführung mathematischer Verfahren/Algorithmen;
- *argumentative Komponente*: ableiten, begründen, beweisen mathematischer Aussagen/Schlüsse,
- *relationale Komponente*: Beziehungen zwischen mathematischen Begriffen;
- *elaborative Komponente*: lösen (außer-)mathematischer Probleme, Lösungspläne konzipieren;

- *reflexive Komponente*: metakognitive Aktivitäten, auch: Einschätzungen, Beurteilungen.

Diese Unterteilung weist fraglos Bezüge zur *Handlungsdimension* der Kompetenzmodelle auf, wie sie etwa den Standards M8 und M12 zu Grunde gelegt wurden und in abgewandelter Form als „O-M-A-Modell“ auch bei der Zusammenstellung der Aufgaben für die Zentralmatura ergänzend zum Grundkompetenz-Katalog Anwendung finden (vgl. Siller et al., 2016). Hier erscheint zunächst einmal wesentlich, dass die Diskussion um diese Komponenten in der Mathematikdidaktik vor die große Welle der „Kompetenzorientierung“ zurück reicht und (zunächst) auch ganz unabhängig von (zentralen) Prüfungen zur Beschreibung von kognitiven Prozessen beim Betreiben von Mathematik auf Schulniveau Verwendung gefunden haben – eben als ein möglicher Vorschlag zur Konkretisierung dessen, was als handfestere Beschreibung des unscharfen Konstrukts „mathematisches Verständnis“ aufgefasst werden kann.

Es ist weiters kaum von der Hand zu weisen, dass sich mit der AHS-Zentralmatura verbundene Intentionen zur Weiterentwicklung der Aufgabekultur in Prüfungen (und entsprechende Rückwirkungen auf die Aufgabekultur im Mathematikunterricht selbst) auch auf eine veränderte Berücksichtigung eben dieser Komponenten mathematischer Verständnisprodukte beziehen. Zentral erscheinen hier mit Blick auf die spezifische bildungstheoretische Rahmung, und zwar relativ konstant über die Anfänge im Klagenfurter Pilotprojekt bis hin zur Umsetzung in der heute noch gültigen Rahmenvorgabe, drei Aspekte (vgl. Peschek, 2011, bmbwf, 2019):

1. Zurückdrängen von rein *syntaktischem Verständnis* („Dressur des Unverstandenen“) zu Gunsten von *semantischem Verständnis*,
2. weniger Fokus auf operatives Wissen / *prozedurale Verständnisebene*, mehr Fokus auf Grundwissen (*konzeptuelle Verständnisebene*) und Reflexion (*reflexive Verständnisebene*),
3. Betonung *elaborativer Verständniselemente*, im Sinne einer flexiblen Anwendung mathematischen Wissens in unterschiedlichen, auch weniger vertrauten außermathematischen Kontexten, die als für alle Heranwachsenden unabhängig von weiteren Studien- und Berufszielen als material qualifizierend angenommen werden (also eine bestimmte Spielart von dem, was man international gemeinhin als “mathematical literacy” bezeichnen würde), gegenüber einem im deutschsprachigen Raum traditionell eher anti-utilitaristisch aufgestelltem, stark innermathematisch orientiertem Fachunterricht, welcher allenfalls noch ausgewählte Standardanwendungen in den Naturwissenschaften im Sinne einer Studierfähigkeit für den MINT-Bereich im Blick hatte.

Dass die Frage, welche Verständnisebenen bzw. -komponenten eine zentrale Prüfung ansprechen sollte, eng mit den Bildungszielen des Mathematikunterrichts zusammenhängt, ist dabei insofern von Bedeutung, als die unterschiedlichen Komponenten mathematischen Verständnisses eben auch in unterschiedlicher Weise mit sprachlichem Verständnis und sprachlicher Ausdrucksfähigkeit verknüpft sind. Während es durchaus möglich ist, isolierte prozedurale Fähigkeiten relativ „spracharm“ appellativ (Leite ab! Löse nach . . . auf! Berechne!), bisweilen sogar völlig „sprachlos“ zu prüfen, sind nahezu alle weiteren Komponenten mathematischen Verständnisses deutlich stärker auf Sprache als Medium der Aufgabenformulierung oder der Präsentation von Ergebnissen angewiesen. Wo es nicht ausreicht, Symbolfolgen mit Rechenappellen zu verbinden, sind in einer schriftlichen Prüfung etwa Beschreibungen der Situationen erforderlich, in denen Begriffe und Verfahren eingesetzt werden, es sind Erklärungen und Begründungen zu verfassen bzw. in geschlossenen Fragetypen zu rezipieren und auf Korrektheit zu überprüfen oder Einschätzungen zur Angemessenheit von Interpretationen zu treffen. Das stillschweigende Übereinkommen beim schriftlichen Prüfen *in allen Fächern* (mit Ausnahme des Sprachunterrichts in der Unterrichtssprache) ist faktisch, dass die Geprüften über soweit ausreichende Fähigkeiten in der Unterrichtssprache verfügen sollten, dass die Sprachlichkeit der Prüfung als solche die Lösung nicht verhindern sollte. Wo man an diesem stillschweigenden Übereinkommen Zweifel hegt, wird es kompliziert: Es stellt sich nämlich die Frage, inwieweit sich sprachliche Anforderungen einer Prüfungsarbeit so variieren lassen, dass keine wesentlichen Rückwirkungen auf das durch die Prüfung erfasste mathematische Verständnis die Folge sind.

## 2.2. Beziehungen zwischen sprachlichem und mathematischem Verständnis

Obwohl das, was man gemeinhin als „Textaufgaben“ bezeichnet, vermutlich in etwa genau so alt ist, wie (organisiertes) mathematisches Lehren und Lernen, stellt die systematische Untersuchung der Bedeutung sprachlicher Aspekte für das mathematische Lernen ein vergleichsweise junges Forschungsfeld dar. International wird als Initialzündung hier meist auf das UNESCO Symposium „Interactions between linguistics and mathematical education“ von 1974 verwiesen, im deutschsprachigen Raum hat die Diskussion wohl erst so richtig mit der Veröffentlichung der Ergebnisse der PISA-Studien an Fahrt aufgenommen. Einen sehr guten Überblick über den aktuellen Stand der Forschung liefert hier Schilcher, Röhl und Krauss (2017), einige Forschungsdesiderate seien im Folgenden kurz angeführt.

Es ist zunächst einmal wichtig festzuhalten, dass es generell nicht *das eine* Text-/Lese-/Sprachverständnis gibt, sondern unterschiedliche Fähigkeiten, die sich auf Sprachrezeption und Sprachproduktion beziehen können und die verschiedene *Sprachregister* (insbes. Alltags-, Bildungs-, Unterrichts- und Fachsprache) und *Darstellungsebenen* (verbalsprachlich, symbolsprachlich, grafisch, etc.) betreffen. Wenn im Umfeld der Zentralmatura von „sprachlichen Schwierigkeiten“ die Rede ist, so darf man wohl davon ausgehen, dass dabei vor allem an die *Verbalsprache* gedacht ist, Sprachwissenschaftlerinnen und -wissenschaftler beziehen in ihre Überlegungen sehr wohl auch nicht-verbalsprachliche Darstellungsebenen ein, etwa (die im Mathematikunterricht weit verbreitete) symbol- und formalsprachliche Darstellungsebene, Bildsprache und bisweilen auch non-verbale Formen (Handlungen, Gestik, Mimik; vgl. Leisen, 2011). Die Unterrichtssprache im Mathematikunterricht bedient sich dann auch im verbalsprachlichen Bereich verschiedener Sprachregister, neben der Alltags- bzw. Umgangssprache und der mathematischen Fachsprache ist in den letzten Jahrzehnten zunehmend dem *bildungssprachlichen Register* Aufmerksamkeit gewidmet worden. Bildungssprache als die im Bildungssystem typischerweise verwendete Sprache weist gegenüber der Alltagssprache einige markante Unterschiede hinsichtlich Wortschatz, grammatikalischer und stilistischer Eigenschaften auf, sie integriert insbesondere fachsprachliche Elemente verschiedener Wissenschaften, der Berufswelt und des gesellschaftlichen Bereichs jenseits des rein privaten Alltags. Während es weitgehend unstrittig ist, dass einerseits die Beherrschung des alltagssprachlichen Registers der Verkehrssprache des Unterrichts eine notwendige *Voraussetzung* fachlichen Lernens ist und andererseits auch wenig Streit darüber besteht, dass der Erwerb der mathematischen Fachsprache ein wichtiger *Lerninhalt* des Mathematikunterrichts darstellt, so kann im Umgang mit der nicht-mathematischen Bildungssprache die zentrale *Herausforderung* für einen sich gegenüber Zielen von “mathematical literacy” zunehmend öffnenden Fachunterricht gesehen werden, insbesondere dann, wenn man zunehmende Heterogenität oder jedenfalls Sensibilität gegenüber der Heterogenität der bildungssprachlichen Ressourcen der Lernenden in Rechnung stellt. Akzeptiert man, dass Schule für viele Lernende den zentralen Ort des Erwerbs von Bildungssprache darstellt, so kann sich auch Fachunterricht kaum der Verantwortung entziehen, am Erwerb ebendieser mitzuwirken und es erschiene wenig angemessen, Bildungssprache künstlich aus dem Fachunterricht herauszuhalten. Gleichwohl bleibt die Frage bestehen, inwieweit insbesondere in Prüfungsaufgaben heterogenen bildungssprachlichen Ressourcen der Lernenden durch einen bewusst moderaten Rückgriff auf dieses Sprachregister Rechnung getragen werden sollte.

Jüngere mathematikdidaktische Untersuchungen zu Zusammenhängen zwischen sprachlichen und mathematischen Fähigkeiten der Lernenden bzw. sprachlichen und mathematischen Anforderungen von Mathematikaufgaben liefern, wie in der Einleitung erwähnt, im deutschsprachigen Raum ein uneinheitliches Bild. Sowohl in den Auswertungen der PISA-Studie als auch in einigen in der Folge durchgeführten Längsschnittstudien konnten immer wieder relativ starke statistische Zusammenhänge zwischen sprachlichen und mathematischen Kompetenzen der Lernenden nachgewiesen werden, insbesondere wirkt Unterricht auch hinsichtlich der sprachlichen Kompetenzen keineswegs durchgängig kompensatorisch – entsprechende Disparitäten akkumulieren sich über die Schulzeit (vgl. Gürsoy, Benholz, Renk, Prediger und Büchter, 2013; Prediger, Wilhelm, Büchter, Gürsoy und Benholz, 2015; Ufer, Reiss und Mehlinger, 2013; Wilhelm, 2016). Umfangreichere Aufgabenbearbeitungsstudien im Umfeld der zentralen Prüfungen am Ende der 10. Schulstufe in Nordrhein-Westfalen belegen etwa, dass sprachliche Hürden für die erfolgreiche Bearbeitung mathematischer Prüfungsaufgaben auf unterschiedlichen Ebenen des

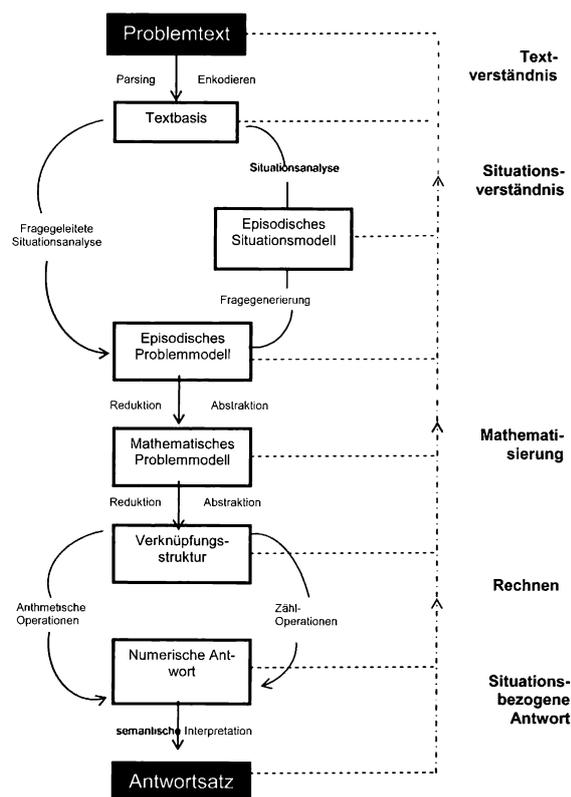


Abbildung 1: Stufen im Mathematisierungsprozess (Reusser, 1997)

Textverständnisses liegen (Wort, Satz, Text) bzw. sich auf unterschiedliche Eigenschaften der Texte (neben deren reiner Länge) beziehen können (vgl. Gürsoy et al., 2013).

Es ist allerdings einzuräumen, dass die Stärke des Zusammenhangs mathematischer Kompetenzen und sprachlicher Kompetenzen sich dort besonders deutlich zeigt, wo beide Konstrukte "literacy"-basierten Konzeptionen folgen (vgl. Schukajlow und Leiß, 2008; Ufer et al., 2013). Wenn, wie in PISA, etwa die Entnahme von Informationen aus nicht-kontinuierlichen Texten und Grafiken ausdrücklich den sprachlichen Kompetenzen zugewiesen wird, so entsteht ein Graubereich, in dem bestimmte Anforderungen gleichermaßen dem mathematischen wie auch dem sprachlichen Verständnis zugeordnet werden können. Im Rahmen des DISUM-Projektes konnten Schukajlow und Leiß (2008) andererseits auch nachweisen, dass Korrelationen zwischen gemessenen mathematischen und sprachlichen Fähigkeiten sehr gering ausfallen, wenn beide Konstrukte klar voneinander abgegrenzten Operationalisierungen folgen. Dass im Bereich der Analyse der Zusammenhänge zwischen sprachlichen Herausforderungen mathematischer Aufgaben und deren erfolgreicher Bearbeitung noch einiger Forschungsbedarf besteht, zeigt sich schließlich vor allem darin, dass man zwar auf quantitativer Ebene Zusammenhänge zwischen beiden Kompetenzbereichen nachweisen kann und auf qualitativer Ebene auch wiederholt das Scheitern von Bearbeitungsprozessen mit konkreten sprachlichen Eigenschaften von Aufgabenstellungen in Beziehung setzen konnte (etwa bei Wilhelm, 2016), gleichzeitig aber bislang wenig Erfolg dabei hatte, potentielle schwierigkeitsgenerierende, klar auf sprachlicher Ebene verorteter Merkmale von Aufgaben so gut zu isolieren, dass eine gezielte Variation dieser Eigenschaften nachweislich positiven Einfluss auf die erfolgreiche Aufgabebearbeitung hat (vgl. Leiß, Domenech, Ehmke und Schwippert, 2017; Plath, 2019).

Für die Frage der Trennbarkeit vorwiegend verbalsprachlich-linguistischer Herausforderungen gegenüber solchen, die man einer (durchaus intendierten) Lese- und Interpretationsfähigkeit mathemathikhaltiger Situationsbeschreibungen zuordnen würde, erscheint ein Bezug auf das auch modifiziert von Wilhelm (2016, S. 297) verwendete Modell von Reusser (1997) aufschlussreich (vgl. Abb. 1). Betrachtet man etwa die eingangs geschilderte Behauptung, viele Lernende seien im Prinzip in der Lage die Aufgaben zu

lösen, scheiterten aber an den Texten, so kann dieses Scheitern insbesondere bei in außermathematischen Kontexten gestellten Aufgaben auf wenigstens drei Ebenen des Mathematisierungsprozesses liegen:

1. Den Lernenden gelingt es bereits nicht, dem Aufgabentext die für die Bildung eines adäquaten *episodischen Situationsmodells* nötigen Informationen zu entnehmen – ein solches Scheitern würde man vermutlich eher auf der Ebene des rein sprachlichen Textverständnisses verorten.
2. Die Lernenden können dem Text zwar die entsprechenden Informationen entnehmen und die außermathematische Situation soweit korrekt erfassen, scheitern dann aber an den nötigen Vereinfachungen und Strukturierungen zur erfolgreichen Bildung eines mathematisierbaren *episodischen Problemmodells* oder an dessen Übertragung in ein adäquates *mathematisches Problemmodell* – ein solches Scheitern würde man fachdidaktisch eher einem Mangel an *elaborativem Wissen* zuordnen wollen.
3. Bei fehlendem *konzeptuellen Wissen* kann die Bearbeitung einer Aufgabe schließlich noch an der Reduktion/Abstraktion zur passenden *Verknüpfungsstruktur* scheitern, hier wäre dann aber bereits sehr klar die Ebene des sprachlichen Verständnisses verlassen.

In allen drei Fällen wäre es zudem durchaus möglich, dass an einer der drei Anforderungen scheiternde Lernende *prinzipiell* das (isolierte, innermathematische) Arbeiten mit den relevanten mathematischen Modellen durchaus beherrschen. Während man in mathematischen Prüfungen berechtigterweise versuchen würde, ein Scheitern im Sinne von 1.) durch möglichst prägnante und zugängliche Formulierungen zu verhindern, lassen sich die mit 2.) und 3.) verbundenen Schwierigkeiten hingegen nicht mindern, ohne gleichzeitig den mathematischen Anspruch der Aufgabenstellungen qualitativ zu verändern: Es ist eine andere Anforderung, die Ableitung einer Funktion zu bestimmen, als zu wissen, dass für eine außermathematische Situation die Lösung in der Bestimmung der Ableitung einer Funktion besteht. Beide Anforderungen können berechtigterweise als „mathematische Kompetenzen“ interpretiert werden. Es ist fraglich, ob der öffentlich geäußerten Kritik an mangelnder Verständlichkeit der Mathematik-Maturaaufgaben solche Differenzierungen bewusst sind bzw. falls sie den Kritikerinnen und Kritikern bewusst sind, inwiefern ihnen besonders viel an Differenzierungen liegt. Verbirgt sich hinter der Kritik an der sprachlichen Gestaltung nicht teilweise eher ein diffuses Unbehagen mit der *Art des mathematischen Verständnisses*, auf welches die AHS-Zentralmatura mit Blick auf die zugrunde liegende, stark an Aspekten von „mathematical literacy“ orientierte Allgemeinbildungsvorstellung rechtfertigt, bei welcher man nun auch nicht zwingend davon ausgehen kann, dass diese von Eltern, Lehrpersonen und MINT-Hochschullehrenden breit akzeptiert werden?

### 3. Studienanlage und Methodologie

Gegenstand der vorliegenden Studie sind zuvorderst die insgesamt 96 Aufgabenstellungen der AHS-Maturajahrgänge 2015-2018 (Haupttermine), für die dem IDM im Rahmen eines Kooperationsvertrages mit dem BMBWF jeweils die relativen Lösungshäufigkeiten in der Gesamtgruppe (ohne weitere Aufschlüsselungen nach Bundesland, Geschlecht oder anderen Merkmalen der Prüflinge) zur Verfügung gestellt wurden. Als „Kontrolltexte“ wurden zudem auch die Mathematik BHS-Maturajahrgänge 2016-2018, die bayrische Abiturprüfung 2016-2018, sowie als mathematikfremde, sprachlich jeweils potentiell anders angelegte Texte die Deutsch AHS-Maturajahrgänge 2016-2018, das erste Kapitel aus „Harry Potter und der Stein der Weisen“ (als gängigem Text der Kinder- und Jugendliteratur) und ein Mietvertrag des VKI (als Paradebeispiel für einen sprachlich herausfordernden Text) mitanalysiert.

An der Studie waren neben Andreas Vohns als Projektleitung noch Tamara Obereder als wissenschaftliche und Janine Egger und Stefan Scheiber als studentische Mitarbeiter/innen beteiligt. Die Finanzierung erfolgte zu 100 % aus In-Kind-Leistungen der Universität Klagenfurt, seitens BMBWF wurden nur Daten geliefert und Auskünfte erteilt. Für einige informelle Fragen zu linguistischen Aspekten konnten wir auf Markus Pissarek zurückgreifen, Professor für Deutschdidaktik an der Universität Klagenfurt.

### 3.1. Forschungsfragen

Gemäß den oben beschriebenen Motivationen und Forschungsdesideraten wurden drei Hauptforschungsfragen bearbeitet:

1. Wie gestaltet sich das sprachliche Anforderungsprofil der Aufgabenstellungen des ersten Teils der AHS-Zentralmatura (Haupttermine) im Vergleich zu anderen mathematischen Prüfungstexten und fachfremden Vergleichstexten?

Von uns näher operationalisiert zu:

Wie sind die Aufgaben bzgl. gut isolierbarer, standardisiert bzw. niedrig-inferentiell bestimmbarer Merkmale von Textverständlichkeit im Vergleich zu relevanten Vergleichstexten (Mathematikmatura BHS, Zentralabitur Mathematik Bayern, Deutschmatura AHS, Mietverträge, Harry Potter) einzuordnen?

2. Welche (statistischen) Zusammenhänge dieser Merkmale mit den Lösungshäufigkeiten ist bei den für das Bestehen relevanten Aufgaben der AHS-Zentralmatura (Haupttermine) nachweisbar?
3. Welche (statistischen) Zusammenhänge von Merkmalen des avisierten mathematischen Verständnisses (Kontextualisierung, Ebenen des Verstehens nach Maier und Steinbring (1998), Inhaltsbereiche) mit den Lösungshäufigkeiten ist bei den für das Bestehen relevanten Aufgaben der AHS-Zentralmatura (Haupttermine) nachweisbar?

Es läge u. U. nahe, auch nach einer Verknüpfung von 2.) und 3.), also der Interaktion sprachlicher und mathematischer Aspekte zu fragen. Dies haben wir zumindest teilweise dadurch umgesetzt, dass wir für bestimmte Teilgruppen getrennt Untersuchungen vorgenommen haben (z. B.: Spielt die Textlänge im Bereich „Algebra, Geometrie“ eine Rolle? Wirken sich niedrige Anteile alltagssprachlichen Wortschatzes bei Aufgaben in außermathematischen Kontexten aus?). Man stößt hier allerdings sehr rasch an Grenzen, da bei Verknüpfung mehrerer Merkmale (etwa die drei Unterscheidungen: offen vs. geschlossen, mit vs. ohne Kontext, Analysis vs. Stochastik) eben nur sehr wenige Beispiele übrig bleiben. Wir haben ebenso auf multivariate Regressionsverfahren letztlich verzichtet. Einige explorativ gerechnete Modelle lieferten keine signifikanten Ergebnisse und es ist auch dort fraglich, ob die Anzahl der Aufgaben überhaupt ausreicht, um über „typische“ Eigenschaften von Aufgaben zu treffen, die Träger mehrerer Merkmalsausprägungen sind. Man bedenke etwa, dass je Inhaltsbereich je Jahr nur 8 Beispiele zur Prüfung kommen, man redet dann spätestens bei Verknüpfungen von drei Merkmalen u. U. nur noch über Einzelaufgaben innerhalb eines Prüfungsjahrgangs.

### 3.2. Datensammlung und -aufbereitung

Was die Sammlung der Daten angeht, liegen die hier analysierten Texte selbst allesamt als offene Daten vor, allerdings weit überwiegend in Form von PDF-Dateien. Für die Datenanalyse war es nötig, die Daten in reine Textformate (TXT) zu konvertieren, da die teilweise automatisierbare Bestimmung linguistischer Merkmale durch entsprechende Software eine solche Formatierung voraussetzt. Für die untersuchten mathematischen Texte ergibt sich als weitere Herausforderung der Datenaufbereitung, dass sämtliche nicht in natürlicher Sprache bzw. als linearer Text vorliegenden Elemente (im Weiteren kurz als „nicht-textuelle Token“ bezeichnet) wie Grafiken, Tabellen und Formel ausdrücke manuell entfernt bzw. durch entsprechende Platzhalter ersetzt werden mussten (z. B. <va> für auftretende Variablennamen, <gl> für Gleichungen, <sd> statistische Darstellung; insgesamt wurden 36 solcher Platzhalter verwendet).

Für die weiteren Analysen insbesondere betreffend des Wortschatzes wurden die so aufbereiteten Dateien mit Hilfe des Programms TreeTagger (Schmid, 1994) computerlinguistisch weiterverarbeitet, welche drei Aufgabenstellungen übernimmt:

1. **Tokenisierung:** Der Text wird in einzelne *Token* zerlegt, das können Worte sein, nicht-textuelle Token (Zahlen bzw. unsere Platzhalter) oder Elemente der Interpunktion.

<b>Merkmal</b>	<b>Auswertung</b>	<b>Kurzbeschreibung</b>
Typ	vorliegend	Aufgabentyp (offen, halboffen, geschlossen, MC, etc.)
Kompetenz	vorliegend	Zuordnung zu einer Grundkompetenz (daraus abgeleitet Inhaltsbereich; BAP, BHS: nur Inhaltsbereich)
Stufe	vorliegend	Zuordnung zu einer Schulstufe
Lösungshäufigkeit	vorliegend	Prozentanteil korrekter Lösungen
O-M-A-Modell	vorliegend	Zuordnung zu den Ebenen des Modells von Siller et al. (2016)
Textlänge	automatisiert	Anzahl der Worte
gSMOG, RIX	automatisiert	Lesbarkeitsindizes (s. u.)
Häufigkeit Token je Wortart	automatisiert	Anzahl/Anteil der Wortarten (Präposition, Substantiv, Adjektiv etc.), nicht-textuellen Token
Häufigkeit Token je Register	teils automatisiert	Alltagssprache durch Abgleich mit Wortlisten (automatisiert, s. u.), Bildungs- und Fachsprache (manuell, s. u.)
Checkliste Sprachhürden	manuell	Identifikation von sieben potentiell schwierigkeitsgenerierenden Merkmalen (s. u.)
Kontext	manuell	Zuordnung eines Kontextbereichs (s. u.) bzw. ob überhaupt außermathematisch kontextualisiert
Verstehensebene	manuell	Verstehensebene nach Maier und Steinbring (1998)

Tabelle 1: Erfasste Text- bzw. Aufgabenmerkmale und Auswertungsmethodik

- 2. Lemmatisierung:** Für jedes textuelle Token wird versucht, eine Grundform (Lexem) zu bestimmen (aus „Mannes“ wird „Mann“; aus allen bestimmten Artikeln wird „die“), was vor allem für Wortschatzanalysen hilfreich ist.
- 3. Linguistisches Tagging:** Für jedes textuelle Token wird versucht, die Wortform (Präposition, Substantiv, Adjektiv etc.) zu ermitteln, was insofern interessant sein kann, als gewisse Häufungen (etwa bei nicht lokalisierenden Präpositionen, vgl. Gürsoy, 2014, S. 137) für mathematische Aufgabentexte als typisch und potentiell schwierigkeitsgenerierend gelten.

Es ist darauf hinzuweisen, dass die Software 2.) und 3.) mit probabilistischen Methoden implementiert. Das Programm ist an Zeitungskorpora trainiert worden, für alltagssprachliche Texte und nicht-mathematische Bildungssprache erhält man damit in der Regel sehr gute Erkennungsraten (> 98%), es mussten allerdings eine Reihe fachsprachlicher Termini, Eigennamen und spezifische österreichische Wortvarianten sowie die vergebenen Platzhalter für nicht-textuelle Elemente in ein Benutzerwörterbuch eingepflegt werden (umfasst etwas über 400 Token).

Lösungshäufigkeitsdaten lagen für die AHS-Mathematikmatura und die BHS-Mathematikmatura (jeweils nur Aufgaben des ersten Teils) vor, verwendet wurden allerdings nur die Daten für die AHS. Für die BHS wurden ebenfalls nur Aufgaben des ersten, schultypübergreifenden Teils analysiert, für die bayrische Abiturprüfung nur diejenigen des ersten, technologiefreien 90-minütigen Teils, um über alle drei Prüfungen hinweg vergleichbare Bearbeitungszeiten zu erhalten.

### 3.3. Datenauswertung

In Tabelle 1 listen wir die wesentlichen Merkmale der Aufgaben bzw. Texte auf, die von uns bestimmt wurden und erläutern jeweils die Art der Bestimmung und soweit erforderlich auch grob, was unter dem Merkmal verstanden wird (Detailerläuterungen erfolgen ggf. anschließend). Mathematikbezogene Merkmale wurden dabei natürlich nur für die Mathematikaufgaben bestimmt, einige davon (bei denen vergleichende Analysen aufgrund der unterschiedlichen Anlage von AHS-Matura, BHS-Matura und bayrischer Abiturprüfung (BAP) nicht zielführend erschienen) nur für die AHS-Aufgabenstellungen.

Zur Erfassung der automatisiert ermittelbaren linguistischen Merkmale wurden neben dem schon erwähnten TreeTagger noch das Paket `quanteda` (Benoit et al., 2018) unter R (R Core Team, 2018) eingesetzt

sowie für den RIX das Regensburger Analysetool für Texte (RATTE; Wild und Pissarek, o.J.). Für alle weiteren statistischen Auswertungen wurden die Basispakete von R verwendet, ergänzend ggpubr (Kassambara, 2018) unter R und für Häufigkeitsanalysen und Abgleiche mit Wortlisten auch Excel, für verschiedene explorative Analysen ferner noch die online zugängliche Common Online Data Analysis Platform (CODAP).

*Lesbarkeitsindizes* versuchen die Schwierigkeit eines Textes auf leicht bestimmbare linguistische Merkmale wie Wort- und Satzlängen oder andere Oberflächenmerkmale des Textes zurückzuführen. Der von uns auf Empfehlung von Markus Pissarek verwendete gSMOG (Simple Measure of Gobbledygook – german) berechnet sich gemäß der Formel: 
$$\text{gSMOG} = \sqrt{\frac{\text{Wörter mit drei oder mehr Silben} \cdot 30}{\text{Zahl der Sätze}}} - 2.$$
 Der gSMOG liefert ein Resultat, welches „näherungsweise das Lesealter (in Schulstufen), für das der Text geeignet ist“ (Wild & Pissarek, o.J.) angeben soll. Seine Erklärungskraft wird darauf zurückgeführt, dass mit „der steigenden Schwierigkeit der Texte [...] die Zahl der längeren Worte und auch der langen Sätze“ (Bamberger & Vanecek, 1984, S. 58) wächst. Eine solche Einschätzung dürfte sich wohl auch auf insbesondere Situationsbeschreibungen in mathematischen Aufgabenstellungen übertragen lassen, d. h. tendenziell dürfte die Textverarbeitung auch in mathematischen Texten schwieriger werden, wenn Sätze länger werden und längere Worte (insbes. Komposita) enthalten. Das angegebene Lesealter ist aber wohl kaum so übertragbar, da das Verstehen der Aufgabentexte u. a. von der Kenntnis enthaltener mathematischer Fachbegriffe und nicht-textueller Token abhängt, die den gSMOG nicht beeinflussen. Relativ höhere Werte geben allerdings auch für mathematische Aufgabentexte immer noch an, wie sehr diese Texte durch die genannten Aspekte geprägt und somit rein auf der Textebene anspruchsvoller sind. Die Verwendung eines solchen Maßes für einzelne Aufgabentexte ist weiters nicht unproblematisch, weil die Normierung ursprünglich an deutlich längeren Texten erfolgte (man findet als Empfehlungen mindestens 30 Sätze oder weit weniger strikt zumindest 100 Worte; eine Schranke, die nicht wenige Aufgaben immer noch unterschreiten). Es ist nicht ganz auszuschließen, dass Variationen der Werte für kurze Aufgabentexte eher als zufälliges „Rauschen“ zu interpretieren wären. Wir haben daher ergänzend auch noch den von Wild und Pissarek (o. J.) entwickelten Index RIX angesehen, konnten aber im Wesentlichen die durch den gSMOG getroffenen Einschätzungen reproduzieren. Tatsächlich hat sich die Verwendung der Indizes zumindest als potentiell schwierigkeitsklärender Faktor auf Aufgabenebene als im Nachhinein vernachlässigbar herausgestellt (s. u.); bedingt anwendbar sind diese Maße eher als Vergleichswerte für die unterschiedlichen Gesamtprüfungstexte (Anspruch AHS und BHS Mathematik-Prüfungsangaben insgesamt gegenüber Deutschmatura und BAP).

Eine Einteilung der Lexeme in unterschiedliche *Sprachregister* gestaltet sich auf Einzelwortebene insofern nicht vollständig durch Wortlisten automatisierbar, als Lexeme aus der Alltagssprache wie „Funktion“ oder „ähnlich“ in mathematischen Kontexten mit spezifischen Fachbedeutungen belegt sein können. Eingedenk dieser Problematik wurde pragmatisch ein mehrschichtiges Verfahren angewandt: Alle Lexeme wurden mit drei Wortlisten verglichen: (a) den tausend 10-Jährigen geläufigsten Vokabeln nach Bamberger und Vanecek (1984, S. 176-179), (b) dem Basiswortschatz für das Sprachniveau B1 (Glaboniat, Perlmann-Balme & Studer, 2013, S. 16-102) und (c) einem repräsentativen Auszug (100.000 Lexeme) aus dem Korpus childLex (Schroeder, Würzner, Heister, Geyken & Kliegl, 2015), der deutlich breiter Kinder- und Jugendlektüre für bis zu 12-Jährige beinhaltet (u. a. auch die Harry-Potter-Bücher). Wörter, die in (a) oder (b) enthalten sind, wurden dem Sprachregister „Alltagssprache“ zugeordnet. In einem zweiten Schritt wurden (zunächst ohne Berücksichtigung des Kontextes) in den verbliebenen Lexemen solche identifiziert, die eindeutig der mathematischen Fachsprache (z. B. „Integral“) oder eindeutig nicht der mathematischen Fachsprache (z. B. „Schneeschuhhase“) zuzuordnen waren, das zweite Register wurde als „nicht-mathematische Bildungssprache“ bezeichnet. Die verbliebenen Lexeme wurden jeweils im Kontext betrachtet (z. B. für „steigend“ oder „Wachstum“ erforderlich) und es wurde als dritte Kategorie „mathematische Unterrichtssprache“ eingeführt, für solche Worte, die zwar nicht im eigentlichen Sinne als mathematische Fachsprache deklarierbar sind, von deren Verwendung im Mathematikunterricht aber nahezu sicher ausgegangen werden kann (etwa: „Halbwertszeit“). Die Zuordnungen erfolgten dabei jeweils zunächst von zwei Personen aus dem Projektteam getrennt, dann im Falle von Unklarheiten

<b>Element</b>	<b>Beschreibung</b>
Komposita	Finden sich im Text mehrere Komposita?
Synonyme	Werden im Text Begriffe synonym verwendet?
getrennte Verben	Befinden sich im Text mehrere getrennte Verben?
Pronomen	Sind im Text mehrere Pronomen zu finden?
Aufgabenstellung	Ist die Aufgabenstellung nicht kurz und prägnant formuliert?
Verhältnisse	Befinden sich im Text Begriffe, die Verhältnisse zwischen Elementen angeben?
Adverbien	Befinden sich im Text mehrere Adverbien?
Präpositionen	Befinden sich im Text mehrere für den Prozess des Mathematisierens entscheidenden Präpositionen?
Passivkonstruktionen	Befinden sich im Text Passivkonstruktionen?
Nebensatzkonstruktionen	Befinden sich im Text Nebensatzkonstruktionen?

Tabelle 2: Verwendete Checkliste potentieller sprachlicher Hürden

auch unter Hinzuziehung des Projektleiters als „Schiedsrichter“. Dasselbe Verfahren der Konsensfindung wurde auch auf alle weiteren manuell ermittelten Aufgabenmerkmale angewandt.

Hinter der *Checkliste Sprachhürden* verbirgt sich eine Liste mit sprachlichen Merkmalen, die sich in der Literatur (genauer: bei Leisen, 2011; Gürsoy et al., 2013; Weis, 2013) insbesondere für zweisprachige Lernende als potentiell relevant herausgestellt haben und welche durch die von uns automatisiert erhobenen Kategorien nicht direkt erfasst werden (vgl. Tabelle 2). Diese Liste ist dementsprechend weniger standardisiert als die linguistischen Maße für Lesbarkeit, geht dafür deutlich stärker auf die Spezifika mathematischer Aufgabentexte ein. Für die Elemente der Checkliste wurden 0 (unzutreffend), 0.5 (mehr oder weniger zutreffend) oder 1 Punkt (deutlich zutreffend) vergeben und anschließend die Summe gebildet. Ein derart hemdsärmelig konstruiertes Maß hat dann notwendig eher heuristischen Charakter.

Bei der Kategorisierung der außermathematischen *Kontextualisierung* haben wir uns für eine leichte Modifikation der Kontexteinteilung von PISA 2015 entschieden (vgl. OECD, 2016, S. 74; Tabelle 3). Von uns wurden allerdings wirtschaftstheoretische Kontexte mit denen des sozialen Bereichs zusammengefasst und aus dem Bereich der als “scientific” bezeichneten Kontexte haben wir zudem die rein innermathematisch-kontextfreien Aufgabenstellungen als eigene Kategorie herausgelöst, da uns ei-

<b>Code</b>	<b>Kurzbezeichnung</b>	<b>Beschreibung</b>
PE	Personal	Problems classified in the personal context category focus on activities of one’s self, one’s family or one’s peer group. The kinds of contexts that may be considered personal include (but are not limited to) those involving food preparation, shopping, games, personal health, personal transportation, sports, travel, personal scheduling and personal finance.
OC	Occupational	Problems classified in the occupational context category are centred on the world of work. Items categorised as occupational may involve (but are not limited to) such things as measuring, costing and ordering materials for building, payroll/accounting, quality control, scheduling/inventory, design/architecture and job-related decision making.
SE	Socio-Economic	Problems classified in the societal context category focus on one’s community (whether local, national or global). They may involve (but are not limited to) such things as voting systems, public transport, government, public policies, demographics, advertising, national statistics and economics. + <i>Anwendungen der Wirtschaftstheorie und -forschung (Ökonomie, BWL, VWL)</i>
SC	Scientific	Problems classified in the scientific category relate to the application of mathematics to the natural world and issues and topics related to science and technology. Particular contexts might include (but are not limited to) such areas as weather or climate, ecology, medicine, space science, genetics, measurement [...].
CL	<i>Contextless</i>	<i>Kein oder rein innermathematischer Kontext</i>

Tabelle 3: Verwendete Einteilung der Kontextbereiche (englische Passagen: OECD, 2016, S. 74)

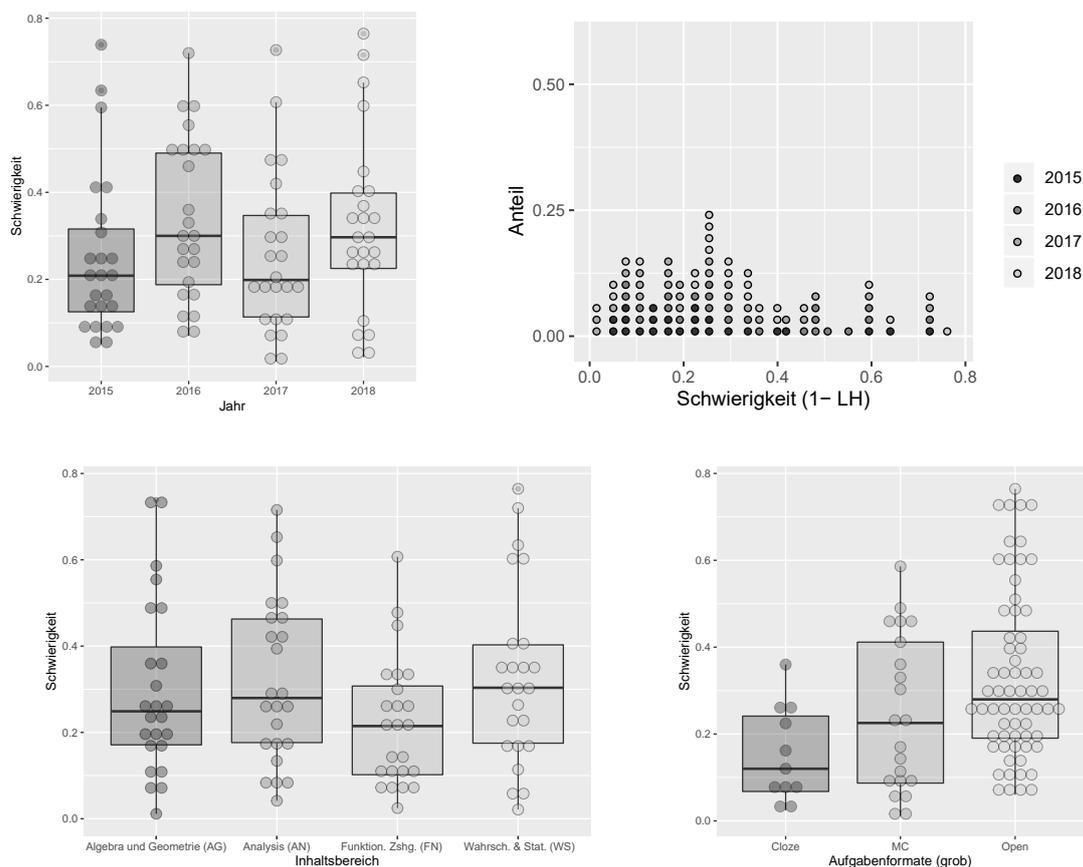


Abbildung 2: Verteilung der empirischen Aufgabenschwierigkeiten

ne Unterscheidung kontextualisierter von kontextfreien Beispielen nicht unwichtig erschien. Für einige Analysen wurde auch eine einfache dichotome Unterscheidung (mit/ohne Kontext) verwendet.

Für die Untersuchung von *Zusammenhangs-* bzw. *Unterschiedshypothesen* wurden die von 1 subtrahierten Lösungshäufigkeiten als *empirische Schwierigkeiten* der Aufgabenstellungen interpretiert. Dazu werden dann auch Aufgabenstellungen miteinander verglichen, die in verschiedenen Jahren geprüft wurden. Eine solche Zusammenfassung ist eine der deutlichen Kompromisse dieser Untersuchung. Forschungslogisch kann man sie dadurch zu rechtfertigen versuchen, dass den Ausgangspunkt des medialen Interesses hohe Schwankungen der Durchfallsquoten darstellen. Diese Schwankungen wären dann problematisch, wenn ihnen nicht unterschiedliche Fähigkeiten der Lernenden zu Grunde liegen. Wir nehmen also hypothetisch an, dass die mathematischen Leistungen der Lernenden im Wesentlichen konstant bleiben. Wenn nun dennoch bei bestimmten Terminen mehr oder weniger Personen durchfallen, ist noch unklar, inwieweit sich dies in den empirischen Aufgabenschwierigkeiten und deren Verteilung (insbes. zentrale Tendenz, Streuung) widerspiegelt. Tatsächlich rechtfertigen die Verteilungen der Schwierigkeiten in den einzelnen Jahren (vgl. Abb. 2, oben links) noch die Annahme einer gemeinsamen Verteilung, d. h. weder die zentrale Tendenz (Kruskal-Wallis-Test) noch die Streuung (Levene-Test) der Daten unterscheiden sich so deutlich, dass die Unterschiede nicht als zufällige Stichprobenunterschiede der Ziehung von je 24 Beispielen aus einer Grundgesamtheit („gemeinsames Aufgabenuniversum“) interpretiert werden könnten. Fasst man sodann alle Beispiele zu einer gemeinsamen Verteilung zusammen, so entspricht diese allerdings eher keiner Normalverteilung (schon grafisch gut erkennbare Gruppe von Ausreißern nach oben, vgl. Abb. 2, oben rechts). Wir haben daher neben Pearson-Korrelationen bei den Zusammenhangsanalysen auch stets Rangkorrelationen berechnet und zusätzlich verschiedene Mittelwertvergleiche in Teilgruppen (Aufteilung nach Schwierigkeit) vorgenommen. Auch die in diesem Sinne verstandene Verteilung der Aufgabenschwierigkeiten über die verschiedenen Inhaltsbereiche (vgl. Abb. 2, unten links) weist nicht auf einen signifikanten Verteilungsunterschied hin, ein solcher findet sich lediglich für die

von uns als „Cloze“ bezeichneten Aufgabenstellungen (Lückentext, Zuordnung) gegenüber den offenen Aufgabenformaten (halboffen, offen, Konstruktion). Den Aufgabenformaten ist schon deshalb bei der Interaktion mit Textmerkmalen eine Bedeutung beizumessen, da die MC-Aufgabenstellungen und die Lückentext-Aufgaben neben der Fragestellung noch mögliche Antworten und somit durchschnittlich deutlich mehr Text enthalten als die offenen Formate.

## 4. Ausgewählte Ergebnisse

### 4.1. Deskriptiver Teil (Forschungsfrage 1)

In Tabelle 4 sind jeweils die Gesamtanzahlen der Lexeme in den Prüfungsangaben (erste Teile) der AHS- und BHS-Mathematikmatura, der bayrischen Abiturprüfung Mathematik (BAP-M) und der AHS-Deutschmatura angegeben, sowie die jeweiligen relativen Anteile gemäß der verwendeten Wortlisten bzw. der aus diesen bzw. durch manuelle Zuordnung ermittelten Sprachregister. Es ist auffällig, dass sich die mathematischen Prüfungsarbeiten hinsichtlich des Anteils an bildungssprachlichen Anteilen insgesamt nur moderat unterscheiden (ca. 51 % AHS-M gegenüber 44 % BAP-M), die Deutschmatura-Prüfungen und der analysierte Mietvertrag sind hinsichtlich dieses Sprachregisters insgesamt anspruchsvoller. Schlüsselt man die Bildungssprache in mathematische Fachsprache, Unterrichtssprache Mathematik und nicht-mathematische Bildungssprache weiter auf, so wird erkennbar, dass der Anteil mathematischer Fachsprache in der BAP-M deutlich höher liegt als in AHS und BHS und umgekehrt diejenigen Lexeme, die eindeutig nicht-mathematische Bildungssprache darstellen, mit knapp einem Drittel in den beiden Maturavarianten um etwa 10 Prozentpunkte höher liegen als bei der bayrischen Abiturprüfung. Hinsichtlich dieser Anforderungen stellen also die österreichischen Prüfungen sprachlich andere, tendenziell höhere Ansprüche als ihr bayrisches Pendant.

Dies gilt auch für den Umfang der jeweils zu verarbeitenden Gesamttextmenge (in der Tabelle als Worte je Dokument), wenn man den Teil 1 der AHS-Mathematikmatura, den A-Teil der BHS-Mathematikmatura und den ersten, technologiefreien, auf 90 Minuten angesetzten Teil der bayrischen Abiturprüfung vergleicht: Die Prüfungsangaben des ersten Teils der AHS-Matura umfassen mehr als die dreifache Textmenge (1320 gegenüber knapp 420 Worten) wie die des ausgewählten Teils der bayrischen Abiturprüfung, auch beim A-Teil der BHS ist es noch mehr als die doppelte Textmenge. Ein Vergleich mit der AHS-Deutschmatura ist aus mehreren Gründen schwierig: In der Deutschmatura werden je drei Vorschläge je Termin an die Prüflinge gegeben, die etwa die 1,5-fache Textmenge der Mathematikmatura ausmachen (insgesamt also ca. die 4,5-fache Textmenge). Nähme man bei der AHS-Mathematikmatura noch den 2. Teil hinzu, wäre diese in etwa so lang wie ein einzelner Vorschlag für Deutsch, u. U. sogar etwas länger. Da Prüflinge im Fach Deutsch aber wohl auch die beiden dann von ihnen nicht bearbeitete-

Sprachregister bzw. Wortliste	AHS-M (15-18)	BHS-M (16-18)	BAP-M (16-18)	AHS-D (16-18)	Harry Potter	Mietvertrag
Bamberger & Vanecek 1000	223 (25,0 %)	193 (28,8 %)	128 (33,7 %)	636 (15,9 %)	420 (38,9%)	132 (21,6 %)
Wortschatz Deutsch B1	423 (47,4 %)	342 (51,0 %)	208 (54,7 %)	1405 (35,1 %)	618 (57,2%)	244 (40,0 %)
<b>Weder-Noch:= Bildungssprache</b>	<b>455 (51,0%)</b>	<b>323 (48,2%)</b>	<b>167 (43,9%)</b>	<b>2541 (63,4%)</b>	<b>437 (40,4%)</b>	<b>357 (58,5%)</b>
Literaturkorpus childLex (11-12)	677 (75,8 %)	527 (78,7 %)	310 (81,6 %)	2953 (73,7 %)	1021 (94,4%)	419 (68,7 %)
Weder-Noch (inkl. childLex)	207 (23,2 %)	130 (19,4 %)	65 (17,1 %)	1013 (25,3 %)	51 (4,7 %)	186 (30,5 %)
Mathematische Fachsprache	118 (13,2 %)	54 (8,1 %)	74 (19,5 %)			
Unterrichtssprache Mathematik	41 (4,6 %)	27 (4,0 %)	9 (2,4 %)			
Nicht-mathematische Bildungssprache	295 (33,0 %)	219 (32,7 %)	84 (22,1 %)			
Gesamt (Lemmata)	893 (100,0 %)	670 (100,0 %)	380 (100,0 %)	4005 (100 %)	1081 (100,0 %)	610 (100,0 %)
Je Dokument (Lemmata)	223,25	223,33	63,33	455	1081	610
Gesamt (Worte)	5283 (591,6 %)	3122 (466,0 %)	2531 (666,0 %)	16380 (409,0 %)	5795 (536,0 %)	1907 (312,6 %)
Je Dokument (Worte)	1320,75	1040,67	421,83	1820	5795	1907

Tabelle 4: Sprachregister auf Basis der Lexeme (Wortgrundformen)

AHS-M	gSMOG	BHS-M	gSMOG	BAP-M	gSMOG
Termin 2015	7.99				
Termin 2016	7.43	Termin 2016	9.41	Termin 2016	7.41
Termin 2017	8.16	Termin 2017	8.58	Termin 2017	7.07
Termin 2018	8.20	Termin 2018	8.95	Termin 2018	7.51
		Zahlen und Maße	8.37		
Algebra & Geometrie	7.19	Algebra & Geometrie	8.91	Analytische Geometrie	6.11
Funktionale Abhängigkeiten	8.02	Funktionale Zusammenhänge	8.61		
Analysis	7.97	Analysis	8.71	Analysis	6.99
Wahrscheinlichkeit & Statistik	8.29	Stochastik	9.37	Stochastik	8.77
Mit Kontext	8.17			Mit Kontext	9.74
Innermathematisch	7.58			Innermathematisch	7.00
Minimum (18-A01)	3.92	Minimum (18-A05)	7.08	Minimum (18-GE-AG02-02)	4.00
Maximum (15-A20)	13.18	Maximum (18-A04)	10.65	Maximum (16-AN-AG01-03)	12.49

Tabelle 5: Standardisierte Lesbarkeit (nur textuelle Token berücksichtigt)

ten Aufgabenvorschläge nicht vollkommen ungelesen bei Seite legen dürften, ist wohl doch eine deutlich größere Textmenge zu verarbeiten und, ganz unabhängig davon, laufen die Lösungen der Aufgaben auf deutlich umfangreichere Textproduktionen hinaus.

Wir haben oben bereits darauf hingewiesen, dass die Bestimmung des *Lesbarkeitsindex gSMOG* für kurze Aufgabentexte, wie sie jedenfalls in der AHS-Mathematikmatura vorkommen, für zufällige Schwankungen anfällig ist (es sind jeweils Maxima und Minima der jeweiligen Einzelaufgaben angeführt, bei deren Interpretation ist aber Vorsicht geboten). Tabelle 5 enthält vornehmlich die besser interpretierbaren Werte für die Gesamtdokumente eines Prüfungsjahrgangs bzw. aus allen Aufgabenstellungen eines Inhaltsbereichs zusammengefasste Dokumente (die Aufschluss über typischerweise nach Inhaltsbereich unterschiedlicher Lesbarkeit geben könnten), deren Textumfang zufällige Schwankungen weitgehend dämpfen sollte. Von der gemäß gSMOG vor allem an der Wort- und Satzlänge festgemachten Lesbarkeit her unterscheiden sich BHS- und AHS-Matura eher moderat, beide unterscheiden sich wiederum deutlicher von der bayrischen Abiturprüfung. In allen Prüfungsarbeiten werden die Texte aus dem Bereich der Stochastik durch den gSMOG als moderat (AHS-M, BHS-M) bzw. deutlich (BAP-M) überdurchschnittlich anspruchsvoll eingeschätzt. Dies dürfte sich auch durch die unterschiedliche Kontextualisierung der Aufgaben erklären lassen (s. u. Tabelle 6), denn wiederum sind bei der AHS-Matura die kontextualisierten Beispiele moderat anspruchsvoller als die kontextfreien, bei der bayrischen Abiturprüfung sogar als deutlich anspruchsvoller eingeschätzt worden. Tatsächlich enthält die BHS keine einzige rein innermathematische Fragestellung und die (wenigen) Anwendungen in den bayrischen Abiturprüfungen konzentrieren sich im Bereich Stochastik. Man kann also die unterschiedlichen Anforderungen hier durchaus als durch die unterschiedliche Zuneigung der jeweiligen Prüfungsarbeiten zu „mathematical literacy“ bzw. „angewandter Mathematik“ mitverursacht ansehen. Erwartungsgemäß gilt der Text von „Harry Potter und der Stein der Weisen“ (Kapitel 1) gemäß gSMOG als klar weniger anspruchsvoll (5.74), der Mietvertrag als klar anspruchsvoller (11.54) wie sämtliche Gesamtprüfungstexte. Für die AHS-Deutschmatura liegen die Lesetexte (8.75) auf einem mit der BHS-Mathematikmatura vergleichbaren Niveau, die Aufgabenstellungen selbst wären als etwas anspruchsvoller einzustufen (10.1).

Hinsichtlich der *Kontexte* (Tabelle 6) ist mit Blick auf die AHS festzustellen, dass Kontexte des persönlichen, gesellschaftlich-ökonomischen und naturwissenschaftlichen Bereichs ähnlich häufig vorkommen (zwischen 16 % und 21 % Prozent mit einem Ausreißer in 2015 bei den persönlichen Kontexten), berufliche Kontexte hingegen nur eine sehr untergeordnete Rolle spielen. Im hier analysierten allgemeinen Teil der BHS-Mathematikmatura dominieren ganz eindeutig die naturwissenschaftlichen Kontexte. Überhaupt sind dort (der Schulfachbezeichnung „Angewandte Mathematik“ durchaus entsprechend) nur 0 bis 11 % der Aufgaben rein innermathematisch, während sich das Verhältnis bei der bayrischen Abi-

Prüfung	Personal	Occupational	Socio-economic	Scientific	Contextless
AHS-M 15	29,2 %	4,2 %	16,7 %	16,7 %	33,3 %
AHS-M 16	16,7 %	4,2 %	12,5 %	20,8 %	50,0 %
AHS-M 17	16,7 %	4,2 %	12,5 %	16,7 %	50,0 %
AHS-M 18	20,8 %	0,0 %	12,5 %	12,5 %	54,2 %
BHS-M 16	0,0 %	16,7 %	0,0 %	72,2 %	11,1 %
BHS-M 17	30,0 %	0,0 %	30,0 %	40,0 %	0,0 %
BHS-M 18	0,0 %	10,0 %	20,0 %	60,0 %	10,0 %
BAP-M 16	5,9 %	0,0 %	0,0 %	0,0 %	94,1 %
BAP-M 17	6,3 %	0,0 %	0,0 %	12,5 %	87,5 %
BAP-M 18	0,0 %	0,0 %	11,8 %	0,0 %	88,2 %

Tabelle 6: Aufgabenanteile nach Kontextbereichen (vereinzelte Mehrfachzählungen)

turprüfung faktisch umkehrt. Die AHS-Mathematikmatura hat abgesehen vom Jahr 2015 einen recht konstanten Anteil von ca. 50 % kontextfreien Fragestellungen. Eine nicht unwichtige Information geht allerdings aus der Tabelle nicht hervor: Während die BHS-Matura mehrere Fragestellungen innerhalb eines übergreifenden Kontextes stellt (5-6 größere Aufgaben mit Teilaufgaben), sind sämtliche etwa 12 kontextualisierten Teilaufgaben je Termin der AHS-Matura unabhängig voneinander. Den Lernenden begegnen also etwa doppelt so viele außermathematische Kontexte je Prüfungsarbeit.

Als nicht ganz einfach hat sich die Zuordnung der Beispiele der AHS-Zentralmatura zu den *Verstehens-ebenen* nach Maier und Steinbring (1998) erwiesen. Unter den 96 Aufgaben konnten wir faktisch keine finden, welche im engeren Sinne der *elaborativen oder reflexiven Komponente* zuzuordnen wären, ebenso zielen kaum Beispiele deutlich abgrenzbar von der *argumentativen Komponente* auf die *relationale Komponente* ab, so dass wir diese in einer Kategorie zusammengefasst haben; ihr wurden 18 Beispiele (bzw. 18,8 %) zugeordnet. Wir haben weiters alle Beispiele der *konzeptuellen Komponente* zugeordnet, bei denen wir davon ausgehen, dass ein rein syntaktisches Verständnis eines durch die Aufgabenstellung direkt oder indirekt referenzierten mathematischen Begriffs nicht ausreicht; in dieser Kategorie finden sind 43 Beispiele (bzw. 44,8 %). Es verbleiben dann noch 35 Beispiele (bzw. 36,5 %), welche klar auf die *prozedurale Komponente* abzielen.

#### 4.2. Zusammenhangs- und Unterschiedsanalysen (Forschungsfragen 2, 3)

Hinsichtlich Forschungsfrage 2 ist in Summe festzuhalten, dass sich weder insgesamt, noch unter Kontrolle von Prüfungsjahrgang, mathematischem Inhaltsgebiet, Aufgabenformat und/oder Kontextbereich signifikante Zusammenhänge zwischen der empirischen Aufgabenschwierigkeit und (a) Anteil/Anzahl bildungs- und fachsprachlicher Elemente, (b) Anteil/Anzahl nicht textueller Elemente (Formeln, Grafiken), (c) standardisierter Lesbarkeit oder (d) Anzahl sprachlicher Schwierigkeiten (gemäß Checkliste) nachweisen lassen (Niveau jeweils  $p < 0.05$ , sowohl für Pearson- als auch für Rang-Korrelationen). Die Ergebnisse bestätigen sich auch bei Test auf Mittelwertunterschiede dieser Merkmale für eher schwierige und eher leichte Aufgaben. In Tabelle 7 sind für die oben genannten Merkmale Mittelwerte und Standardabweichungen für die jeweils leichtesten 16 Aufgaben (höchstens 10 % nicht korrekte Bearbeitungen) und die schwierigsten 17 Aufgaben (mindestens 49 % nicht korrekte Bearbeitungen) angegeben. Keine der aufgelisteten Mittelwertunterschiede ist signifikant ( $p < 0.05$ ). Der erkennbar höhere bzw.

Merkmal	Leichte Aufgaben		Schwierige Aufgaben	
	M	SD	M	SD
Anteil bildungssprachlicher Token	9,1 %	4,9 %	11,6 %	5,2 %
Anteil fachsprachlicher Token	11,4 %	8,9 %	8,0 %	7,5 %
Anteil nicht-textueller Token	22,8 %	24,5 %	10,3 %	9,7 %
Textlänge (Wörter)	61.9	46.4	55.7	23.5
Werte gSMOG	7.9	1.7	7.5	2.0
Werte Checkliste	2.3	2.0	3.1	1.5

Tabelle 7: Sprachliche Merkmale der leichten und schwierigen Beispiele

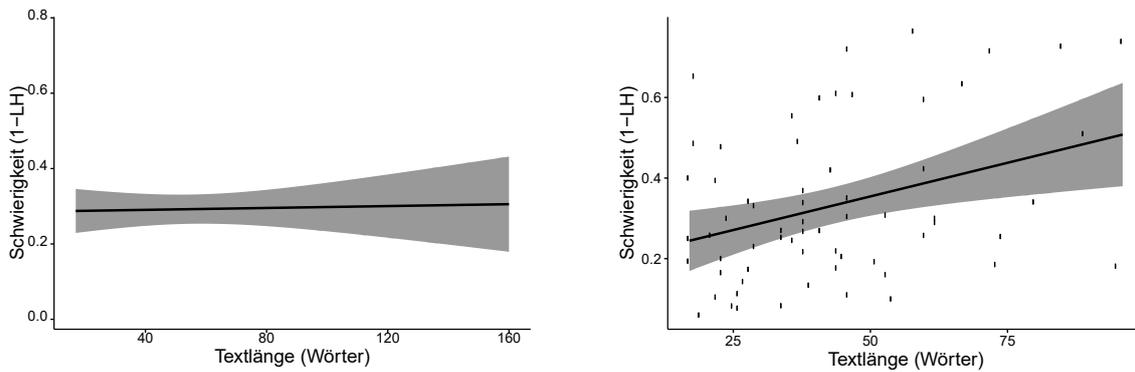


Abbildung 3: Zusammenhang Textlänge und Schwierigkeit für alle (li.) bzw. offene (re.) Aufgaben

auch deutlich stärker schwankende Anteil nicht-textueller Elemente erklärt sich nahezu vollständig aus den bei den schwierigen Beispielen fast völlig fehlenden geschlossenen Aufgabenformaten (s. u.; einige MC-Antworten und die Lösungen bei Zuordnungs-Aufgaben bestehen fast nur aus Formeln oder Funktionsgraphen, eine Aufgabe zu Potenzfunktionen (Nr. 10, 2016) enthält sogar über 50 % nicht-textueller Token). Zur Vorhersage der Aufgabenschwierigkeit taugt allerdings auch dieser Indikator nicht.

Der einzige signifikante ( $p < 0.05$ ), wenn auch mit ca. 12 % Varianzaufklärung (bzw. einem  $R$ -Wert von 0.35 bei Rangkorrelation) schwache Zusammenhang lässt sich zwischen der Schwierigkeit und der Textlänge in der Teilgruppe der offenen Aufgabenstellungen (halboffen, offen, Konstruktion) nachweisen (s. Abb. 3, rechts). Dass dieser schwache Zusammenhang in der Gesamtgruppe nicht durchschlägt, liegt am geringen Anteil der relativ wortreichen geschlossenen Formate unter den schwierigen Aufgaben.

Ein nur leicht anderes Bild ergibt sich hinsichtlich Forschungsfrage 3, also dem Einfluss der mathematischen Anforderungsprofile. Da hier lediglich kategoriale Daten vorliegen, betrachten wir in Tabelle 8 jeweils die relativen Anteile der leichtesten/schwierigsten Aufgaben und geben die Werte der Chi-Quadrat-Tests für die jeweiligen Kontingenztabelle an. Signifikante Unterschiede gibt es hier vor allem hinsichtlich der Aufgabenformate: Unter den schwierigen Aufgaben kommen geschlossene Formate (MC, Lückentext, Zuordnung) kaum vor. Hinzu kommt ein deutlich erhöhter Anteil von prozeduralen gegenüber konzeptuellen Aufgaben. Es sind also tendenziell solche Aufgaben schwierig, in denen die Lösungen aktiv prozedural ermittelt werden müssen, wohingegen unter den leichten Aufgaben solche Aufgaben überrepräsentiert sind, in denen zwar konzeptuelles Wissen angesprochen ist, aber zwischen Antwortalternativen zu wählen ist. Was das offiziell verwendete O-M-A-Modell angeht, so lassen sich

Merkmal	Leichte Aufgaben	Schwierige Aufgaben
Anteil offene Aufgaben	37,5 %	88,2 %
$\chi^2$ : 9.17; Cramer's C: 0.53; $p < 0.01$		
Anteile Verstehensebenen		
– prozedural	12,5 %	53,0 %
– konzeptuell	62,5 %	35,3 %
– argumentativ-relational	25,0 %	11,8 %
$\chi^2$ : 6.09; Cramer's C: 0.43; $p = 0.02$		
Anteile O-M-A-Modell (Bereich)		
– Operieren	43,8 %	41,2 %
– Modellbilden	25,0 %	41,2 %
– Argumentieren	31,3 %	17,6 %
$\chi^2$ : 1.29; Cramer's C: 0.19; $p = 0.43$		
Anteil O-M-A-Modell, Ebene 2	6,3 %	17,6 %
$\chi^2$ : 1.01; Cramer's C: 0.17; $p = 0.31$		
Anteil kontextualisierte Aufgaben	37,5 %	64,7 %
$\chi^2$ : 2.44; Cramer's C: 0.27; $p = 0.12$		

Tabelle 8: Weitere Merkmale der leichten und schwierigen Beispiele

keine unterschiedlichen Anteile der verschiedenen Bereiche (Operieren, Modellbilden, Argumentieren) und auch die Unterscheidung in zwei Stufen, die dieses Modell vorsieht, weist keinen signifikanten Zusammenhang mit den Aufgabenschwierigkeiten auf.

## 5. Diskussion und Ausblick

In Summe liefert die hier vorgestellte Studie wenig Anhaltspunkte dafür, dass *nicht-intendierte sprachliche Hürden* die Erhebung der *intendierten Art mathematischen Verstehens* bei der AHS-Zentralmatura Mathematik in einem für die Gesamtpopulation relevanten Ausmaß überlagern würden. Es ist darauf hinzuweisen, dass sowohl die ersten Teile der AHS- wie auch der BHS-Mathematikmatura sich hinsichtlich ihres sprachlichen Anforderungsprofils sehr deutlich vom ersten Teil der bayrischen Abiturprüfung unterscheiden. Es ist eine deutlich größere Menge an Text zu verarbeiten, die Anzahl bildungssprachlicher, nicht-mathematischer Wendungen ist deutlich größer und insbesondere den Prüflingen an der AHS begegnet eine erheblich größere Anzahl außermathematischer Kontexte, in die sich diese potentiell einfinden müssen. Diese Ergebnisse waren aufgrund der deutlich unterschiedlichen Zuneigung der Prüfungsarbeiten zu angewandter Mathematik bzw. "mathematical literacy" so oder ähnlich durchaus zu erwarten und sprechen als solche nicht für unnötig lange oder unzureichend prägnante Aufgabenstellungen.

Während sich ein gewisser, eher schwacher Zusammenhang zwischen den Textlängen offener Aufgabenstellungen und den empirischen Aufgabenschwierigkeiten herstellen lässt, weist nahezu kein anderes potentiell als schwierigkeitsgenerierend angenommenes sprachliches Merkmal der Aufgabentexte einen erkennbaren Zusammenhang mit den Aufgabenschwierigkeiten auf. Insbesondere ist nachdrücklich zu bestärken, dass unter den schwierigen Aufgaben nicht die kontextualisierten oder diejenigen überrepräsentiert sind, die auf die konzeptuelle Verstehensebene abzielen, auch nicht die von Lehrpersonen und Eltern häufig problematisierten „ungewohnten Aufgabenformate“ (MC, Lückentext, Zuordnung), sondern es viel mehr die offenen Aufgabenstellungen sind, bei denen es eher auf prozedurales Wissen und Können ankommt. Der Vergleich der sprachlichen Anforderungsprofile über die drei mathematischen Prüfungen hinweg legt nahe, dass sich deutlich andere sprachliche Anforderung umsetzen ließen, in dem zu einem der bayrischen Abiturprüfung ähnlichen, deutlich traditionelleren Verständnis von mathematischer Leistung übergegangen würde. Es ist allerdings fraglich, ob so eine Prüfung noch dem aktuellen Lehrplan gerecht würde, da sie kaum mehr mit den bildungstheoretischen Vorstellungen konform ginge, welche die aktuelle Prüfungskonzeption für sich umzusetzen beansprucht. Da aber kaum Effekte des sprachlichen Anforderungsprofils beobachtet wurden und in einem traditionelleren Verständnis mathematischer Leistungen üblicherweise offene Aufgabenstellungen und prozedurales Wissen Prüfungen dominieren, welche sich derzeit überproportional häufig unter den schwierigen Aufgaben finden, ist auch alles andere als klar, dass eine solcher Schritt überhaupt zu sozial erwünschteren Durchfallsquoten führen würde.

Es verbleibt abschließend noch einmal deutlich auf die Beschränkungen der vorliegenden Studie hinzuweisen, die sich aus dem verfügbaren Datenmaterial und/oder den getroffenen Entscheidungen in dessen Aufbereitung und Analyse ergeben: Wir haben lediglich Gesamtlösungshäufigkeiten der Aufgabenstellungen verwenden können, so dass u. U. interessante Aussagen zu Teilpopulationen (Gender, einzelne Schulklassen oder Schulstandorte, etc. pp) nicht möglich sind wie überhaupt keine Effekte untersuchbar sind, für die Daten auf Personenebene nötig wären (etwa Korrelationen zwischen Lösungshäufigkeiten einzelner Aufgaben). Es wurde heuristisch unterstellt, die Fähigkeiten der Schülerinnen und Schüler wären über die Maturajahrgänge hinweg im Wesentlichen unverändert geblieben, um Aufgabenschwierigkeiten über die Jahre hinweg vergleichen zu können. Auf multiple Regressionen wurde verzichtet. Diese Einschränkungen nehmen wir bewusst in Kauf, weil wir davon ausgehen, dass jedenfalls dann, wenn es korrekt wäre, dass die Mathematikmatura für die Mehrzahl der österreichischen Schülerinnen und Schüler eine (unzulässige) zweite Deutschprüfung wäre, Einflüsse des sprachlichen Anforderungsprofils der Aufgaben selbst trotz dieser Einschränkungen und selbst mit den vergleichsweise bescheidenen Verfahren sichtbar gemacht werden können sollten, die wir angewendet haben.

Offen muss leider die wichtige Frage bleiben, inwieweit es möglich ist, ein und dieselbe Grundkompetenz in zulässiger Weise sprachlich unterschiedlich anspruchsvoll zu erheben und ob bei einer gezielten Variation des sprachlichen Anspruchs dann auch wirklich bessere Lösungsquoten erzielt werden können – Antworten auf diese Frage lässt der Datensatz einfach nicht zu. Ebenso sind aufgrund des vorliegenden Datensatzes keine Aussagen zur Frage möglich, inwieweit für bestimmte gegenüber sprachlichen Herausforderungen von Aufgaben besonders sensible Teilpopulationen spezifische Effekte auftreten, also ob etwa für mehrsprachig aufgewachsene Lernende oder Lernende aus bildungsfernen Elternhäusern abweichende Aufgabenschwierigkeiten beobachtet werden können und etwa gewisse potentielle sprachliche Hürden für diese Teilpopulationen relevant werden, obwohl sich in der Gesamtpopulation keine solchen Effekte nachweisen lassen. Hier und in der Analyse konkreter Aufgabenbearbeitungen, die näheren Aufschluss über tatsächlich relevante sprachliche wie konzeptuelle Hürden geben könnten, wird von uns weiterer dringender Forschungsbedarf gesehen.

## Literatur

- APA. (2018). Mathe-Zentralmatura soll „verständlicher“ werden. *Salzburger Nachrichten*, (15.11.2018). Zugriff unter <http://bit.ly/2IBdQqz>
- Bamberger, R. & Vanecek, E. (1984). *Lesen-Verstehen-Lernen-Schreiben: Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Wien: Jugend und Volk.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S. & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774. doi:10.21105/joss.00774
- bmbwf. (2019). *Die standardisierte schriftliche Reifepfprüfung in Mathematik: Inhaltliche und organisatorische Grundlagen zur Sicherung mathematischer Grundkompetenzen*. Wien. Zugriff unter <http://bit.ly/2IBdAI0>
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (1997). *Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“*. Materialien zur Bildungsplanung und zur Forschungsförderung. Bonn: BLK, Geschäftsstelle. Zugriff unter <http://www.blk-bonn.de/papers/heft60.pdf>
- Chi, M. T. H. (1987). Representing Knowledge and Metaknowledge: Implications for Interpreting Metamemory Research. In F. Weinert & R. Kluwe (Hrsg.), *Metacognition, motivation, and understanding*. Hillsdale, NJ: Erlbaum.
- Egyed, M.-T. & Mittelstaedt, K. (2018). Warum Schüler bei der Mathe-Matura schlechter abschneiden. *Der Standard*, (28.05.2018). Zugriff unter <http://bit.ly/2UoiLNG>
- Glaboniat, M., Perlmann-Balme, M. & Studer, T. (2013). *Zertifikat B1: Deutschprüfung für Jugendliche und Erwachsene : Prüfungsziele, Testbeschreibung*. Ismaning: Hueber Verlag. Zugriff unter <http://bit.ly/2ZkyAsm>
- Gürsoy, E. (2014). *Kohäsion und Kohärenz in mathematischen Prüfungstexten türkisch-deutschsprachiger Schülerinnen und Schüler: Eine multiperspektivische Untersuchung*. Münster: Waxmann.
- Gürsoy, E., Benholz, C., Renk, N., Prediger, S. & Büchler, A. (2013). Erlös = Erlösung? Sprachliche und konzeptuelle Hürden in Prüfungsaufgaben zur Mathematik. *Deutsch als Zweitsprache*, (1), 14–24.
- Kassambara, A. (2018). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.2. Zugriff unter <http://bit.ly/2UrTgLn>
- Leisen, J. (2011). Sprachsensibler Fachunterricht. Ein Ansatz zur Sprachförderung im mathematisch-naturwissenschaftlichen Unterricht. In S. Prediger & E. Özdil (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit* (S. 143–162). Münster: Waxmann.
- Leiß, D., Domenech, M., Ehmke, T. & Schwippert, K. (2017). Schwer – schwierig – diffizil: Zum Einfluss sprachlicher Komplexität von Aufgaben auf fachliche Leistungen in der Sekundarstufe I. In D. Leiß, M. Hagena, A. Neumann & K. Schwippert (Hrsg.), *Mathematik und Sprache. Empirischer Forschungsstand und unterrichtliche Herausforderungen* (S. 99–125). Münster: Waxmann.
- Maier, H. (1995). Verstehen von Lehrerinstruktionen und -erklärungen durch Schüler im Mathematikunterricht. Unveröffentlichter Projektbericht. Regensburg.

- Maier, H. & Steinbring, H. (1998). Begriffsbildung im alltäglichen Mathematikunterricht – Darstellung und Vergleich zweier Theorieansätze zur Analyse von Verstehensprozessen. *Journal für Mathematik-Didaktik*, 19(4), 292–329.
- Neubrand, M., Biehler, R., Blum, W., Cohors-Fresenborg, E., Flade, L. & Knoche, N. (2004). Der Prozess der Itementwicklung bei der nationalen Ergänzungsuntersuchung von PISA 2000: Vom theoretischen Rahmen zu den konkreten Aufgaben. In M. Neubrand (Hrsg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland* (S. 31–50). Wiesbaden: VS Verlag für Sozialwissenschaften.
- OECD. (2016). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy*. PISA. Paris: OECD Publishing. Zugriff unter <http://bit.ly/2ZnUHhS>
- Peschek, W. (2011). Sicherung mathematischer Grundkompetenzen am Beispiel des österreichischen Zentralabiturs. In M. Helmerich, K. Lengnink, G. Nickel & M. Rathgeb (Hrsg.), *Mathematik Verstehen* (S. 211–220). Wiesbaden: Vieweg+Teubner.
- Plath, J. (2019). Verstehensprozesse bei der Bearbeitung realitätsbezogener Mathematikaufgaben: Klassische Textaufgaben vs. Zeitungstexte. *Journal für Mathematik-Didaktik*. doi:10.1007/s13138-019-00148-w
- Prediger, S., Wilhelm, N., Büchter, A., Gürsoy, E. & Benholz, C. (2015). Sprachkompetenz und Mathematikleistung – Empirische Untersuchung sprachlich bedingter Hürden in den Zentralen Prüfungen 10. *Journal für Mathematik-Didaktik*, 36(1), 77–104.
- Prediger, S. & Wittmann, G. (2009). Aus Fehlern lernen – (wie) ist das möglich? *Praxis der Mathematik in der Schule*, 51(27), 1–8.
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Zugriff unter <https://www.R-project.org/>
- Reusser, K. (1997). Erwerb mathematischer Kompetenzen: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 141–155). Weinheim: Beltz Psychologie-Verl.-Union.
- Rieger, L. & APA. (2018). Die Mathematik-Zentralmatura auf dem Prüfstand: Bisher schlechtestes Ergebnis an den BHS. Die Ergebnisse der Bundesländer im Überblick. *Kurier*, (26.06.2018). Zugriff unter <http://bit.ly/2PcxYRa>
- Schilcher, A., Röhrli, S. & Krauss, S. (2017). Sprache im Mathematikunterricht – eine Bestandsaufnahme des aktuellen didaktischen Diskurses. In D. Leiß, M. Hagena, A. Neumann & K. Schwippert (Hrsg.), *Mathematik und Sprache* (S. 11–43). Münster: Waxmann.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*. Manchester, UK. Zugriff unter <http://bit.ly/2vc7ANS>
- Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A. & Kliegl, R. (2015). childLex: A lexical database of German read by children. *Behavior research methods*, 47(4), 1085–1094. doi:10.3758/s13428-014-0528-1
- Schukajlow, S. & Leiß, D. (2008). Textverstehen als Voraussetzung für erfolgreiches mathematisches Modellieren – Ergebnisse aus dem DISUM-Projekt. In *Beiträge zum Mathematikunterricht online. Vorträge auf der 42. Tagung für Didaktik der Mathematik. Jahrestagung der Gesellschaft für Didaktik der Mathematik vom 13.3. bis 18.3.2008 in Budapest*. (4 S.). Zugriff unter <http://bit.ly/2KHgMoa>
- Siller, H.-S., Bruder, R., Hascher, T., Linnemann, T., Steinfeld, J. & Sattlberger, E. (2016). Kompetenzstufenmodell zu Reifeprüfungsaufgaben und deren Eignung für einen kompetenzorientierten Mathematikunterricht. In S. Keller & C. Reintjes (Hrsg.), *Aufgaben als Schlüssel zur Kompetenz* (S. 371–388). Münster: Waxmann.
- Skemp, R. R. (1976). Relational Understanding and Instrumental Understanding. *Mathematics Teaching*, (77), 20–26. Zugriff unter <https://goo.gl/reXJTJ>
- Ufer, S., Reiss, K. & Mehringer, V. (2013). Sprachstand, soziale Herkunft und Bilingualität: Effekte auf Facetten mathematischer Kompetenz. In M. Becker-Mrotzek, K. Schramm, E. Thürmann & H. J. Vollmer (Hrsg.), *Sprache im Fach* (S. 185–201). Münster: Waxmann Verlag GmbH.

- Weis, I. (2013). Wie viel Sprache hat Mathematik in der Grundschule? Über die Notwendigkeit der Verbindung von sprachlichem und fachlichem Lernen im Mathematikunterricht der Grundschule. Duisburg. Zugriff unter <http://bit.ly/2mQfagA>
- Wild, J. & Pissarek, M. (o. J.). Ratte. Regensburger Analysetool für Texte. Zugriff unter <http://bit.ly/2UpRUBb>
- Wilhelm, N. (2016). *Zusammenhänge zwischen Sprachkompetenz und Bearbeitung mathematischer Textaufgaben: Quantitative und qualitative Analysen sprachlicher und konzeptueller Hürden*. Wiesbaden: Springer.

#### Anschrift des korrespondierenden Autors

Andreas Vohns

Institut für Didaktik der Mathematik  
Universität Klagenfurt  
Sterneckstraße 15  
9020 Klagenfurt  
Österreich  
[andreas.vohns@aau.at](mailto:andreas.vohns@aau.at)