

Eine diachronische Analyse der Inhaltsvalidität des Prüfungsteils ‚Leseverstehen‘ in der ‚Prüfung für das Germanistik-Grundstudium‘ (PGG)

Zhu Jia und Kong Deming
(Nanjing)

Kurzzusammenfassung: Der vorliegende Beitrag untersucht im Rahmen des Modells von Bachman und Palmer (1996) die Inhaltsvalidität der Testaufgaben im Abschnitt ‚Leseverstehen‘ in der PGG von 2012 bis 2018. Dabei handelt es sich um die Merkmale des Textinputs, welcher durch Textlänge, Thema, Gattung und Lesbarkeit charakterisiert ist, sowie die der erwarteten Testantwort, welche Aufgabentypen und geforderte Qualifikationen umfasst. Nach der Untersuchung wird festgestellt, dass außer der Lesbarkeit die Merkmale Textlänge, Themen und Textsorten des Prüfungsabschnitts ‚Leseverstehen‘ in der PGG hauptsächlich der Prüfungsordnung entsprechen. Unter den Aufgabentypen tritt das Lückenfüllen tendenziell in den Vordergrund. Der Umgang mit expliziten Informationen und die Inferenz aus dem Kontext gelten als wichtige Fähigkeit in der Überprüfung des Leseverständnisses. Zu beobachten ist zudem die Korrektheitsproportion der Antworten, wobei eine statistische Analyse durchgeführt wird, um einige Vorschläge für die Lehre wie auch für den Testentwurf des Leseverstehens geben zu können.

1. Einleitung

Unter der Leitung des Anlenkungskomitees für Germanistik des Bildungsministeriums wird die ‚Prüfung für das Germanistik-Grundstudium‘ (PGG) von der PGG-Arbeitsgruppe durchgeführt und gilt als staatlicher Standardtest für Studierende der Germanistik im vierten Semester. 1997 erließ die PGG-Arbeitsgruppe die Prüfungsordnung für das Germanistik-Grundstudium, seit 1998 findet die Prüfung einmal im Jahr statt. Die Zahl der teilnehmenden Hochschulen sowie Prüflinge nimmt seitdem von Jahr zu Jahr zu. Nach Angaben des Prüfungszentrums lag die Zahl der Hochschulen, die 2018 an der PGG teilnahmen, bei 113, und die Zahl Teilnehmenden ist im Vergleich zu 1998 auf das Fünfzehnfache angestiegen.

Um sich den Ausbildungszielen und den Anforderungen für Germanisten in der neuen Zeit anzupassen, überarbeitete das Prüfungszentrum 2012 die Prüfungsordnung, wodurch sich einige Elemente der Prüfung änderten. Das Leseverstehen ist in zwei Teile gegliedert und die Bearbeitungszeit dafür erhöhte sich von 40 Minuten auf 50 Minuten. Die Anzahl der Items bzw. die Punkte bleiben jeweils 20, aber die Aufgaben beim Leseverstehen haben sich

teilweise geändert. Der erste Teil besteht aus vier kurzen Abschnitten. Nach dem Lesen soll man die passenden Überschriften dazu auswählen. Die Richtig / Falsch-Fragen, die früher im Teil „Leseverstehen“ vorkamen, wurden durch Mehrfachfragen ersetzt.

Die Gesamtzahl der Studenten, welche die PGG seit der Veränderung der Prüfungsordnung im Jahre 2012 abgelegt haben, hat die Marke von 40.000 längst überschritten. Wie hoch ist die Inhaltsvalidität des Leseverstehens in der PGG? Können Prüflinge die Leseaufgaben innerhalb einer vorgegebenen Zeit erfüllen? Welche Lesefähigkeiten werden beim Lösen der Aufgaben verwendet? Ausgehend von diesen Fragen wird ein Korpus von PGG-Leseteilen von 2012 bis 2018 erstellt und die Validität des Leseverstehens mit Hilfe des Modells von Bachman und Palmer (1996) analysiert.¹ Einbezogen ist eine statistische Analyse der Antworten, um die auf die Richtigkeit beim Leseverstehen Einfluss nehmenden Faktoren herauszufinden. Ziel ist, hilfreiche Anmerkungen für die Prüfungserstellung herauszuarbeiten.

2. Forschungsstand

Im Bereich der Sprachtests tauchten in den 90er-Jahren eine Reihe neuer Theorien auf, unter anderem das erste Modell der kommunikativen Sprachkompetenz (engl. *communicative language ability* – CLA) von Bachman.² Nach seiner Darstellung setzt sich die kommunikative Sprachkompetenz aus dem Sprachwissen und der Fähigkeit, das Sprachwissen in angemessene Sprachverwendung umzusetzen, zusammen. Die CLA lässt sich in drei Komponenten untergliedern, nämlich in Sprachkompetenz, strategische Kompetenz und psycho-physiologische Mechanismen. Dennoch konzentriert sich das modifizierte Modell der kommunikativen Sprachkompetenz auf das Sprachwissen (Umbenennung der Komponente ‚Sprachkompetenz‘) und die strategische Kompetenz.

Im Allgemeinen kann ein kommunikativer Sprachtest als Rückmeldung für Lernende, als ein diagnostisches Instrument und zur gesellschaftlichen Selektion und Chancenzuweisung dienen.³ Um nützliche kommunikative Sprachtests zu entwickeln, ist es zielführend, Teilnehmer sowie den Sprachgebrauch zu berücksichtigen. Unter dem Gebrauch der Zielsprache (engl. *target language use* – TLU) wird „eine Reihe von Sprachgebrauchsaufgaben, welchen die Prüfungsteilnehmer außerhalb des Tests wohl begegnen wer-

¹ Lyle F. Bachman / Adrian S. Palmer, *Language Testing in Practice. Designing and Developing Useful Language Tests*. Oxford 1996.

² Lyle F. Bachman, *Fundamental Considerations in Language Testing*. Oxford 1990, S. 84.

³ Vgl. Dietmar Rösler, *Deutsch als Fremdsprache. Eine Einführung*. Stuttgart 2012, S. 255.

den“ verstanden.⁴ Nach dieser Definition ist die Sprachgebrauchsaufgabe eine Aktivität, in der Individuen die Sprache zu einem bestimmten Ziel in einer bestimmten Situation verwenden. Ein kommunikativer Sprachtest soll aus mehreren Sprachgebrauchsaufgaben bestehen. Nach Bachman und Palmer liegt der Schlüssel zur Testentwicklung darin, dass die Testaufgaben entsprechend den Gebrauchsaufgaben der Zielsprache entworfen werden.⁵

Um sowohl Gebrauchsaufgaben der Zielsprache als auch Testaufgaben adäquat zu beschreiben und der Testentwicklung eine Grundlage zu geben, schufen Bachman und Palmer für die Beschreibung eines Sprachtests einen Rahmen für die Aufgabenmerkmale, welcher fünf Aspekte umfasst:

- a. Rahmenbedingung (engl. setting): Physische Merkmale in einem Test inklusive des Orts, der Anlage, Materialien und weiteres; Testpersonen; Zeitpunkt des Tests.
- b. Testdarstellung (engl. test rubrics): Anleitung; Struktur der Prüfungsteile sowie Testaufgaben; Zeitdauer; Bewertungsmethode.
- c. Input (engl. input): Länge, Geschwindigkeit des Inputs; sprachliche Merkmale; thematische Merkmale. Zusammengefasst geht es dabei um Materialien, die man im Test bearbeitet und auf die man reagiert.
- d. Erwartete Testantwort (engl. expected response): Form, Art, Geschwindigkeit der Testantwort.
- e. Beziehung zwischen Input und erwarteter Testantwort (engl. relationship between input and response): Reaktivität, Umfang und Unmittelbarkeit der Beziehung.⁶

Basierend auf den Testaufgabenmerkmalen haben sich mit der inhaltlichen Validität des englischen Leseverstehens bislang viele chinesische Untersuchungen auseinandergesetzt, denen hauptsächlich der TEM 4 (Test for English Majors, Bd. 4) und die englische Hochschulaufnahmeprüfung zugrunde liegen. In Kombination mit dem englischen Lehrplan und der Prüfungsordnung führte Hou (2012) eine Untersuchung zur inhaltlichen Validität des Leseverstehens in TEM 4 von 2005 bis 2011 durch und stellte fest, dass die inhaltliche Validität des TEM 4 im untersuchten Zeitraum hoch sei, der Inhalt des Tests im Wesentlichen den Anforderungen des Lehrplans entspreche und eine gute Kontinuität zeige.⁷ Nach einer experimentellen Studie zur Validität der Antworten im Teil ‚Leseverstehen‘ von TEM 4 kamen Zou et al. (2002) zu dem Ergebnis, dass Aufgaben wie Paraphrase und Erklärung die Lernenden am besten dazu ermutigen könnten, Lesestrategien anzuwen-

⁴ Ebenda, S. 256.

⁵ Vgl. Lyle F. Bachman / Adrian S. Palmer, a. a. O., S. 45.

⁶ Vgl. ebenda, S. 47-57.

⁷ Hou Yanping, Evaluating the Content Validity of the Innovated TEM 4 Reading in Recent Seven Years, in: Journal of Hebei University (Philosophy and Social Science), 37 / 4 (2012), S. 143-147.

den.⁸ Deren Ergebnisse überprüfte Yang (2011) und ergänzte, dass der Lesetest im Wesentlichen die Erwartungen erfülle, es sei jedoch anzumerken, dass einige Strategien wie z. B. das Ausarbeiten einer Gliederung, die Umwandlung von Informationen in eine bildliche Darstellung oder umgekehrt kaum verwendet wurden.⁹ Liu und Hu (2018) analysierten die inhaltliche Validität des Leseverstehentests in der TEM 4 Musterprüfung 2016, insbesondere die der Multiple-Choice-Fragen bzw. der Kurzantworten, was belege, dass die Inhaltsvalidität in diesem Jahr hoch sei und diejenige von Aufgaben, die eine kurze Antwort verlangten, höher als die von Multiple-Choice-Fragen sei.¹⁰ Nach den Untersuchungen bezüglich der staatlichen englischen Hochschulaufnahmeprüfung¹¹ sowie der in der Provinz Anhui¹², der Provinz Jiangsu¹³ und der Provinz Zhejiang¹⁴ lässt sich festhalten, dass das Leseverstehen in der Hochschulaufnahmeprüfung dem Lehrplan entspricht und über eine hohe Inhaltsvalidität verfügt.

Zusammenfassend lässt sich feststellen, dass der Validität des Leseverstehensinhalts in dem englischen TEM 4 und der Hochschulaufnahmeprüfung umfassend Aufmerksamkeit geschenkt wird, während weniger Studien zu Prüfungen anderer Fremdsprachen vorliegen, deren Inhaltsvalidität überprüft werden soll. Ebenfalls ist die Angemessenheit des Leseverstehensinhalts in der PGG, die als einzige staatliche Prüfung für Germanistikstudierende gilt, zu untersuchen.

⁸ Zou Shen et al., The Relationship Between Item Type, Strategy and Score in Reading Test. A Study on the Validity of Reading Items in TEM4, in: *Foreign Languages and Their Teaching*, 5 (2002), S. 19-22.

⁹ Yang Weijian, Verification of Response Validity of Reading Comprehension in TEM 4, in: *Foreign Language Education*, 32 / 6 (2011), S. 53-56.

¹⁰ Liu Baoquan / Hu Fangfang, A Study on the Validity of Reading Comprehension in TEM4. The Case of Reading Comprehension in TEM4 2016 Sample Test, in: *Technology Enhanced Foreign Languages*, 179 (2018), S. 70-75.

¹¹ Huang Juan, Research on Content Validity of Reading Comprehension of National Matriculation English Test in 2017, in: *Educational Measurement and Evaluation*, 4 (2018), S. 143-147.

¹² He Jing, A Diachronic Study of Content Validity of Reading Comprehension in NMET of Anhui Province, in: *Journal of Yangtze Normal University*, 28 / 10 (2012), S. 126-130, S. 140.

¹³ Liu Ying u. a., A Diachronic Study of Content Validity of Reading Comprehension in NMET of Jiangsu Province, in: *Educational Measurement and Evaluation*, 2 (2016), S. 23-29.

¹⁴ Tan Xiaohui / Zhou Zhiyuan, An Analysis of Content Validity of Reading Comprehension in NMET of Zhejiang Province (2012-2014), in: *Elementary Education Studies*, 5 (2015), S. 62-66.

3. Methodisches Vorgehen

Prüfungen sind wichtige Hilfsmittel zur Messung des Sprachniveaus, zur Bewertung der Unterrichtsqualität und zur Evaluation des Rahmenplans, welche für die Germanistik an den Hochschulen durch die Inhaltsvalidität der staatlichen PGG bestätigt werden. Beim Prozess der Validierung steht der Begriff „Validität“ im Vordergrund. Zur Validität gehört, dass „die Prüfung adäquat das misst, was sie messen soll, und die Interpretation der Ergebnisse, die mit dem Verwendungszweck der Prüfung verbundene Entscheidungen rechtfertigt.“¹⁵ Aus traditioneller Sicht gliedert sich die Validität in vielerlei Aspekte. Einen davon bildet die Inhaltsvalidität, die „durch umfangreiche Sprachtests erhöht werden kann.“ Testaufgaben mit hoher Inhaltsvalidität „stellen eine repräsentative Stichprobe aller Aufgaben aus dem zu prüfenden Sprachbereich da.“¹⁶

Der Untersuchung liegen die Items zum Leseverstehen in der PGG von 2012 bis 2018 zu Grunde. Mit Hilfe der Statistik-Software R wird die Inhaltsvalidität des Prüfungsteils „Leseverstehen“ in der PGG quantitativ untersucht.

Das Leseverstehen folgt in der schriftlichen PGG auf das Diktat und das Hörverstehen und ist in zwei Teile unterteilt. Zum Lesen werden sieben Materialien angeboten. Im ersten Teil werden die Prüflinge aufgefordert, zweimal zwei Absätze oder kurze Texte zu lesen und aus je vier Optionen passende Überschriften auszuwählen. Der zweite Teil besteht aus drei relativ vollständigen Texten, nach denen die Teilnehmenden 16 Single-Choice-Fragen lösen müssen.

Das Korpus umfasst 49 Texte und 140 Items, die von 2012 bis 2018 in der PGG erschienen. Anhand der Anforderungen und Aufgaben des Abschnitts „Leseverstehen“ in der PGG setzt sich der Beitrag mit den Merkmalen der Eingabe und der erwarteten Antwort (Fragetyp, Antwortfähigkeit) sowie mit der Korrektheit der Antworten auseinander.

4. Analyse der Inhaltsvalidität

4.1. Eingabemerkmale

Hinsichtlich der Eingabemerkmale werden Lesetexte in der PGG unter den vier folgenden Aspekten berücksichtigt, nämlich Textlänge, Thema, Textsorte und Lesbarkeit.

¹⁵ Gabriele Kecker, Validierung von Sprachprüfungen. Die Zuordnung des TestDaF zum Gemeinsamen Europäischen Referenzrahmen für Sprachen. Frankfurt a. M. 2011, S. 16.

¹⁶ Dietmar Rösler, Deutsch als Fremdsprache. Eine Einführung. Stuttgart 2012, S. 259.

4.1.1. Textlänge

Teil 1 in Tab. 1 bezieht sich auf die Wortanzahl des ersten Teils im „Leseverstehen“. Teil 2 in der Tabelle entspricht der Wortanzahl des zweiten Teils vom „Leseverstehen“ in der Prüfung. Bei der Wortanzahl von ‚Aufgaben‘ geht es um die Wörter der zu lösenden Aufgaben nach den Lesetexten. In Bezug auf die Gesamtzahl der verwendeten Wörter im Aufgabenteil ‚Leseverstehen‘ der PGG von 2012 bis 2018 liegt das Jahr 2013 an der Spitze, während die Kandidaten des Jahres 2015 die niedrigste Wortanzahl zu lesen hatten.

JAHR WORTANZAHL in	2012	2013	2014	2015	2016	2017	2018
Teil 1	426	373	364	342	334	311	364
Teil 2	916	1004	851	747	822	803	785
Aufgaben	501	623	452	541	702	638	590
Gesamtzahl	1843	2000	1667	1630	1858	1752	1739

Tab.1: Wortanzahl der Lesetexte und Aufgaben in der PGG

Gemäß der Prüfungsordnung besteht die Anforderung an die Lesefähigkeit darin, kurze Artikel und einfache literarische Texte zu überfliegen und den Hauptinhalt erfassen zu können. Die Lesegeschwindigkeit beim Überfliegen beträgt ungefähr 70 Wörter pro Minute. Beim detaillierten Lesen reichen ungefähr 50 Wörter pro Minute und man soll in der Lage sein, den Hauptinhalt und wichtige Details eines Textes zu verstehen.¹⁷

Die Zeitdauer des Leseverstehens in der PGG ist auf 50 Minuten begrenzt. Wenn alle Texte auf umfangreiches Lesen berechnet würden, könnte die Gesamtzahl der Wörter 3500 (70×50) erreichen. Rechnet man mit detailliertem Lesen, liegt die Summe unter 2500 (50×50) Wörtern. Aus Tab.1 ist ersichtlich, dass die Anzahl der Wörter in den sieben Jahren die Durchschnittszahl nicht überschritten hat. Eine gewisse Erhöhung bei der Wortanzahl im Abschnitt „Leseverstehen“ ist zu erwarten.

¹⁷ Vgl. Prüfungsordnungen für das Germanistik-Grundstudium und -Hauptstudium im Hochschulwesen Chinas. Shanghai 2013, S. 5.

4.1.2. Themen

Nach der Prüfungsordnung stammen die Themen der Lesematerialien hauptsächlich aus dem alltäglichen Leben und sozialen Angelegenheiten und sollen zudem den Inhalt des Leseunterrichts gemäß dem Lehrplan widerspiegeln.¹⁸ Die Themen der Lesematerialien in der PGG von 2012 bis 2018 sind nach Textinhalten klassifiziert und in Tab. 2 dargestellt:

JAHR THEMA	2012	2013	2014	2015	2016	2017	2018
Gesundheit und Medizin	2	1	1	2	1	0	0
Urlaub und Verkehr	2	2	2	1	1	0	1
Studium und Studienleben	1	0	0	1	0	1	0
Soziale Themen	0	3	3	2	4	4	4
Landeskunde und Kultur	1	0	1	1	1	1	1
Persönlich- keiten	1	1	0	0	0	1	1

Tab. 2: Themen der Lesetexte in der PGG

Aus Tabelle 2 wird ersichtlich, dass sich die Themen in der Regel an der Prüfungsordnung orientieren. ‚Gesundheit und Medizin‘, ‚Urlaub und Verkehr‘, ‚Studium und Studienleben‘ gehören zum Alltagsleben. Außer 2012 traten soziale Themen im Teil ‚Leseverstehen‘ jedes Jahr in der Prüfung auf. In vier Jahren beziehen sich mehr als die Hälfte der Lesetexte in der Prüfung auf aktuelle soziale Themen. Bei ‚Landeskunde und Kultur‘ geht es hauptsächlich um die deutsche Landeskunde und Kultur, wie beispielsweise die Vorstellung einer deutschen Stadt. Was ‚Persönlichkeiten‘ angeht, werden sowohl gegenwärtige als auch historische vorgestellt.

Die Themenauswahl der Lesematerialien zeigt eine relativ große Vielfalt, welche eng mit dem Alltag und Studium zusammenhängt. Je vielfältiger die Themen sind, desto besser können sie dazu beitragen, die im Studium er-

¹⁸ Vgl. ebenda, S. 12.

worbene Lesefähigkeit zu prüfen. Deswegen ist es empfehlenswert, dass sich die Lesematerialien in der Prüfung auf mehr Themen beziehen.

4.1.3. Textsorten

Laut der Prüfungsordnung werden Sachtexte und literarische Texte erfasst. Unter die Sachtexte fallen Beschreibung, Bericht, Nachricht, Erörterung, Interview u.a. Unter literarischen Gattungen versteht man hier Kurzgeschichte, Geschichte, Märchen.¹⁹

JAHR TEXTSORTE	2012	2013	2014	2015	2016	2017	2018
Anzeige	1	3	1	0	2	2	2
Nachricht	3	0	2	0	0	0	0
Beschreibung	1	1	4	7	3	4	3
Forschungsbericht	1	2	0	0	2	0	1
Interview	1	0	0	0	0	0	1
Literarische Texte	0	1	0	0	0	1	0

Tab. 3: Textsorten der Lesetexte in der PGG

Literarische Texte sind im Allgemeinen lang, deswegen finden sich diese von 2012 bis 2018 aufgrund der Testform und -zeit lediglich zweimal. Dabei handelt es sich um Anekdoten, alle anderen sind Sachtexte (s. Tab. 3).

2012 war das erste Jahr, in dem die neue Prüfungsordnung in Kraft war. Die Verteilung der Textsorten in diesem Jahr ist ebenfalls am ausgewogensten. Außer literarischen Texten erscheinen alle fünf Textsorten von Sachtexten. Im Gegenteil dazu ist die Themenaufstellung im Jahr 2015 einförmig. Alle sieben Texte sind der Beschreibung unterzuordnen.

Zusammenfassend lässt sich feststellen, dass literarische Texte aufgrund der Länge kaum im Prüfungsteil ‚Leseverstehen‘ vorkommen. Sachtexte eignen sich besser aufgrund der zeitlichen Begrenzung der Prüfung, und auch für die Prüfung der Lesegeschwindigkeit und der Fähigkeit zum Herausfiltern der Hauptinformationen, weshalb Sachtexte in der PGG bevorzugt werden. Allerdings ist es nötig, beim Entwurf der Prüfung einen größeren

¹⁹ Vgl. ebenda, S. 12.

Umfang der Textsorten zu berücksichtigen, damit der Umgang der Prüflinge mit verschiedenen Themen festgestellt werden kann.

4.1.4. Lesbarkeit

Die Lesbarkeit dient dazu, die Leseschwierigkeiten eines Textes zu messen. Das Forschungsinteresse daran entstand früh in Europa. Bereits in den 1920er Jahren begannen westliche Wissenschaftler, den Schwierigkeitsgrad eines Textes durch Messung der Lesbarkeit des Vokabulars, der Wort- und Satzlänge auszuloten und entwickelten die erste Lesbarkeitsformel.²⁰ Daraus ergaben sich viele berühmte Formeln wie z. B. die SMOG-Formel²¹, die Fry-Formel²², die Flesch-Formel²³ u. a. Unter diesen ist die Flesch-Lesbarkeitsformel am gebräuchlichsten, die aus der Berechnung des Wortschatzes und der syntaktischen Schwierigkeit des Textes besteht. Die Formel lautet wie folgt:

$$RE \text{ (Reading Ease)} = 206,84 - 0,85wl - 1,02sl \text{ (} wl \text{ bezieht sich auf die Anzahl der Silben pro 100 Wörter, } sl \text{ bezieht sich auf die durchschnittliche Anzahl der Wörter pro Satz)}$$

Nach den RE-Werten bildet sich eine Lesbarkeitsskala. Danach ist der Schwierigkeitsgrad eines Textes

- zwischen 30 und 50 „schwierig“
- zwischen 50 und 60 „relativ schwierig“
- zwischen 60 und 70 „standardmäßig“
- zwischen 70 und 80 „relativ einfach“
- zwischen 80 und 90 „einfach“

Errechnet nach der Flesch-Formel wird die Lesbarkeit von zwei Teilen sowie die Gesamtlesbarkeit des PGG-Lesetests von 2012 bis 2018 in Tabelle 4 dargestellt:

²⁰ Bertha A. Lively / Sidney L. Pressey, A Method for Measuring the „Vocabulary Burden“ of Textbooks, in: Educational Administration and Supervision, 9 (1923), S. 389-398.

²¹ G. H. McLaughlin, SMOG Grading - a New Readability Formula, in: Journal of reading, 12 / 8 (1969), S. 639-646.

²² Edward Fry, A Readability Formula That Saves Time, in: Journal of Reading, 11 / 7 (1968), S. 513-516, S. 575-578.

²³ Rudolph Flesch, A New Readability Yardstick, in: Journal of Applied Psychology, 32 / 3 (1948), S. 221-233.

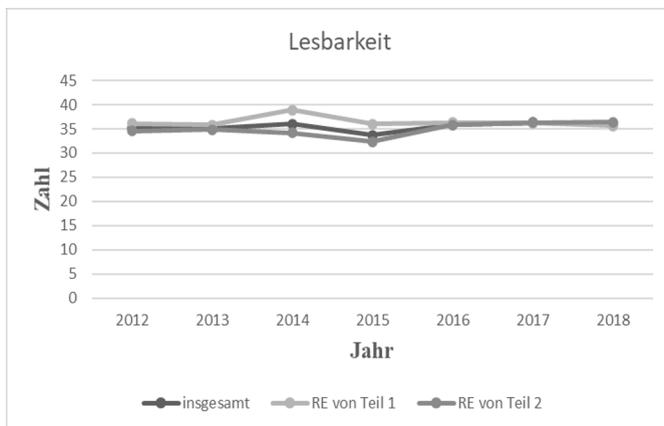


Abb. 1: Lesbarkeit in der PGG

Abb.1 zeigt, dass die Lesbarkeit der Texte im untersuchten Zeitraum zwischen 30 und 40 variiert, was nach der Klassifikation als „schwierig“ zu bezeichnen ist. Allerdings sind die Texte des ersten Teils im Jahr 2014 deutlich einfacher zu lesen als in den anderen Jahren, während die Gesamtlesbarkeit und die des zweiten Teils im Jahr 2015 niedriger ist als in den anderen Jahren. Im Allgemeinen weist die Lesbarkeit in dem Zeitraum keinen signifikanten Unterschied auf. Die Lesbarkeit hängt jedoch auch von Faktoren wie Thema, Gattung, Wortschatz, grammatikalischer Struktur u. ä. ab. Die Einschränkung der Lesbarkeitsformel von Flesch besteht darin, nur die Anzahl der Silben und die durchschnittliche Anzahl der Wörter zu berücksichtigen.

4.2. Erwartete Antwortmerkmale

Die Untersuchung der erwarteten Antwortmerkmale spiegelt sich hauptsächlich in der Analyse von Testaufgaben wider, einschließlich der Aufgabentypen und der zu prüfenden Lesefähigkeit.

4.2.1. Aufgabentypen

Der Prüfungsteil „Leseverstehen“ umfasst 20 Single-Choice-Fragen, bei denen gefordert wird, für die ersten vier Texte je eine Überschrift aus vier Angeboten auszuwählen, und bei den anderen Fragen eine richtige Lösung aus drei möglichen Antworten zu bestimmen. Von 2012 bis 2018 werden insgesamt 140 Aufgaben erledigt. Der Form der Fragen zufolge repräsentieren die Aufgaben vier Typen: Überschrift, Frage nach Information, Auswahl richtiger oder falscher Aussage und Lückenfüllen.

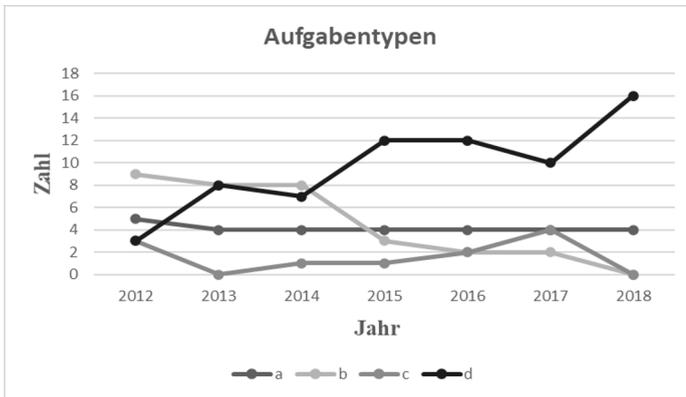


Abb. 2: Aufgabentypen (a. Überschrift; b. Ergänzungsfrage; c. Richtige/falsche Aussage; d. Lückenfüllen)

Infolge der Aufgabenverteilung ist es üblich, dass jedes Jahr die ersten vier Aufgaben im ersten Teil vom „Leseverstehen“ zu Typ a zählen. Im untersuchten Zeitraum offenbart lediglich 2012 eine Ausnahme, denn in diesem Jahr erscheint dieser Typ außer im ersten Teil erneut im zweiten Teil, wobei die Frage nach einer passenden Überschrift zu Text B gestellt wird (Item 30). Typ c macht durchschnittlich den geringsten Anteil aus, 2013 und 2018 kommt er überhaupt nicht vor. Bei diesem Typ geht es darum, eine richtige oder eine falsche Aussage aus drei Formulierungen auszuwählen, z. B. „Welche Aussage ist falsch?“ (Item 26 von 2012), „Welche Aussage ist richtig?“ (Item 25 von 2017).

Typ b ist durch Ergänzungsfragen gekennzeichnet, bei denen nach einer bestimmten Information im Lesetext gefragt wird, beispielsweise „Wie viele Sprachen gelten in Schleswig-Holstein als offizielle Landessprachen?“ (Item 15 von 2012). Abb. 2 zeigt in der PGG eine signifikant sinkende Tendenz von Typ b an, d. h. von 2012 fast die Hälfte der gesamten Aufgaben auf keine einzige im Jahr 2018. Im Gegensatz zu dem Rückgang von Typ b offenbart sich Jahr für Jahr eine Bevorzugung des Lückenfüllens. Die Aufgaben des Teils „Leseverstehen“ zeichnen sich im Jahr 2018 dadurch aus, dass im zweiten Teil ausschließlich Typ d vorkommt.

4.2.2. Lesefähigkeit

Nach der Prüfungsordnung werden die Leseverstehensaufgaben in zwei Disziplinen unterteilt: Fragen nach angegebener Information und nach semantischer Bedeutung. Die detaillierten Informationsfragen werden weiter in zwei Typen unterteilt. Bei dem einen Typ geht es um das Verstehen eines Ausdrucks. Die Lösung kann gefunden werden, indem die Kandidaten die entsprechenden Wörter oder Ausdrücke im gelesenen Text gemäß den Schlüsselwörtern in Fragen finden. Bei dem anderen Typ ist die Lösung im-

plizit. Es ist schwierig, die Lösung unmittelbar aus den Lesetexten herauszuarbeiten. Die Antworten setzen das richtige Verstehen des Textes voraus. Hinsichtlich der impliziten Kategorie handelt es sich bei der richtigen Option im Allgemeinen darum, dass die Wörter im Text durch Ausdrücke mit gleicher oder umfangreicherer semantischer Bedeutung ersetzt werden. In anderen Fällen unterscheidet sich die richtige Option deutlich von den beiden anderen. Im Gegensatz dazu weisen die Optionen zur Inferenz die folgenden Merkmale auf: (1) unvollständig; (2) anders als im Lesetext; (3) zwar im Text gefunden, aber den Fragen nicht entsprechend; (4) anscheinend dem Textinhalt ähnelnd, aber zu absolut klingend. Gemäß der obigen Erläuterung wird die geforderte Lesefähigkeit in der PGG durch die folgenden sechs Kriterien verdeutlicht:

1. Hauptinhalt des Textes zusammenfassen
2. Explizite Informationen im Text finden
3. Ausdrücke verstehen
4. Richtige oder falsche Aussage auswählen
5. Inferenz und Analyse durchführen
6. Autorenabsicht erfassen

Nach der Lesefähigkeit sind alle Aufgaben von 2012 bis 2018 in Tab. 4 klassifiziert. Wegen der Aufgabentypen machen die Items mit A jedes Jahr mindestens vier aus. 2012 ist das einzige Jahr, in dem Fragen nach dem Titel bzw. Hauptinhalt des Textes nicht nur im ersten Teil gestellt werden. Am meisten hängen die Items eng mit den Texten zusammen, während kein großes Interesse an der Fähigkeit besteht, die Autorenmeinungen zu erfassen.

JAHR KRITERIUM	2012	2013	2014	2015	2016	2017	2018
A	6	4	4	4	4	4	4
B	7	5	8	8	8	6	9
C	0	0	1	3	2	3	3
D	2	0	0	1	2	3	0
E	3	10	7	4	4	4	4
F	2	1	0	0	0	0	0

Tab. 4: Lesefähigkeit in der PPG-Prüfung

Die Auswahl eines geeigneten Titels erfordert die Fähigkeit, die Hauptideen des gesamten Artikels zu verstehen und zusammenzufassen. B-F dienen dem Testen des Verstehens des Textinhalts und der Erfassung einzelner Ausdrücke. Darunter gilt B als die einfachste Fähigkeit. Die Aufgabe ist gelöst, sobald man entsprechende Informationen oder Ausdrücke im Lesetext findet. C verlangt, dass die Kandidaten vor allem einzelne Wörter oder Sätze im Text richtig verstehen und dann ihre Alternativen aus angegebenen Optionen identifizieren. D wird als die Fähigkeit angesehen, zu beurteilen, welche Formulierung dem Textinhalt entspricht. In E muss man auf der Grundlage des Verstehens des Textes analysieren und begründen, und F fordert die Kandidaten dazu auf, die Absichten, Meinungen und Einstellungen des Autors zu erfassen.

5. Analyse der Antwortkorrektheit

Mit Hilfe der R-Software wurde eine statistische Untersuchung der Antworten in der Praxis der PPG-Prüfung durchgeführt. In den sieben Jahren finden sich insgesamt 140 Items, darunter sind 28 Überschrift-Aufgaben und 112 Single-Choice-Fragen. Die Ergebnisse lassen sich durch die folgenden Box-Plots²⁴ anschaulich darstellen, die jeweils die Korrektheit der Antworten von allen 20 Aufgaben, vom ersten sowie vom zweiten Teil des Leseverstehens in Prozent widerspiegeln:

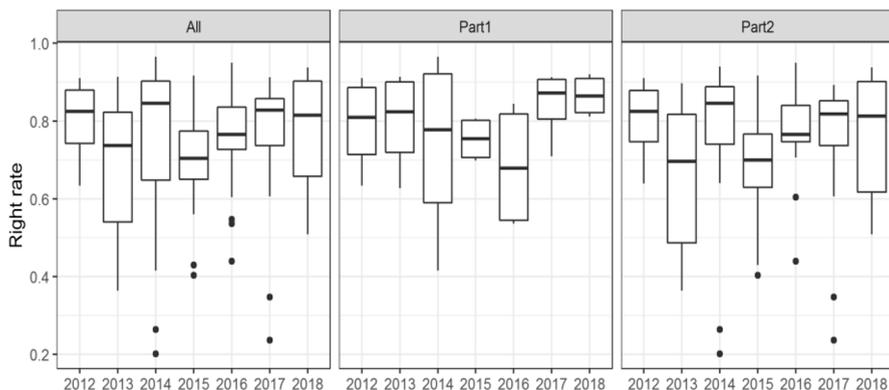


Abb. 3: Korrektheitsquote der Antworten

²⁴ Ein Box-Plot, auch Whisker-Plot genannt, zeigt die Streuung, die Zentren sowie die Distribution eines Datensatzes an. Bei einer Box-Plot-Grafik sind grundsätzlich sechs Zahlenwerte enthalten: das Maximum, das erste und dritte Quartil, der Median sowie das Minimum. Insbesondere lassen sich die ungewöhnlich niedrigen Werte veranschaulichen, die als „Ausreißer“ bezeichnet werden.

Gemäß der Verteilung der Korrektheitsrate aller Items ist die Distanz von 2015 bis 2017 gering, während die Korrektheitsraten im Jahr 2013, 2014 und 2018 auseinanderliegen. Die große Entfernung findet sich ebenfalls im ersten Teil vom Jahr 2014 und 2016. Im Vergleich dazu ist die Verteilung der Korrektheitsraten 2015, 2017 und 2018 deutlich konzentrierter. Im zweiten Teil tritt der große Abstand vor allem im Jahr 2013 und darauffolgend 2018 in den Vordergrund. Sonst ergibt sich kein auffälliger Unterschied bei der Korrektheit. Nach der Statistik vom Box-Plot sind die Werte zwischen dem Maximum und dem Minimum als regulär anzusehen. Von 2014 bis 2017 treten außer den regulären Werten einige unübersehbare Ausreißer auf.

Im Jahr 2014 beispielsweise liegt die Korrektheit des Items 21 mit 96.563% auf dem ersten Platz, während die Korrektheitsrate vom 24. Item nicht einmal 41.5% erreicht. Die Prozentzahlen der anderen drei Optionen von Item 24 bewegen sich zwischen 16% und 23%, was zeigt, dass keine von ihnen die Prüflinge definitiv überzeugt. Obwohl es bei Item 21 und 24 um die gleiche Aufgabe geht, d. h. einem Text seine passende Überschrift zuzuordnen, sind die Prozentzahlen der richtigen Antworten der beiden sehr unterschiedlich. Es mag daran liegen, dass das Stichwort „Sport“, welches mehrmals im Lesetext vorkommt, nur in der richtigen Option erscheint. Bei Item 24 kommen die Begriffe „Handys“, „Mobil telefonieren“, „Mobiltelefon“ in allen Optionen vor, was die Prüflinge leicht verwirren kann.

An der Korrektheitsquote lässt sich ablesen, dass sich die gesamte Korrektheit in den Jahren 2012, 2015, 2016, 2017 homogen verteilt. Dennoch scheinen den Prüflingen die Aufgaben von 2015 schwerer als in den anderen drei Jahren zu fallen. Im Allgemeinen bleibt die Distribution des ersten Teils normal, während die des zweiten Teils eher schief dazu liegen. Die Beziehung zwischen der Korrektheit und den vier Aufgabentypen lässt sich anhand von Abb. 4 verdeutlichen:

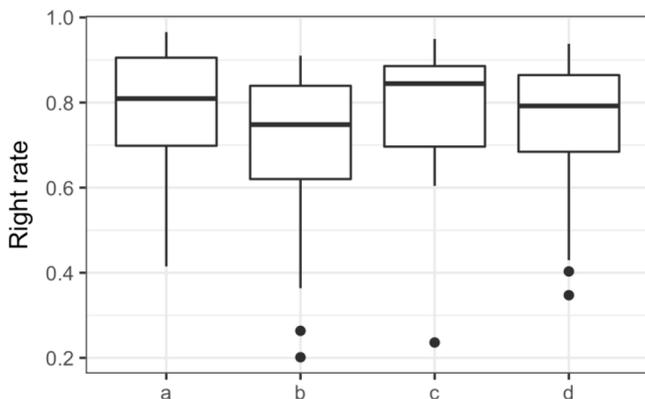


Abb. 4: Korrektheitsquote der Aufgabentypen
(a. Überschrift; b. Ergänzungsfrage; c. Richtige/falsche Aussage; d. Lückenfüllen)

Die Distribution der Korrektheit von a, b und d ist als normal anzusehen. Der Unterschied über und unter dem Median ist gering, d. h. die überdurchschnittlichen und unterdurchschnittlichen Prozentzahlen verteilen sich relativ homogen. Dagegen konzentrieren sich die höheren Korrektheitszahlen bei dem Typ c viel mehr als die Niedrigeren und sie befinden sich in einer engen Distanz, was bedeutet, dass die Prüflinge diesen Aufgabentyp besser im Griff haben.

Immerhin sind die Ausreißer auffällig, die in Typ b, c und d auftreten. Ein Beispiel findet sich bei Item 35 in der Prüfung 2014, wobei lediglich etwa 20% der Prüflinge die richtige Entscheidung trafen. Das ist eine Ergänzungsfrage nach dem Begriff „Megacity“. Der Anglizismus hilft hier beim Verstehen, jedoch nicht bei der Auswahl einer richtigen Definition.

Die Beziehung zwischen der Korrektheit und den sechs Lesefähigkeiten lässt sich durch Abb. 5 darstellen:

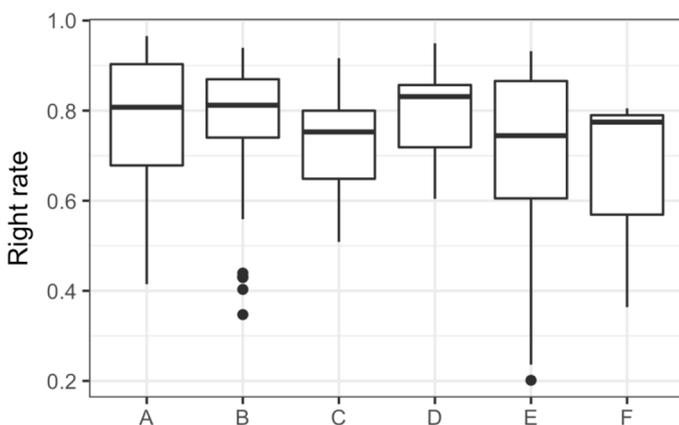


Abb. 5: Korrektheitsquote der Lesefähigkeiten

(A. Inhalt; B. Information; C. Ausdruck; D. Aussage; E. Inferenz; F. Absicht)

Wie Abb. 5 zeigt, sind die Mediane von A, B, D höher als die von C, E, F. Das heißt, im Gegensatz zu den Aufgaben, bei denen die Fähigkeit erforderlich ist, Ausdrücke zu verstehen, den Inhalt zu analysieren und die Autorenabsicht zu erfassen, konnten die Prüflinge solche Aufgaben besser lösen, bei denen die Fähigkeit gefragt ist, den Hauptinhalt des Textes zusammenzufassen, explizite Informationen im Lesetext zu finden sowie richtige oder falsche Aussagen auszuwählen. Hierbei zeichnet sich eine normale Distribution der Korrektheitsquoten bei den Ausgaben mit Lesefähigkeit B ab. Trotzdem sind die vier ungewöhnlich niedrigen Werte nicht zu übersehen. Bei Item 37 in der Prüfung 2017 beispielsweise liegt die Nuancierung zwischen dem Text und der Aufgabe in zwei Wörtern, nämlich „Wirtschaftskraft“ im Text und „Wissenschaftskraft“ in der Option, was bei mangelnder Sorgfalt leicht zu einer falschen Wahl führt.

6. Schlussfolgerung

Der Analyse der Merkmale der Eingabe und der erwarteten Antworten zufolge kann festgestellt werden, dass der Prüfungsteil ‚Leseverstehen‘ der PGG von 2012 bis 2018 im Wesentlichen den Anforderungen der Prüfungsordnung entspricht und die Inhaltsvalidität allgemein hoch ist. Die Lesetexte der PGG weisen verschiedene Textsorten, eine reichhaltige Themenauswahl sowie eine angemessene Gesamtlänge auf und beurteilen multiple Fähigkeiten beim Lesen. Im Gegensatz dazu ist der Schwierigkeitsgrad relativ hoch, was darauf zurückzuführen ist, dass es überdurchschnittlich anspruchsvoll ist, Texte in deutschen Prüfungen zu lesen. Allerdings müssen dabei noch die Einschränkungen der Flesch-Lesbarkeitsformel selbst und die vollständige Anwendbarkeit der Formel auf das Deutsche erörtert werden. In Bezug auf die erwarteten Antworten nimmt der Aufgabentyp, die richtigen Informationen zum Ausfüllen der Lücken auszuwählen, zu. Unter den von den Fragen geforderten Fähigkeiten wird bevorzugt, Äquivalente im Text zu finden und Schlussfolgerungen aus dem Kontext zu ziehen.

Gemessen an der Korrektheitsquote der Antworten finden sich in den Jahren von 2014 bis 2018 ungewöhnlich niedrige Werte, wobei diese Ausreißer hauptsächlich im zweiten Teil vom ‚Leseverstehen‘ auftauchen. Unter den verschiedenen Aufgabentypen bietet die Auswahl einer richtigen oder falschen Aussage im Vergleich zu anderen Typen weniger Schwierigkeiten. Bezüglich der Korrektheit der geprüften Lesefähigkeit lässt sich feststellen, dass die Teilnehmer der PGG von 2012 bis 2018 mit den Lesefähigkeiten vertraut sind, den Hauptinhalt eines Textes zusammenzufassen, explizite Informationen im Lesetext zu finden wie auch richtige oder falsche Aussage auszuwählen.

Ausgehend von den Untersuchungsergebnissen ist für den Deutschunterricht sowie für den Entwurf einer Deutschprüfung im Rahmen des chinesischen Germanistikstudiums folgendes vorzuschlagen:

- a. Themen weiter anreichern, um die Lesefähigkeit bei Texten mit verschiedenen Themen zu testen.
- b. Mit mehr Textsorten in Berührung kommen, damit die Studierenden nicht nur mit einer einzelnen Textsorte vertraut werden.
- c. Vielfältige Aufgaben konzipieren, so dass unterschiedliche Lesefähigkeiten beansprucht werden.

Werden die oben genannten Vorschläge sowohl bei der Vermittlung als auch beim Prüfungsentwurf berücksichtigt, erhöht sich die Qualität des Lesunterrichts sowie die Inhaltsvalidität der PGG, wodurch die Lesefähigkeit der Studierenden umfassend geschult und besser bewertet werden kann.