

Maßnahmen der Qualitätssicherung beim TestAS – Ergebnisse wissenschaftlicher Studien

1 Überblick

Der Test für Ausländische Studierende – TestAS – wurde von der ITB Consulting (Bonn) entwickelt, basierend auf einem Zuwendungsvertrag des Deutschen Akademischen Austauschdienstes und aus Mitteln des Bundesministeriums für Bildung und Forschung. Der Studierfähigkeitstest ermöglicht es Hochschulen in Deutschland, die Eignung ausländischer Studienbewerberinnen und -bewerber, die nicht aus der EU oder dem EWR stammen, zu überprüfen und Entscheidungen über die Zulassung zu treffen. Der TestAS (bzw. ein bestimmtes TestAS-Mindestergebnis) kann beispielsweise als Zulassungsvoraussetzung oder als Bonus im Zulassungsverfahren eingesetzt werden.

TestAS-Ergebnisse können also erhebliche Auswirkungen für Bewerberinnen und Bewerber einerseits und für Hochschulen andererseits („High-Stakes Test“) haben. Umso wichtiger ist es, die Qualität des Tests fortlaufend zu überprüfen und zu sichern. Hierbei spielen statistische Methoden der Testanalyse und Testevaluation eine wichtige Rolle. Qualitätssichernde Maßnahmen werden vom TestDaF-Institut und von der ITB Consulting kontinuierlich vorgenommen. Unter anderem hat die ITB Consulting im Auftrag der Gesellschaft für Akademische Studienvorbereitung und Testentwicklung e. V. (TestDaF-Institut, Bochum) zu diesem Zweck zwei Validitätsstudien durchgeführt (s. unten, Abschnitte 2.1 und 3.1)

Insbesondere ist nachzuweisen, dass die psychometrischen Qualitätskriterien erfüllt werden, die auch international für Tests gelten und die u. a. von der *International Test Commission (ITC)* definiert werden.

Fairness: Ist der TestAS ein *fares* Kriterium zur Auswahl von Studienbewerbern?

Für die Auswahl und Zulassung von Studienbewerberinnen und -bewerbern aus Deutschland (und der EU) ist die Schulabschlussnote (Hochschulzugangsberechtigung, HZB) ein entscheidendes Kriterium. Für viele ausländische Studieninteressierte steht die Bewertung anhand eines solchen *einheitlichen* und *generalisierbaren* Bewertungsmaßstabs (d.h. Notensystems) jedoch nicht zur Verfügung. Es bedarf also eines (zusätzlichen) Maßes, um die Studierfähigkeit dieser Studienbewerberinnen und -bewerber fair einzuschätzen. Der TestAS bietet durch seine Objektivität, Unverfälschbarkeit, Standardisierung und Vergleichbarkeit diese Möglichkeit.

Ein wichtiger Aspekt zur Qualitätssicherung des TestAS ist es daher, dafür zu sorgen, dass das Ergebnis des TestAS in aussagenkräftigem Verhältnis zu Leistungen von Oberstufenschülerinnen und -schülern deutscher Schulen steht: Ein Abiturient oder eine Abiturientin einer deutschen Schule mit guten Schulnoten (und somit guten Chancen auf einen Studienplatz) sollte auch ein gutes TestAS-Ergebnis erzielen. Eine von der ITB Consulting in den Jahren 2013 und 2014 durchgeführte Studie bestätigt, dass diese Voraussetzung erfüllt wird (Abschnitt 2.1).

Zur weiteren Sicherung der Testfairness sorgen wir dafür, dass Teilnehmende verschiedener Personengruppen (z. B. Frauen versus Männer, verschiedene Kulturen) und aus verschiedenen Testläufen beim TestAS dieselben Chancen haben (Abschnitt 2.2).

Reliabilität: Ist der TestAS ein *zuverlässiges* Kriterium zu Auswahl von Studienbewerberinnen und -bewerbern?

Die TestAS-Aufgaben sollen reliabel sein, das heißt, möglichst genau (mit möglichst geringem Messfehler) die im TestAS enthaltenen Facetten der Studierfähigkeit abbilden. Hierzu wird nach der Aufgabenerprobung und bei jedem Testlauf ausgewertet, wie stark die Aufgaben innerhalb der

einzelnen Aufgabengruppen des TestAS (wie der Aufgabengruppe *Muster ergänzen* im Kerntest) miteinander zusammenhängen. Dieses statistische Maß wird *interne Konsistenz* genannt und wird häufig zur Bestimmung der Reliabilität herangezogen. Durch die sorgfältige Konstruktion und Auswahl der einzelnen Aufgaben in einer Aufgabengruppe für einen Testsatz weisen die Reliabilitäten im TestAS kontinuierlich akzeptable bis gute Werte auf.

Validität: Ist der TestAS ein *gültiges* Kriterium zur Auswahl von Studienbewerbern?

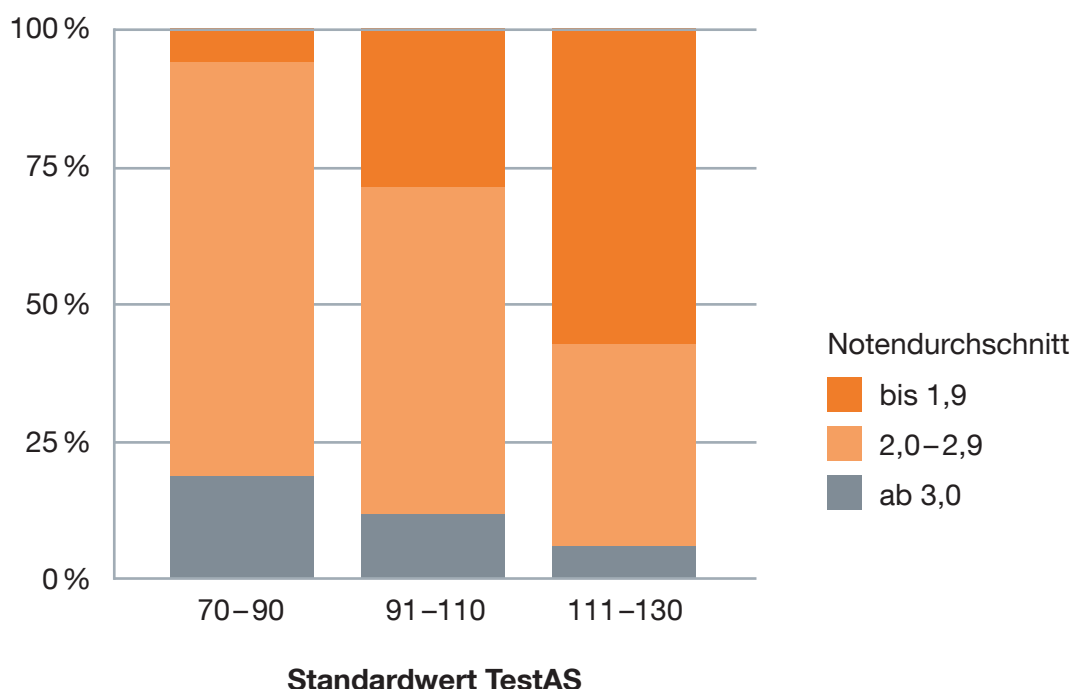
Die Validität (Gültigkeit) eines Tests wird dadurch bestimmt, wie genau der Test das gewünschte Konstrukt abbildet. Die Validität des TestAS lässt sich vor allem durch seine Vorhersagekraft für tatsächlich erzielte Studienleistungen bestimmen. Können Hochschulen auf Grundlage des TestAS systematisch die Studienbewerberinnen und -bewerber auswählen, die in einem Studium erfolgreich agieren werden? Eine Studie, die die ITB Consulting und das TestDaF-Institut gemeinsam durchgeführt haben, bestätigt die Annahme, dass dies der Fall ist (Abschnitt 3.1). Zudem legen die Ergebnisse nahe, dass der TestAS auch eine Entscheidungshilfe für die Studieninteressierten sein kann.

2. Testfairness

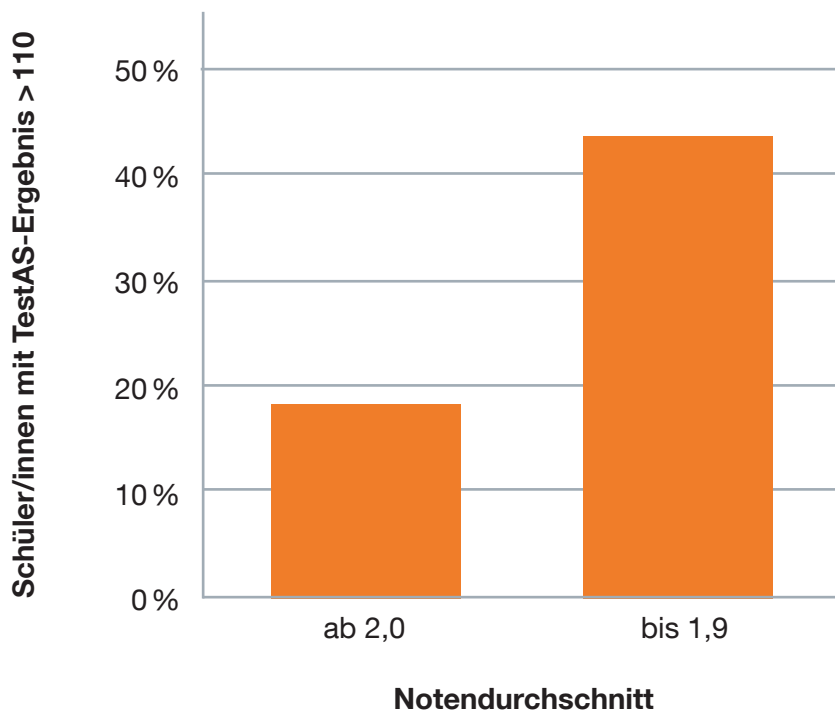
2.1 Zusammenhang zwischen TestAS-Ergebnissen und Schulnoten

Der TestAS steht im Zusammenhang mit der Schulleistung von Oberstufenschülern. Der TestAS wird von deutschen Hochschulen zur Auswahl ausländischer Studienbewerberinnen und -bewerber (die nicht aus der EU oder dem EWR stammen) genutzt. Studienbewerberinnen und -bewerber aus Deutschland (und der EU) werden (unter anderem) auf Grundlage ihrer Schulabschlussnote bewertet. Eine wichtige Fragestellung ist daher, inwieweit die TestAS-Ergebnisse mit den Schulleistungen von Oberstufenschülerinnen und -schülern vergleichbar sind und inwieweit der TestAS eine geeignete und angemessene Schwierigkeit aufweist.

Dieser Frage ist die ITB Consulting im Auftrag von g.a.s.t./dem TestDaF-Institut in einer Studie mit insgesamt 696 Oberstufenschülerinnen und -schülern der Klassen 11, 12 und 13 von 16 Gymnasien im Raum Bonn, Köln und Aachen nachgegangen. Die Studie wurde im Zeitraum von Januar 2013 bis Februar 2014 durchgeführt. Die Ergebnisse der Studie belegen, dass die erreichte Punktzahl des TestAS klar mit dem selbst berichteten Notendurchschnitt der Gymnasiasten in Zusammenhang steht: Schülerinnen und Schüler, die gute Schulnoten hatten, erreichten auch gute Ergebnisse im TestAS.



Die Wahrscheinlichkeit, dass eine Schülerin oder ein Schüler mit der Durchschnittsnote unter 2,0 auch mit seinem TestAS-Ergebnis im überdurchschnittlichen Bereich (über Standardwert 110) liegt, ist beispielsweise mehr als doppelt so hoch wie bei einer Schülerin oder einem Schüler mit einer schlechteren Schulnote (ca. 44 % vs. ca. 18 %).



Der Notendurchschnitt korreliert hoch sowohl mit dem Kerntest (.51) als auch mit den Fachmodulen (.40–.41).

Das bedeutet, dass der TestAS dazu geeignet ist, Studienbewerberinnen und -bewerber zu beurteilen, die nicht über einen direkt mit der deutschen Hochschulreife vergleichbaren Abschluss verfügen.

Die Ergebnisse der Gymnasiasten wurden auch mit denen einer internationalen Stichprobe von TestAS-Teilnehmenden verglichen (N = 2.121). Es zeigt sich, dass die erreichten Punktzahlen beider Gruppen nah beieinander liegen. Im Kerntest bestand (bis zur ersten Nachkommastelle) kein Unterschied; in den drei Fachmodulen „Wirtschaftswissenschaften“ (WIWI), „Mathematik, Informatik und Naturwissenschaften“ (MIN) sowie „Geistes-, Kultur- und Gesellschaftswissenschaften“ (GKG) schnitten die Schülerinnen und Schüler aus deutschen Schulen durchschnittlich etwas besser ab. Im Fachmodul „Ingenieurwissenschaften“ (ING) erzielten hingegen die ausländischen Teilnehmenden etwas bessere Ergebnisse. Dieser Befund legt nahe, dass der TestAS angemessen ist, um ausländische Studienbewerber mit den Anforderungen des deutschen Bildungssystems zu vergleichen.

Der TestAS ermöglicht demnach eine Beurteilung von ausländischen Studienbewerberinnen und -bewerbern mit Blick auf die Anforderungen des deutschen Bildungssystems. Das heißt, dass der TestAS den an ihn gestellten Anforderungen gerecht wird.

2.2 Alle TestAS-Teilnehmer werden mit demselben Maßstab gemessen

Die Schwierigkeit des TestAS ist für alle Teilnehmenden gleich. Der TestAS wird sowohl in deutscher als auch in englischer Sprache angeboten; die Aufgabenzusammensetzung erfolgt für jeden Testlauf neu. Für die Testfairness ist es sehr wichtig sicherzustellen, dass die Ergebnisse aller Testvarianten miteinander vergleichbar sind. So soll weder die Testsprache zu Unterschieden in der Schwierigkeit führen noch ein Testergebnis von der Testversion abhängen.

Um die Schwierigkeit über Testläufe hinweg konstant zu halten, werden alle Testsätze statistisch miteinander in Beziehung gesetzt (Rasch-Modellierung). Dies geschieht über sogenannte Anker-Items. Das sind Aufgaben, die zwischen den Testsätzen gleich sind. Anhand dieser Anker-Items kann das generelle Fähigkeitsniveau der Testgruppen errechnet und die Schwierigkeit aller anderen Items und damit des Testsatzes bestimmt werden.

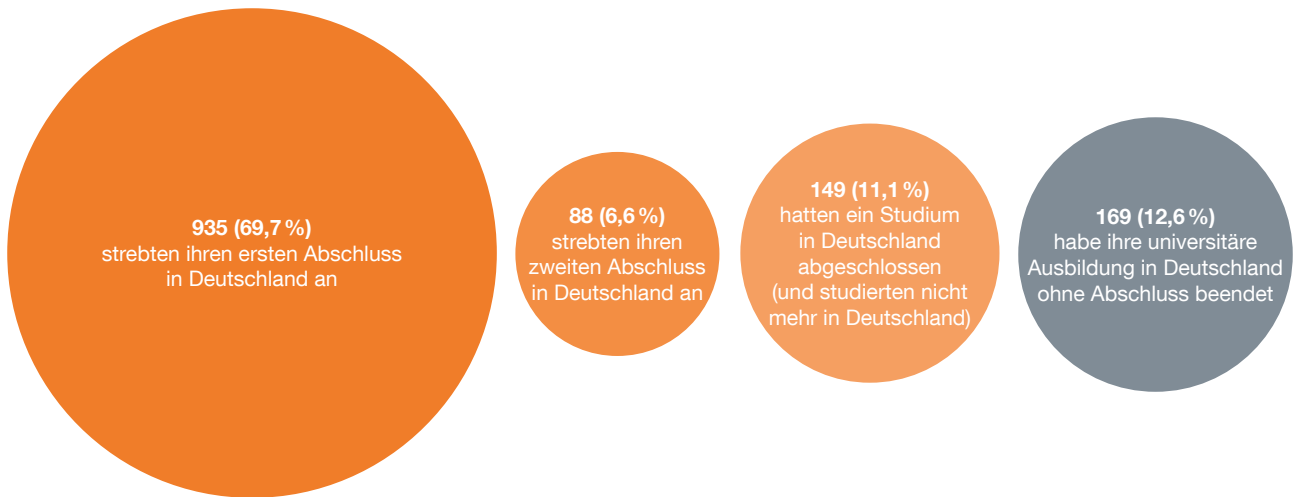
Auch Unterschiede zwischen einzelnen Teilnehmergruppen, die nicht auf Unterschiede in der Fähigkeit zurückgehen, sollen vermieden werden. So dürfen (bei gleicher Fähigkeit) die Aufgaben für einen Teilnehmer zum Beispiel nicht weniger gut lösbar sein als für eine Teilnehmerin (oder umgekehrt). Nach jedem Testlauf wird daher überprüft, ob bei den Items gruppenspezifische Unterschiede bestehen; die Ergebnisse dieser Analysen fließen dann in die Itemauswahl, Itementwicklung und Testsatzerstellung mit ein. Die folgende Abbildung zeigt ein Beispiel für die Ausgeglichenheit in den Standardwerten männlicher und weiblicher Teilnehmender der Vergleichsstudie im Fachmodul MIN. Die minimale Abweichung (Standardwerte 108 vs. 109) ist statistisch nicht signifikant (d. h., sie weist nicht auf einen tatsächlich bestehenden Unterschied).



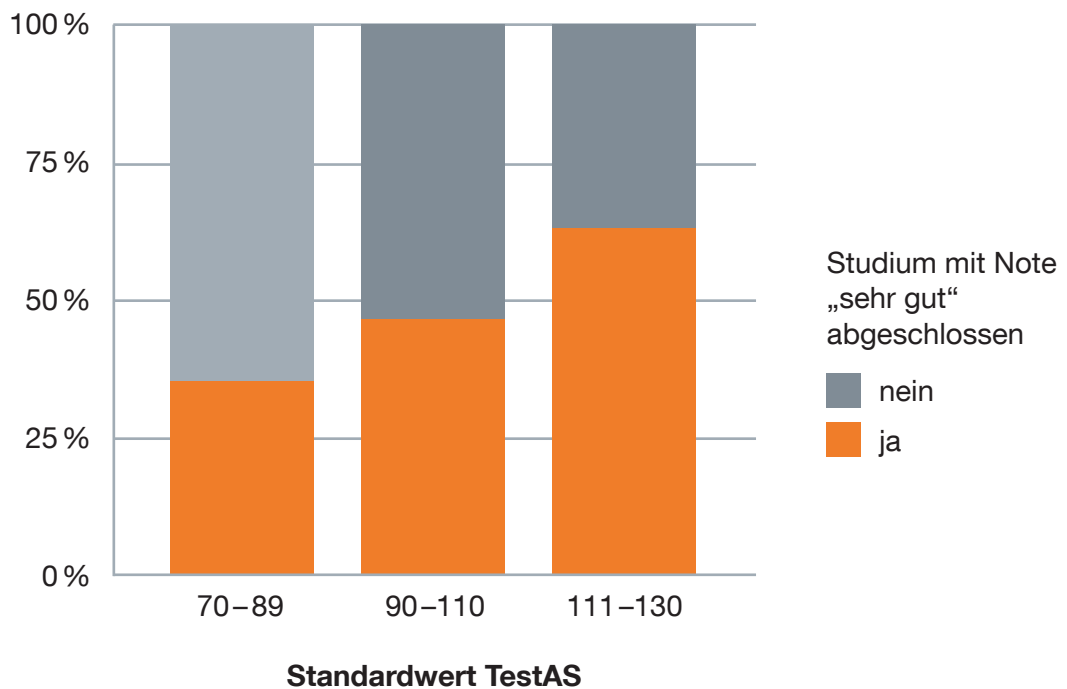
3. Validität

3.1. Prognostische Validität

Der TestAS ermöglicht eine passgenaue Bewerberauswahl. Eine Studie von der ITB Consulting und dem TestDaF-Institut belegt, dass sich der Erfolg im Studium gut durch den TestAS vorhersagen lässt. Die studienfeldspezifischen Module des TestAS erlauben besonders gute Prognosen für den akademischen Erfolg im entsprechenden Studienfeld. Insgesamt bearbeiteten mehr als 4.000 TestAS-Teilnehmende einen Fragebogen über ihren Studienverlauf und -erfolg. 1.341 davon hatten in Deutschland studiert oder studierten dort zum Zeitpunkt der Erhebung noch. 70,5 % (935) der Teilnehmenden strebten ihren ersten Abschluss an. 17,7 % (237) hatten bereits einen Abschluss in Deutschland erreicht, 88 von ihnen strebten sogar ihren zweiten Abschluss an. Nur 12,6 % (169) dieser TestAS-Teilnehmenden beendeten ihre universitäre Ausbildung in Deutschland ohne Abschluss.



Hinsichtlich der Vorhersagekraft des TestAS für den Erfolg in einzelnen Studiengängen lässt sich festhalten: Mit dem TestAS kann der Anteil der erfolgreichen Studierenden an den ausländischen Studierenden erhöht werden. Das bedeutet, dass auf Grundlage des TestAS genau die ausländischen Studierenden ausgewählt werden können, für die die Aussicht auf Studienerfolg am höchsten ist. Die Korrelation zwischen Studienerfolg und dem zum Studienfeld passenden TestAS-Fachmodul ist statistisch signifikant. Wie die nachfolgende Grafik deutlich macht, steigt die Wahrscheinlichkeit, dass ein ausländischer Studierender sein Studium mit der Note „sehr gut“ abschließt, mit einem höheren TestAS-Ergebnis.

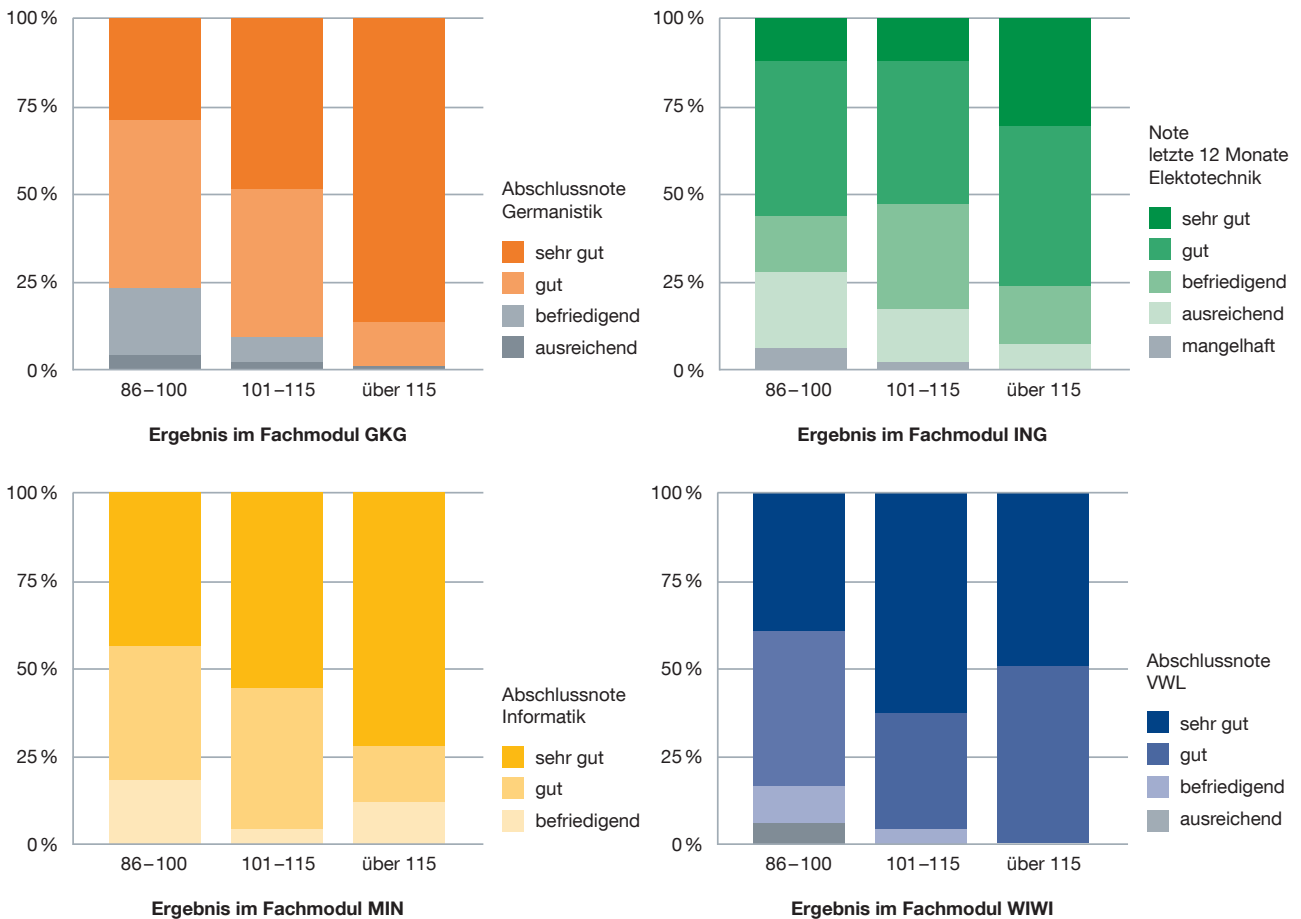


Zusammen mit den Schulabschlussnoten ergeben sich für die einzelnen TestAS-Module Korrelationen von .23 (MIN), .45 (ING), .44 (GKG) und .38 (WiWi) mit der Studienabschlussnote.

Neben den kognitiven Fähigkeiten gibt es weitere Faktoren, die sich auf die erbrachten Leistungen eines Studierenden auswirken können (z. B. Persönlichkeitsmerkmale wie Gewissenhaftigkeit und Selbstdisziplin). Dies gilt umso mehr für ausländische Studierende, bei denen zusätzlich weitere Einflüsse wie Umgang mit Sprache und Kultur eine (z. T. sehr gewichtige) Rolle spielen. Das heißt, dass die kognitiven Fähigkeiten, die ein Studierfähigkeitstest erfasst, nur einen Teil der Unterschiede zwischen Studierenden aufklären können. Die Auswahl erfolgreicher Studierender stellt für

Hochschulen weltweit eine sehr große Herausforderung dar, auch weil viele der relevanten Faktoren gar nicht zuverlässig erfasst werden können. Daher ist jede Verbesserung der Prognosefähigkeit gewinnbringend. Bei der Auswahl ausländischer Studierender kommt hinzu, dass Bewerberinnen und Bewerber mancher Herkunftsländer nur schwer interpretierbare Nachweise über ihre Schulbildung vorlegen können, die zur Auswahl kaum genutzt werden können. Daher ist der TestAS umso bedeutender für die Auswahl dieser Studierenden.

Die nachfolgende Darstellung veranschaulicht: Der TestAS hilft vorherzusagen, welche Bewerberinnen und Bewerber ein Studium erfolgreich absolvieren werden: Je besser das TestAS-Ergebnis ist, desto höher ist die Wahrscheinlichkeit, gute Studienergebnisse zu erzielen.



Bei der Interpretation der Ergebnisse sollte auch beachtet werden, dass ein Teil der Teilnehmenden den TestAS möglicherweise nur mit geringer Motivation abgelegt hat, da der TestAS zum Zeitpunkt der Datenerhebung gebührenfrei war und von nur wenigen Hochschulen genutzt wurde. Dies kann dazu führen, dass die Zusammenhänge zwischen TestAS und Studienleistungen in dieser Studie geringer ausfallen als sie tatsächlich sind. Neue Studien zur Prüfung der Validität des TestAS sind daher geplant.

Festzuhalten bleibt: Ein signifikanter Zusammenhang zwischen Studierfähigkeitstest und Studienleistungen bedeutet, dass die Vorhersage von Studienerfolg durch den TestAS verbessert (und im Falle nicht interpretierbarer Schulabschlussnoten bei ausländischen Studierenden vielfach erst ermöglicht) werden kann.

3.2 Weitere Ergebnisse der Validitätsstudie

Der TestAS dient als Entscheidungshilfe – auch für die Studienbewerber selbst. Die Daten der Validitätsstudie deuten auch darauf hin, dass der TestAS eine steuernde Funktion hat und Studieninteressierte bei ihrer Entscheidung für oder gegen ein Studium in Deutschland oder bei dem Studium einer bestimmten Fachrichtung unterstützt: Zum einen gaben 40 % der Teilnehmenden an, den TestAS abgelegt zu haben, um eine Selbsteinschätzung (Self-Assessment) zu erhalten (und nicht, weil sie das Zertifikat für ihre Bewerbungen genutzt haben). Zum anderen zeigt sich, dass die Wahrscheinlichkeit, tatsächlich ein Studium in Deutschland aufzunehmen, bei den Teilnehmenden mit einem guten TestAS-Ergebnis höher war als bei Teilnehmenden mit einem weniger guten Ergebnis. Da der TestAS zum Zeitpunkt der Studie nur von wenigen deutschen Hochschulen als Auswahlkriterium bei der Zulassung herangezogen wurde, kann dies nicht dadurch begründet werden, dass die TestAS-Ergebnisse unmittelbar für ein aufgenommenes Studium entscheidend waren. Dieser Zusammenhang besteht auch über den Einfluss der Schulabschlussnoten hinaus; tatsächlich sagt das TestAS-Ergebnis sogar besser voraus, ob ein Teilnehmender ein Studium in Deutschland aufgenommen hat, als es die Schulabschlussnote tut. Dies ist ein Hinweis darauf, dass Studieninteressierte das TestAS-Ergebnis für ihre Entscheidung für oder gegen ein Studium in Deutschland nutzten.

