

# Eignung von HAM-Nat und TMS-Modul "Medizinisch-naturwissenschaftliches Grundverständnis" für die Studienbewerberauswahl in der Medizin

## Zusammenfassung

**Ziele:** Tests mit naturwissenschaftlichen Inhalten sind prädiktiv für den Studienerfolg in den ersten Semestern des Medizinstudiums. Einige deutschsprachige Universitäten verwenden für die Studienbewerberauswahl den Test für Medizinische Studiengänge (TMS), dessen Test-Modul „Medizinisch-naturwissenschaftliches Grundverständnis“ die Fähigkeit zu schlussfolgerndem Denken erfasst. Im Hamburger Auswahlverfahren für Medizinische Studiengänge - Naturwissenschaftsteil (HAM-Nat) werden dagegen naturwissenschaftliche Kenntnisse überprüft. In dieser Studie vergleichen wir die prädiktive Stärke des HAM-Nat mit einem Test (NatDenk), der dem TMS-Modul „Medizinisch-naturwissenschaftliches Grundverständnis“ inhaltlich und strukturell ähnelt.

**Methoden:** Im Jahr 2007 nahmen 162 Studienanfänger der Humanmedizin freiwillig am HAM-Nat (N=77) oder am NatDenk (N=85) teil. Bis zum Frühjahr 2011 hatten 84,2% der Testteilnehmer das Physikum in Hamburg bestanden. Mittels verschiedener logistischer Regressionsmodelle überprüften wir die Vorhersagekraft der Abiturnote und der Testergebnisse (HAM-Nat und NatDenk) auf das Studienerfolgskriterium „Physikum nach 7 Studiensemestern bestanden“. Berichtet werden die Odds Ratios (OR) für Studienerfolg.

**Ergebnisse:** Für beide Testgruppen bestand ein signifikanter Zusammenhang zwischen Testergebnis und Studienerfolg (HAM-Nat: OR=2,07; NatDenk: OR=2,58). Werden beide Zulassungskriterien in einem Modell geschätzt, sind in der HAM-Nat Gruppe beide Haupteffekte (Abitur: OR=2,45; Test: OR=2,32) und deren Interaktionseffekt (OR=1,80) signifikant, während in der NatDenk Gruppe nur das Testergebnis (OR=2,21) allein einen signifikanten Beitrag zur Aufklärung der Varianz liefert.

**Schlussfolgerungen:** Für sich genommen haben sowohl HAM-Nat als auch NatDenk prädiktive Information für den Studienerfolg, allerdings klärt die Abiturnote nur bei Verwendung des HAM-Nat zusätzliche Varianz auf. Die Auswahl nach HAM-Nat und Abiturnote hat unter den aktuellen Bedingungen des Mediziner Auswahlverfahrens mit vielen guten Bewerbern aber nur wenigen Studienplätzen die höchste Vorhersagekraft aller Modelle.

**Schlüsselwörter:** HAM-Nat, Studienbewerberauswahl, Medizin, Naturwissenschaftstest, TMS

Johanna Hissbach<sup>1</sup>

Lena Feddersen<sup>1</sup>

Susanne Sehner<sup>1,2</sup>

Wolfgang Hampe<sup>1</sup>

1 Universitätsklinikum  
Hamburg Eppendorf (UKE),  
Institut für Biochemie und  
Molekulare Zellbiologie,  
Hamburg, Deutschland

2 Universitätsklinikum  
Hamburg Eppendorf (UKE),  
Institut für medizinische  
Biometrie und Epidemiologie,  
Hamburg, Deutschland

## Autorenschaft

Johanna Hissbach und Lena Feddersen trugen in gleichen Teilen zu dieser Arbeit bei und teilen sich die Erstautorschaft.

## Einleitung

Studienbewerber für das Fach Medizin sind in Deutschland mit vielen unterschiedlichen Zulassungsmodalitäten konfrontiert [1]. Während die Abiturnote als maßgebliches

Auswahlkriterium gesetzlich vorgeschrieben ist [2], können weitere, hochschulspezifische Kriterien wie Tests, Interviews oder Boni für bestimmte Leistungen in die Auswahlentscheidung mit einbezogen werden. Der wohl bekannteste deutsche Studieneingangstest ist der „Test für Medizinische Studiengänge“ (TMS) [3]. Nachdem der TMS in Deutschland nach 1996 nicht mehr eingesetzt wurde, begannen im Jahr 2007 die baden-württembergischen Universitäten damit, gute Testergebnisse bei der Zulassung zu berücksichtigen. Im Wintersemester 2011/12 vergaben 10 von 34 medizinischen Fakultäten Bonuspunkte für gute TMS-Leistungen.

Seit 2008 wird am Universitätsklinikum Eppendorf das Hamburger Auswahlverfahren für medizinische Studiengänge, Naturwissenschaftsteil (HAM-Nat) zur Studienbewerbersauswahl eingesetzt. Beide Tests, der HAM-Nat und der TMS, wurden eigens für die Auswahl von Medizinstudierenden entwickelt, zielen jedoch auf unterschiedliche Konstrukte ab. Der HAM-Nat zielt darauf ab, auf Oberstufenniveau naturwissenschaftliche Kenntnisse und Fähigkeiten, die für das Medizinstudium relevant sind, zu überprüfen. Der TMS enthält mit dem Testmodul „medizinisch-naturwissenschaftliches Grundverständnis“ zwar auch naturwissenschaftliche Fragen, diese sollen aber „verbal-schlußfolgerndes, deduktives Denken in fachbezogenen Kontexten“ abbilden, es ist kein spezielles Fachwissen erforderlich [4]. Dieses Modul ist eine von vier medizinischen Aufgabengruppen, auf denen die Vorhersagekraft des TMS im Wesentlichen beruht [3]. Im Gegensatz zum HAM-Nat, bei dem die Vorbereitung motivierter Bewerber dazu führt, dass für die ersten Semester relevante schulische Themen vor Studienbeginn wiederholt werden, hat die Vorbereitung auf den TMS keinen inhaltlichen Bezug zum Studium und soll nach Aussage der Testentwickler zudem nur zu einer geringfügigen Ergebnisverbesserung führen [3].

Der HAM-Nat wurde mit dem Ziel entwickelt, die Abbrecherquote im 1. Studienabschnitt zu reduzieren. Da die Abiturdurchschnittsnote mit einer korrigierten prädiktiven Stärke von  $r=0,58$  [5] für Studiennoten im vorklinischen Studienabschnitt ein solider Prädiktor für Studienerfolg (Physikumsnote) ist, stellt sich die Frage, was wir gewinnen, wenn wir weitere Tests zur Studierendenauswahl einsetzen.

Sowohl das HAM-Nat Testergebnis als auch die Abiturnote waren in der 2006er Kohorte signifikante Prädiktoren für Studienerfolg nach dem 2. Semester, operationalisiert als Anzahl der bestandenen Prüfungen ( $r=0,31$  bzw.  $r=0,26$ ), und klärten gemeinsam 13% der Gesamtvarianz auf [6]. Der Beitrag zur erklärten Varianz war für den HAM-Nat allein (9,5%) höher als für die Abiturnote allein (6,6%). Betrachtet man eine Stichprobe, die der Hochschulquote entspricht, ist die Varianz der Abiturnote derart eingeschränkt, dass keine signifikante Korrelation mit dem HAM-Nat zu finden ist ( $r=0,11$ ) [6]. Unterschiede im Bereich der sehr guten Abiturnoten (1,2 – 1,7) haben also keinen wesentlichen Einfluss auf das HAM-Nat Ergebnis. Die Varianzeinschränkung der Abiturnote in der Gruppe der Hochschulquote führt auch dazu, dass sie – im Gegensatz zum HAM-Nat Ergebnis – nicht mehr signifikant mit dem Kriterium Studienerfolg korreliert. Für den HAM-Nat sinkt die prädiktive Stärke hingegen nur leicht von  $r=0,31$  auf  $r=0,26$ .

Die Ergebnisse unterschiedlicher HAM-Nat Testversionen aus den Vorstudien mit Studienanfängern 2006 und 2007 korrelieren mit der Abiturdurchschnittsnote in der Stichprobe aller Studienanfänger zwischen  $r=0,12$  und  $r=0,34$  [6], [7], der TMS-Subtest „medizinisch-naturwissenschaftliches Grundverständnis“ gehört zu einer Gruppe von TMS-Modulen, die mit der Abiturnote zwischen  $r=0,28$  und  $r=0,40$  korrelieren (3). Die Korrelatio-

nen des gesamten TMS mit der Abiturnote sind mit  $r=0,36$  bis  $r=0,48$  höher. Dennoch schlussfolgern die TMS-Autoren, dass Schul- und Testleistung größtenteils unterschiedliche Leistungsaspekte erfassen [3], [4]. Im TMS-Modul „medizinisch-naturwissenschaftliches Grundverständnis“ wird die Fähigkeit erfasst, „aus Geschriebenem rasch das Essentielle zu extrahieren und die richtigen Schlüsse daraus zu ziehen“ (S. 53 in [3]). Untersuchungen zeigten, dass hierfür eine interne Repräsentation der Textinformationen im Arbeitsgedächtnis, die Konstruktion einer Inferenzkette, und eine Umkodierung von Textaussagen in ein Vorstellungsbild wichtig sind [3].

Im Rahmen einer zweiten Studie vor der Einführung des HAM-Nat in das Hamburger Auswahlverfahren wurde der Zusammenhang des HAM-Nat mit dem Testmodul „Naturwissenschaftliches Denken“ (NatDenk) untersucht. Dieses Modul ähnelt strukturell und inhaltlich dem Subtest „Medizinisch-naturwissenschaftliches Grundverständnis“ des TMS. Für den HAM-Nat betrug die Korrelation mit dem NatDenk je nach Testversion  $r=0,34$  oder  $r=0,21$  [7], beide Tests bilden also unterschiedliche Konstrukte ab.

In Deutschland, Österreich und der Schweiz werden von unterschiedlichen Fakultäten sowohl Kenntnis- und Fähigkeitstests mit naturwissenschaftlichen Inhalten wie der HAM-Nat als auch Studierfähigkeitstests wie der TMS von medizinischen Fakultäten zur Studierendenauswahl eingesetzt. Ziel ist jeweils die Optimierung des Studienerfolgs. In der hier beschriebenen Untersuchung vergleichen wir die prädiktive Stärke der beiden Tests bezüglich des Studienerfolgs. Als Studienerfolgssparameter wählten wir das Bestehen des Physikums innerhalb der ersten sieben Semester, da die bis zu diesem Zeitpunkt erfolglosen Studierenden in der Regel das Studium an der Universität Hamburg nicht abschließen. Als zweites Kriterium untersuchten wir „Physikum nach dem 4. Semester bestanden“, da viele Universitäten einen Abschluss in Regelstudienzeit anstreben.

## Methoden

### HAM-Nat

Die 2007er Version des HAM-Nat umfasst 60 Multiple-Choice Fragen aus dem Bereich Mathematik, Chemie, Physik und Biologie. Die Inhalte stammen aus medizinrelevanten Themengebieten auf dem Niveau der gymnasialen Oberstufe. Auf der Internetseite des Universitätsklinikums Eppendorf (UKE) sind der aktuelle Themenkatalog und ein Selbsttest mit Fragen der 2006 und 2007er Testversionen zu finden (<http://www.uke.de/studienbewerber>). Eine von fünf Antwortalternativen ist jeweils richtig und die Teilnehmer hatten pro Frage 1,5 Minuten Zeit zur Bearbeitung. Entwickelt wurden die Fragen von Gymnasiallehrern sowie von Dozenten der klinischen und theoretischen Fächer der medizinischen Fakultät.

## Testmodul „Naturwissenschaftliches Denken“

Für den Vergleich des HAM-Nat mit dem TMS-Modul „medizinisch-naturwissenschaftliches Grundverständnis“ als externem Kriterium entwickelte die ITB-Consulting GmbH, die auch den TMS erstellt, ein inhaltlich und strukturell dem TMS-Modul entsprechendes Testmodul (NatDenk) mit 24 Multiple-Choice Aufgaben. Jede Aufgabe ist so aufgebaut, dass eine von fünf aufgestellten Behauptungen zu einem zuvor dargestellten naturwissenschaftlichen Sachverhalt passt. Für die 24 Aufgaben haben die Testteilnehmer 55 Minuten Zeit. Die Aufgaben setzen kein spezifisch naturwissenschaftliches Wissen voraus, sondern sollen die Durchdringung eines Sachverhalts und die Fähigkeit zu schlussfolgerndem Denken abbilden. Das Recht, das Testmodul durchzuführen, wurde von ITB-Consulting erworben.

## Studienerfolg

Studienerfolg wurde operationalisiert als „Physikum bis nach dem 7. Semester in Hamburg bestanden“ (*Erfolg7Sem*). Wir wählten dieses dichotome Kriterium als Hauptkriterium, da das primäre Ziel bei der Entwicklung des HAM-Nat war, die Zahl der Studienabbrecher im vor-klinischen Studienabschnitt zu verringern. Die Wahrscheinlichkeit, das Physikum noch zu bestehen, sinkt mit zunehmender Anzahl von Fachsemestern. Um die Stabilität der Modelle zu überprüfen, berechneten wir noch den Zusammenhang mit einem weiteren dichotomen Studienerfolgskriterium: Bestehen des Physikums in Regelstudienzeit nach 4 Semestern (*Erfolg4Sem*). Dieses Kriterium wird insbesondere dann verwendet, wenn der Fokus auf monetären Aspekten liegt, da durch plangemäße Studienverläufe Kosten eingespart werden können. Für eine übersichtliche Darstellung diskutieren wir die Ergebnisse für das Kriterium *Erfolg4Sem* nur, wenn sie sich von unserem Hauptkriterium *Erfolg7Sem* unterscheiden.

## Studiendesign

Die Untersuchung fand im Jahr 2007 im Rahmen einer breiter angelegten Studie zur Parallel- und Retestrelativität des HAM-Nat statt [7]. In der Orientierungseinheit wurde den Studienanfängern der Humanmedizin die Teilnahme an dieser Studie angeboten. Vier Wochen nach Semesterbeginn wurden die Teilnehmer randomisiert in zwei Gruppen aufgeteilt: eine Gruppe bearbeitete den HAM-Nat, die andere das Testmodul „Naturwissenschaftliches Denken“. Die Studienteilnahme war freiwillig und die Durchführung des Tests im Anschluss an eine Pflichtlehrveranstaltung wurde durch Mitarbeiter unserer Arbeitsgruppe organisiert und von Dozenten der Medizinischen Fakultät beaufsichtigt. Der Test „Naturwissenschaftliches Denken“ wurde eigens für diese Studie durchgeführt. Im Mai 2011 wurden die Testergebnisse mit den Studienerfolgsergebnissen zusammengeführt.

## Stichprobe

Die Studie wurde mit Studienanfängern der Kohorte 2007 durchgeführt. Diese waren nach ihrer Abiturdurchschnittsnote ausgewählt worden oder über die anderen Quoten (Wartezeit, Ausländerquote) zugelassen worden. Der Numerus clausus lag in diesem Jahr in Hamburg bei 1,6. Berücksichtigt wurden für diese Auswertung nur die Daten derjenigen Studienteilnehmer, denen eine Abiturnote zugeordnet werden konnte. Die Gesamtstichprobe, auf deren Basis diese Analyse erfolgte, umfasste 162 Studienanfänger. Diese Stichprobe unterscheidet sich in Hinsicht auf Geschlecht, Alter und Abiturnote nicht signifikant von der Gesamtheit der Studienanfänger in der Orientierungseinheit. Die eine Hälfte der Testteilnehmer (N=77) bearbeiteten die 2007er HAM-Nat Testversion und die andere Hälfte (N=85) eine Woche später das Modul „Naturwissenschaftliches Denken“ (NatDenk). Alle Studienteilnehmer willigten schriftlich in die Verwendung ihrer Daten ein.

## Statistische Analyse

Um die Testergebnisse und die Abiturnoten auf einer gemeinsamen Skala abbilden zu können, wurden sie z-transformiert. Mit Hilfe eines logistischen Regressionsmodells überprüften wir in drei verschiedenen Modellen die Vorhersagekraft der Abiturnote sowie der Testergebnisse (HAM-Nat und NatDenk) auf die Studienerfolgskriterien *Erfolg7Sem* und *Erfolg4Sem*. Innersubjektive Faktoren waren jeweils die Abiturnote und das Testergebnis, Zwischensubjektive Faktoren die Testgruppenzugehörigkeit (HAM-Nat vs. NatDenk). Wir berichten die Odds Ratios und 95%-Konfidenzintervalle. IBM SPSS Statistics, Version 19.0.0 wurde für die Analysen verwendet.

## Ergebnisse

### Stichprobe

Von allen Studienteilnehmern haben in Hamburg 67,3% das Physikum in Regelstudienzeit und 84,6% bis zum Frühjahr 2011, also in 7 Semestern, bestanden. Einen Überblick über die Charakteristika der Gesamtstichprobe und der beiden Substichproben liefert Tabelle 1. Signifikante Korrelationen mit der Abiturnote gab es für die Physikums Gesamtnote ( $r=0,24$ ;  $p=0,004$ ) und den HAM-Nat ( $r=-0,24$ ;  $p=0,038$ ), nicht aber für den NatDenk ( $r=-0,11$ ;  $p=0,324$ ). Die negativen Korrelationswerte bedeuten, dass gute (niedrige) Abiturnoten mit guten (hohen) Testergebnissen zusammenhängen.

### Prädiktion

Die Auswertung der Ergebnisse erfolgte in mehreren Schritten. Wir haben getrennt für die Abiturnote und die Testergebnisse untersucht, inwieweit sie mit dem Studienerfolg zusammenhängen und schließlich in einem

Tabelle 1: Stichprobencharakteristika

	Stichprobe		Vergleich	
	Gesamt (N=162)	HAM-Nat (N=77)	NatDenk (N=85)	HAM-Nat/NatDenk
	N (%)	N (%)	N (%)	$\chi^2$ (p)
Geschlecht weiblich	108 (66,7%)	48 (62,3%)	57 (67,1%)	0,17 (0,678)
<i>Erfolg4Sem</i>	109 (67,3%)	57 (75,0%)	52 (60,5%)	3,87 (0,049)
<i>Erfolg7Sem</i>	137 (84,6%)	66 (85,7%)	71 (83,5%)	0,10 (0,751)
	M (SD)	M (SD)	M (SD)	T (p)
Alter	21,59 (3,64)	20,63 (2,55)	22,43 (4,21)	3,23 (0,001)
Abiturnote	1,80 (0,64)	1,70 (0,53)	1,89 (0,71)	1,87 (0,063)
Testergebnis (in %)	-	40,44 (12,70)	58,38 (17,12)	-
Physikumsgesamtnote	2,73 (0,75) <sup>a</sup>	2,62 (0,71) <sup>b</sup>	2,83 (0,78) <sup>c</sup>	1,59 (0,115)

<sup>a</sup> N=137, <sup>b</sup> N=66, <sup>c</sup> N=71

Gesamtmodell ergänzend die Interaktion zwischen beiden Kriterien mit eingeschlossen.

### Modell 1: Abiturnote

In der Gesamtstichprobe bestand ein signifikanter Zusammenhang zwischen der Abiturnote und dem Studienerfolg (Odds Ratio: OR=1,72; p=0,006 für *Erfolg7Sem* und OR=1,98; p=0,000 für *Erfolg4Sem*). Das OR von 1,72 bedeutet, dass die Chance, das Physikum nach 7 Semestern bestanden zu haben, mit jeder Verbesserung in der Abiturnote um eine Standardabweichung (ca. 0,6 Notenpunkte) um das 1,72-fache steigt.

Getrennt nach der Testgruppe zeigt sich ein Effekt der Abiturnote auf den *Erfolg7Sem*, der für die HAM-Nat Gruppe signifikant ist und für die NatDenk Gruppe knapp über der Signifikanzschwelle von p=0,05 liegt (siehe Tabelle 2). Abbildung 1 veranschaulicht diesen Effekt: Die unabhängige Variable Abiturnote wurde am Median dichotomisiert in gut (Abi≤1,6) vs. schlecht (Abi≥1,6) und die geschätzten Wahrscheinlichkeiten in Abhängigkeit für den Studienerfolg getrennt für die beiden Gruppen aufgetragen. Für das Kriterium *Erfolg4Sem* bestand in beiden Gruppen ein signifikanter Zusammenhang zwischen Abiturnote und Studienerfolg (siehe Tabelle 2).

### Modell 2: Testergebnis (HAM-Nat vs. NatDenk)

Jeweils für die Tests allein, also ohne den Einfluss der Abiturnote, besteht ein signifikanter Zusammenhang zwischen Testergebnis und Studienerfolg (siehe Tabelle 2). Während sich die Wahrscheinlichkeit, das Physikum zu bestehen, pro Standardabweichung nach oben im HAM-Nat in etwa verdoppelt, steigt sie beim NatDenk etwas stärker um den Faktor 2,58. Die beiden ORs unterscheiden sich nicht signifikant (p=0,909).

### Modell 3: Abiturnote und Testergebnis (HAM-Nat vs. NatDenk) mit Interaktion

Werden die Effekte beider Zulassungskriterien sowie deren Wechselwirkung in einem Modell geschätzt, sind

in der HAM-Nat Gruppe beide Haupteffekte und die Wechselwirkung signifikant, während in der NatDenk Gruppe nur das Testergebnis allein einen signifikanten Beitrag zur Aufklärung der Varianz liefert (siehe Tabelle 2). Das bedeutet, dass beim HAM-Nat die prädiktive Stärke des Testergebnisses von der Abiturnote des Getesteten abhängt. Vor allem bei guten Abiturnoten enthält der HAM-Nat zusätzliche Informationen für den Studienerfolg (siehe Abbildung 1), während die Abiturnote zusätzlich zum NatDenk keine weitere Information liefert.

Für das Kriterium *Erfolg4Sem* ist die Interaktion zwischen HAM-Nat und Abiturnote nicht signifikant und die Abiturnote verliert wenig prädiktive Stärke. In der NatDenk Gruppe tragen nun zwar beide Prädiktoren signifikant zur Vorhersage des Studienerfolgs bei, allerdings ist der Einfluss der Abiturnote größer als der des Testergebnisses.

## Sensitivität und Spezifität

Die *Receiver Operating Characteristic* (ROC) zeigt den Zusammenhang zwischen Sensitivität und Spezifität eines Tests (siehe Abbildung 2). Mithilfe der ROC-Kurven kann man verdeutlichen, was passieren würde, wenn man verschiedene Selektionsquoten bei der Studierendenauswahl anwendet.

Eine hohe Sensitivität bedeutet in unserem Fall, dass wir möglichst wenige Bewerber ablehnen, die erfolgreich studieren können. Die höchste Sensitivität könnten wir erreichen, indem wir alle Bewerber zulassen, was einer Selektionsquote von 100% entspricht (also keine Selektion, rechts auf der x-Achse in Abbildung 2). Dann hätten alle, die das Physikum in diesem Zeitraum schaffen, auch eine Chance bekommen, und wir hätten niemanden fälschlicherweise abgelehnt. Allerdings hätten wir so auch all diejenigen zugelassen, die das Physikum nicht schaffen (geringe Spezifität, ebenfalls rechts auf der x-Achse). Ist unser Ziel in erster Linie, die Studienabbrecher zu identifizieren, brauchen wir einen spezifischen Test. Legen wir eine strengere Selektionsquote zugrunde, bewegen uns also weiter nach links auf der x-Achse, sinkt unsere Sensitivität bei steigender Spezifität. Durch eine strengere

Tabelle 2: Logistische Regression mit den Zielkriterien *Erfolg7Sem* und *Erfolg4Sem*

Modell	Prädiktoren	<i>Erfolg7Sem</i> OR (p)		<i>Erfolg4Sem</i> OR (p)	
		HAM-Nat	NatDenk	HAM-Nat	NatDenk
1	Abiturnote	1,84 (0,031)	1,62 (0,078)	1,74 (0,027)	2,30 (0,001)
2	Testergebnis	2,07 (0,027)	2,58 (0,007)	2,36 (0,006)	1,87 (0,009)
3	Testergebnis	2,32 (0,028)	2,21 (0,029)	2,37 (0,009)	1,80 (0,025)
	Abiturnote	2,45 (0,010)	1,36 (0,393)	1,92 (0,029)	2,51 (0,002)
	Interaktion Abiturnote & Test	1,80 (0,037)	0,74 (0,370)	1,48 (0,119)	0,72 (0,270)

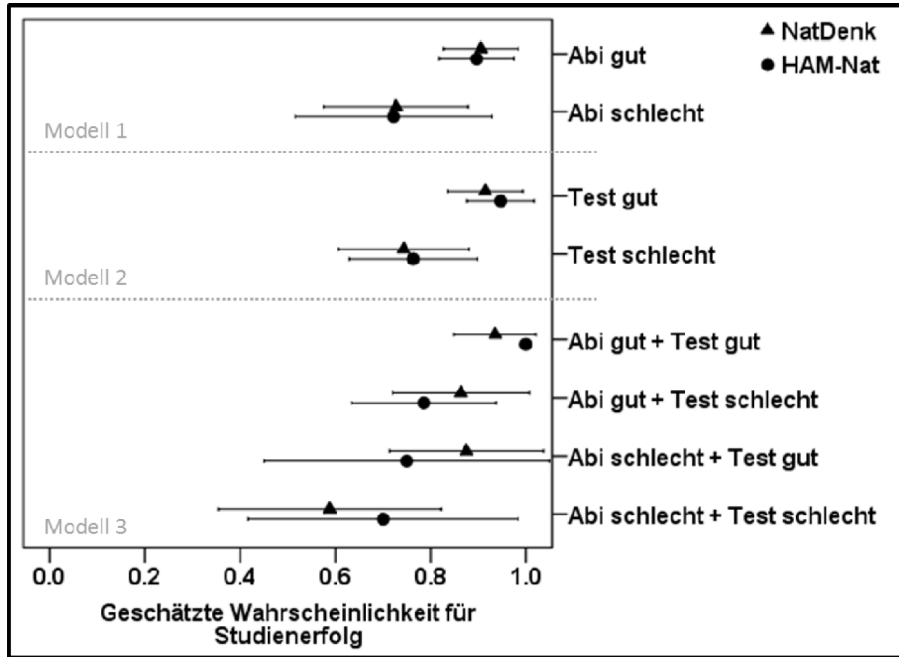


Abbildung 1: Geschätzte Wahrscheinlichkeit für *Erfolg7Sem* in Abhängigkeit von Abiturnote und Testergebnis. Die beiden Prädiktoren wurden am Median dichotomisiert (Abit gut: <1,6 und Test gut: >Median). Die Güte eines Prädiktors lässt sich aus dem Unterschied der Studienerfolgswahrscheinlichkeit bei guten und schlechten Ausprägungen ablesen. Beispielsweise ist im Modell 3 bei gutem Abitur die Studienerfolgswahrscheinlichkeit für gute HAM-Nat-Teilnehmer deutlich höher als für die unterdurchschnittlichen, wohingegen die Differenz für den NatDenk deutlich geringer ist.

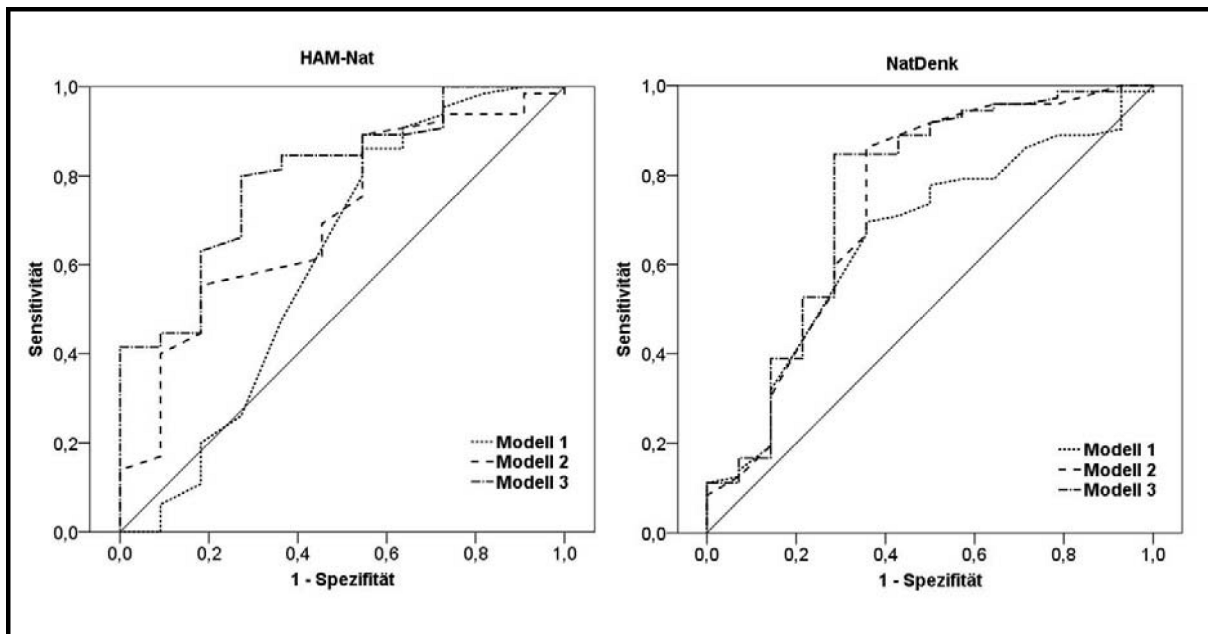


Abbildung 2: ROC Kurven für den HAM-Nat und den NatDenk für das Kriterium *Erfolg7Sem*, getrennt für die unterschiedlichen Regressionsmodelle (Modell 1: Abiturnote allein; Modell 2: Testergebnis allein; Modell 3: Abiturnote und Testergebnis mit Interaktion)

Auswahl können wir demnach mit einer höheren Treffsicherheit diejenigen identifizieren, die das Physikikum nicht schaffen, lehnen allerdings auch viele Studienbewerber ab, die erfolgreich gewesen wären.

Ein Maß für die Güte eines Tests ist die „area under the curve“ (AUC). Sie zeigt an, inwieweit ein Test in der Lage ist, zwei Gruppen (in unserem Fall „erfolgreich“ vs. „nicht erfolgreich“) voneinander zu trennen. Die AUCs für die verschiedenen Modelle und die beiden Tests sind in Tabelle 3 dargestellt. In beiden Testgruppen unterscheidet sich die AUC für die Abiturnote allein (Modell 1) nicht signifikant von 0,5, was einer Zufallsauswahl entsprechen würde. Sowohl für den Test allein, als auch für den Test und die Abiturnote plus deren Interaktion unterscheiden sich die Kurven signifikant von 0,5.

Aus Abbildung 2 wird deutlich, dass durch den HAM-Nat gerade bei einer hohen Selektionsquote, d.h. bei hoher Spezifität (links auf der x-Achse), mehr erfolgreiche Studenten ausgewählt werden würden als durch die Abiturnote allein (Modell 3 vs. Modell 1). Beim NatDenk zeigt sich gegenüber der Abiturnote eine höhere Sensitivität im Bereich geringerer Spezifität. Dieser Bereich steht für ein Auswahlverfahren mit vielen Studienplätzen und wenigen Bewerbern.

In der Gesamtstichprobe hatten 84,6% das Physikikum bis 2011 bestanden. Für das Kriterium *Erfolg4Sem* lag die Quote bei 67,3%. Wir haben hochgerechnet, welche Auswirkungen eine Auswahl der Studierenden nach verschiedenen Selektionsquoten (keine Selektion,  $\frac{3}{4}$ ,  $\frac{1}{2}$  oder  $\frac{1}{4}$  der Stichprobe) auf den *Erfolg7Sem* hätte. In beiden Gruppen war die Bestehensrate mit 85,5% (HAM-Nat) und 83,7% (NatDenk) ähnlich hoch (siehe Tabelle 4). Hätten wir nur die 25% Besten nach Testergebnis und Abiturnote zugelassen, wären in der HAM-Nat Gruppe alle Studierenden erfolgreich gewesen. In der NatDenk Gruppe wären es 90,5% gewesen. Hier wird noch einmal der Effekt deutlich, den man anhand der ROC-Kurven erkennen kann: für den HAM-Nat wäre es sinnvoll gewesen, eine strenge Selektionsquote anzulegen, während für den NatDenk eine Selektionsquote von 75% am günstigsten gewesen wäre.

## Diskussion

Die Studienabbruchquote ist laut Hochschul-Informationssystem (HIS) unter den Medizinstudierenden mit 5% relativ gering [8]. Dennoch müssen wir unter den gegebenen Umständen – es gibt wesentlich mehr Bewerber als Studienplätze – die potenziell geeigneten Studienbewerber auswählen. In Anbetracht der Tatsache, dass die Auswahlverfahren und die Studienabbruchquoten auch in den Fokus im Rahmen der Debatte um den Ärztemangel rücken, ist ein Ziel der Studierendenauswahl, diejenigen unter den Studienbewerbern zu identifizieren, die ihr Studium abschließen, denn dies ist die Voraussetzung dafür, den Arztberuf zu ergreifen. Da Studienabbruch schwierig zu erfassen ist, haben wir Studienerfolg als „Physikum nach 7 Semestern bestanden“ operationali-

siert. In unserer Stichprobe hatten ca. 15% dieses Kriterium nicht erfüllt. Wir haben uns für diese Operationalisierung entschieden, da die Erfahrung zeigt, dass nur sehr wenige Studierende das Physikikum noch nach dem 7. Semester absolvieren. Aufgrund der Dichotomie dieses Studienerfolgsparameters ist ein Vergleich mit den in der Einleitung berichteten Korrelationen zwischen den Testergebnissen und dem metrischen Studienerfolgsparameter Physikumsnote nicht ohne weiteres möglich. Die berichteten Odds Ratios erlauben jedoch einen Vergleich der Prädiktion der einzelnen Parameter.

Beide Tests, der HAM-Nat und der NatDenk, haben für sich genommen eine prädiktive Stärke für das Studienerfolgskriterium *Erfolg7Sem*. Wenn wir uns an der vom Hochschulrahmengesetz vorgegebenen Auswahlpraxis orientieren und das Testergebnis mit der Abiturnote kombinieren, unterscheiden sich die Tests jedoch. Während der HAM-Nat insbesondere in der Gruppe der Studienanfänger mit guten Abiturnoten zusätzliche prädiktive Informationen für den Studienerfolg liefert, differenziert der NatDenk besser zwischen den Bewerbern mit schlechteren Abiturnoten. Die Auswahl nach HAM-Nat und Abiturnote (Modell 3) hat unter den aktuellen Bedingungen des Mediziner Auswahlverfahrens mit vielen guten Bewerbern, aber nur wenigen Studienplätzen, die höchste Vorhersagekraft. In der Gruppe der 25% Besten nach Kombination der Abiturnote mit dem HAM-Nat Ergebnis haben 100% der Studienteilnehmer das Physikikum in sieben Semestern bestanden, bei Kombination mit dem NatDenk Ergebnis erhöht sich die Bestehensquote von unselektiert 85,5% lediglich auf 90,5%.

Eine Übertragung dieses Modells auf die Anwendungspraxis ist nur eingeschränkt möglich, auch wegen der geringen Stichprobengröße, der möglicherweise mangelnden Motivation der Teilnehmer, sich im Test anzustrengen und der Zusammensetzung der Stichprobe, in der auch Studierende der Abiturbesten- und Wartezeitquote enthalten waren. Um unsere Ergebnisse auf Allgemeingültigkeit zu prüfen, müssten diese an einer neuen Studierendekohorte validiert werden.

Ziel der Studie war zu untersuchen, ob der HAM-Nat ähnliche Zusammenhänge zum Studienerfolg aufweist wie das Testmodul NatDenk. Wir wollen keine Aussagen über den TMS im Ganzen treffen. Bestätigt hat sich die Annahme, dass der HAM-Nat und der NatDenk auf unterschiedliche Konstrukte zurückgreifen. Wir sehen einen Vorteil des HAM-Nat in der relativen Unabhängigkeit des Testergebnisses von der Abiturnote. Wir haben somit ein Kriterium eingeführt, welches insbesondere im Bereich der Spitzenleistungen eine Differenzierung von Studienbewerbern erlaubt und auch in dieser schon stark vorselektierten Stichprobe noch prädiktive Stärke für den Studienerfolg besitzt.

## Danksagung

Wir danken dem Dekan Prof. U. Koch-Gromus und Herrn Dr. B. Andresen für Anregungen und lebhaftes Diskussio-

Tabelle 3: Fläche unter der Kurve (AUC) für die verschiedenen Prädiktionsmodelle, getrennt nach HAM-Nat und NatDenk

Modell	Prädiktoren	Erfolg7Sem AUC (p)	
		HAM-Nat	NatDenk
1	Abiturnote	0,61 (0,246)	0,66 (0,060)
2	Test	0,70 (0,032)	0,73 (0,006)
3	Test, Abiturnote, Interaktion	0,77 (0,002)	0,75 (0,003)

Tabelle 4: Hochrechnung des Anteils der erfolgreichen Studierenden nach unterschiedlichen Selektionsquoten

Selektions- quote	HAM-Nat und Abitur Erfolg7Sem				NatDenk und Abitur Erfolg7Sem			
	N	ja	nein	%	N	ja	nein	%
Keine Selektion	76	65	11	85.5	86	72	14	83.7
75%	56	52	4	92.9	64	60	4	93.8
50%	38	36	2	94.7	43	39	4	90.7
25%	19	19	0	100	21	19	2	90.5

nen und ihre Unterstützung, Herrn D. Münch-Harrach und Herrn C. Kothe für ihre Unterstützung bei der Datenverarbeitung sowie Herrn N. Feddersen für Hilfe bei der Übersetzung. Diese Studie wird durch den Förderfonds Lehre des Dekanates der Medizinischen Fakultät Hamburg unterstützt.

## Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

## Literatur

- Hampe W, Hissbach J, Kadmon M, Kadmon G, Klusmann D, Scheutzel P. Wer wird ein guter Arzt? Verfahren zur Auswahl von Studierenden der Human- und Zahnmedizin. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz. 2009;52(8):821-830. DOI: 10.1007/s00103-009-0905-6
- Deutscher Bundestag. Hochschulrahmengesetz. BGBl. 2005;l:3835. Zugänglich unter/available from: [http://www.bmbf.de/pub/HRG\\_20050126.pdf](http://www.bmbf.de/pub/HRG_20050126.pdf)
- Trost G, Flum F, Fay E, Klieme E, Maichle U, Meyer M, Nauels HU. Evaluation des Tests für Medizinische Studiengänge (TMS): Synopse der Ergebnisse. Bonn: ITB; 1998.
- Trost G. Test für Medizinische Studiengänge (TMS): Studien zur Evaluation, 20. Arbeitsbericht: Institut für Test- und Begabungsforschung. Bonn: ITB; 1996.
- Trapmann S, Hell B, Weigand S, Schuler H. Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. Z Padagog Psychol. 2007;21(1):11-27. DOI: 10.1024/1010-0652.21.1.11

- Hampe W, Klusmann D, Buhk H, Muench-Harrach D, Harendza S. Reduzierbarkeit der Abbrecherquote im Humanmedizinstudium durch das Hamburger Auswahlverfahren für Medizinische Studiengänge - Naturwissenschaftsteil (HAM-Nat). GMS Z Med Ausbild. 2008;25(2):Doc82. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000566.shtml>
- Hissbach J, Klusmann D, Hampe W. Reliabilität des Hamburger Auswahlverfahrens für Medizinische Studiengänge, Naturwissenschaftsteil (HAM-Nat). GMS Z Med Ausbild. 2011;28(3):Doc44. DOI: 10.3205/zma000756
- Heublein U, Schmelzer R, Sommer D, Wank J. Die Entwicklung der Schwund- und Studienabbruchquoten an den deutschen Hochschulen. Hannover: HIS Hochschul-Informationssystem; 2008.

### Korrespondenzadresse:

Prof. Dr. Wolfgang Hampe  
Universitätsklinikum Hamburg Eppendorf (UKE), Institut für Biochemie und Molekulare Zellbiologie, Martinistraße 52, 20246 Hamburg, Deutschland  
[hampe@uke.uni-hamburg.de](mailto:hampe@uke.uni-hamburg.de)

### Bitte zitieren als

Hissbach J, Feddersen L, Sehner S, Hampe W. Eignung von HAM-Nat und TMS-Modul "Medizinisch-naturwissenschaftliches Grundverständnis" für die Studienbewerbersauswahl in der Medizin. GMS Z Med Ausbild. 2012;29(5):Doc72.  
DOI: 10.3205/zma000842, URN: urn:nbn:de:0183-zma0008422

### Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2012-29/zma000842.shtml>

**Eingereicht:** 30.11.2011

**Überarbeitet:** 20.06.2012

**Angenommen:** 21.06.2012

**Veröffentlicht:** 15.11.2012

**Copyright**

©2012 Hissbach et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.



# Suitability of the HAM-Nat test and TMS module "basic medical-scientific understanding" for medical school selection

## Abstract

**Aims:** Tests with natural-scientific content are predictive of the success in the first semesters of medical studies. Some universities in the German speaking countries use the 'Test for medical studies' (TMS) for student selection. One of its test modules, namely "medical and scientific comprehension", measures the ability for deductive reasoning. In contrast, the Hamburg Assessment Test for Medicine, Natural Sciences (HAM-Nat) evaluates knowledge in natural sciences. In this study the predictive power of the HAM-Nat test will be compared to that of the NatDenk test, which is similar to the TMS module "medical and scientific comprehension" in content and structure.

**Methods:** 162 medical school beginners volunteered to complete either the HAM-Nat (N=77) or the NatDenk test (N=85) in 2007. Until spring 2011, 84.2% of these successfully completed the first part of the medical state examination in Hamburg. Via different logistic regression models we tested the predictive power of high school grade point average (GPA or "Abiturnote") and the test results (HAM-Nat and NatDenk) with regard to the study success criterion "first part of the medical state examination passed successfully up to the end of the 7<sup>th</sup> semester" (Success7Sem). The Odds Ratios (OR) for study success are reported.

**Results:** For both test groups a significant correlation existed between test results and study success (HAM-Nat: OR=2.07; NatDenk: OR=2.58). If both admission criteria are estimated in one model, the main effects (GPA: OR=2.45; test: OR=2.32) and their interaction effect (OR=1.80) are significant in the HAM-Nat test group, whereas in the NatDenk test group only the test result (OR=2.21) significantly contributes to the variance explained.

**Conclusions:** On their own both HAM-Nat and NatDenk have predictive power for study success, but only the HAM-Nat explains additional variance if combined with GPA. The selection according to HAM-Nat and GPA has under the current circumstances of medical school selection (many good applicants and only a limited number of available spaces) the highest predictive power of all models.

**Keywords:** HAM-Nat, student selection, medicine, test for natural science, TMS

Johanna Hissbach<sup>1</sup>

Lena Feddersen<sup>1</sup>

Susanne Sehner<sup>1,2</sup>

Wolfgang Hampe<sup>1</sup>

1 Universitätsklinikum  
Hamburg Eppendorf (UKE),  
Institut für Biochemie und  
Molekulare Zellbiologie,  
Hamburg, Deutschland

2 Universitätsklinikum  
Hamburg Eppendorf (UKE),  
Institut für medizinische  
Biometrie und Epidemiologie,  
Hamburg, Deutschland

## Authorship

Johanna Hissbach and Lena Feddersen contributed equally to this work.

## Introduction

University applicants for medicine are confronted with a variety of different admission procedures in Germany [1]. While GPA is legally prescribed as decisive selection cri-

terion for all universities [2], other criteria as tests, interviews or special achievements can be considered by single faculties. The best known German test for university medical courses entry is the "Test For Medical Studies" (TMS) [3]. After the TMS ceased to be applied in Germany post 1996, Baden-Wuerttembergian universities began to consider good test results in admission process in 2007. In the winter semester of 2011/12 10 out of 34 German medical schools granted bonus points for good TMS results.

Since 2008 the "Hamburg Assessment Test for Medicine, Natural Sciences" (HAM-Nat) is applied at the University hospital Hamburg-Eppendorf. HAM-Nat and TMS were

specifically developed for the selection process of medical students, but at the same time each test aims at different constructs. The HAM-Nat test aims to examine knowledge of natural science on a higher education level, as relevant to medical studies. The TMS test includes the test module "medical and scientific comprehension" with questions on natural sciences, these however should indicate "deductive reasoning in subject specific areas and contexts", for which no specialised knowledge is required [4]. This test module is one of four test parts related to medicine, on which the predictive efficiency of the TMS is based [3]. To be prepared for the HAM-Nat, motivated applicants repeat high school topics in the field of natural sciences which at the same time means preparation for the first semesters. In contrast the preparation for the TMS has no relation to the content of the university curriculum and, according to the developers of the test, will only lead to minor improvements of results [3].

The HAM-Nat was developed with the aim of reducing the university dropout rate during the first part of the course. Since GPA (Abiturdurchschnittsnote) shows a corrected predictive power of  $r=0.58$  [5] for university marks during the pre-clinical part of the course, it is a solid predictor for success in the course (results of the first part of the medical state examination), and the question arises what benefit is gained if further tests for medical school selection are applied.

HAM-Nat as well GPA were significant predictors for success in course and studies after the second semester of the 2006 cohort in Hamburg ( $r=0.31$  or respectively  $r=0.26$ ), thus 13% of the total variance were explained [6]. In this study success was operationalized as number of passed examinations. The contribution of HAM-Nat to the explained variance (9.5%) was higher than the single contribution of the GPA (6.6%). Considering a sub-sample, corresponding to the quota of places allocated by the university itself, no significant correlation of the GPA with the HAM-Nat can be found ( $r=0.11$ ) due to variance restriction of the GPA [6] indicating that differences in the range of very good GPAs (1.2 -1.7) have no substantial influence on HAM-Nat test results. Due to the variance restriction, GPA no longer has any significant correlation to study success within this group. In contrast, the predictive power of the HAM-Nat in this group falls only slightly from  $r=0.31$  to  $r=0.26$ .

The results of different HAM-Nat test versions from pilot studies with university beginners in 2006 and 2007 correlate with GPA between  $r=0.12$  and  $r=0.34$  [6], [7]. The TMS-module "medical and scientific comprehension" belongs to a group of TMS modules correlating with GPA between  $r=0.28$  and  $r=0.40$  [3]. The correlation of the entire TMS with GPA is higher with  $r=0.36$  up to  $r=0.48$ . Notwithstanding the TMS developers conclude that school and test measure different aspects of achievement [3], [4]. In the TMS module "medical and scientific comprehension" the ability is captured to quickly extradite the essential information from written material and, based on this, to draw the correct conclusions (p. 53 in [3]). Studies show that essential for this are an internal re-

presentation of the text information in the working memory, the construction of an interference chain, and recoding of text statements into an imaginative picture. The correlation between the HAM-Nat test with the test module "scientific reasoning" (NatDenk) was explored in context of a second study before the introduction of the HAM-Nat into the Hamburg selection process. This module is structurally and textually similar to the TMS module "medical and scientific comprehension". The correlation between HAM-Nat and NatDenk was, respective to the test-version,  $r=0.34$  or  $r=0.21$  [7]; thus both tests represent different constructs.

Medical faculties in Austria, Germany, and Switzerland use both knowledge tests like the HAM-Nat as well as ability tests, such as the TMS, in their selection process. The aim is always the optimisation of study success. In this study we compare the predictive efficiency of both tests with regard to study success. As parameter for study success we chose the successful completion of the first part of the medical state examination within the first seven semesters (Success7Sem), since only very few students complete their medical studies at Hamburg University later. The second criterion chosen was the successful completion of the first part of the medical state examination within the first four semesters (Success4Sem) since many universities aim at completion within the standard period of study.

## Methods

### HAM-Nat

The 2007 version of the HAM-Nat consists of 60 multiple choice questions in the field of mathematics, chemistry, physics and biology. The content is formed from thematic areas relevant to medicine and is of German high school standard. Covered topics as well as a self-test consisting of questions from the 2006 and 2007 test versions are available on the webpage of the University Hospital Eppendorf (<http://www.uke.de/studienbewerber>). One out of five answer choices is correct and participants were allowed 1.5 minutes for answering each question. The questions were developed by secondary school teachers and lecturers of the clinical and theoretical subjects of the medical faculty.

### Test-module "scientific reasoning"

For the comparison of the HAM-Nat with the TMS-module "medical and scientific comprehension" as an external criterion, ITB-Consulting GmbH (developers of the TMS) developed a test module ("scientific reasoning", NatDenk) consisting of 24 multiple-choice questions, similar to the TMS-module in content and structure. One of 5 answer choices for each question fits to an aforementioned natural-science issue. For all 24 questions the participants are allowed a total of 55 minutes. The tasks do not require specific natural-science knowledge, but instead are

aimed at comprehensive understanding and the ability of deductive reasoning. The right to conduct the test module was obtained from ITB-Consulting.

## Study success

Study success was operationalized as "first part of the medical state examination passed up to the end of the 7<sup>th</sup> semester" (*Success7Sem*). This dichotomous criterion was chosen as main criterion since reducing the number of university dropouts in the pre-clinical study phase is the primary aim of the HAM-Nat. The success probability after the 7<sup>th</sup> semester is very low. To verify the stability of the models an additional dichotomous criterion for study success was chosen – passing the same examination in the standard period of study after four semesters (*Success4Sem*) - and the correlation between both criteria was calculated. The latter criterion is used specifically in cases when the focus is set on monetary aspects, as scheduled progression of study is financially efficient. For sake of clarity the results for the criterion *Success4Sem* are only discussed if they differ from the main criterion *Success7Sem*.

## Study Design

The study was part of a wider study on the parallel- and retest-reliability of the HAM-Nat in 2007 [7]. Participation in the study was offered to incoming students of medicine during the orientation phase. Four weeks after the start of the semester all participants were randomly assigned to one of two groups: one group took the HAM-Nat, while the other took the test module "scientific reasoning" (NatDenk). Participation in the study was voluntary, the actual taking of the tests followed a compulsory course. It was organised by members of our research group and was supervised by lecturers of the medical faculty. The test "scientific reasoning" was especially conducted for the purpose of the study. The results of the tests were merged with study-success data in May 2011.

## Sample

The study used a sample of university beginners of the year 2007, which had been chosen due to GPA or had been accepted due to other quotas (waiting time since school graduation, foreigner quota). In the GPA quota only applicants with a GPA < 1.6 (Abiturdurchschnittsnote) were admitted. Only data of participants with known GPA were included. The total sample, on which the analysis is based, comprised 162 university beginners. This sample was representative of the entity of university beginners in the orientation phase in terms of gender, age and GPA. Half of the participant (N=77) took the 2007 HAM-Nat, the other half (N=85) took the module "scientific reasoning" (NatDenk) a week later. All participants gave written consent to the use of their data.

## Statistical Analysis

To depict test and GPA on one scale both were z-transformed. The predictive power of the GPA as well as of the test results (HAM-Nat and NatDenk) in relation to the study success criteria *Success7Sem* and *Success4Sem* were examined via three different logistic regression models. Within-subject factors were each GPA and the test results; between-subject factors were the affiliation to the test group (HAM-Nat vs. NatDenk). The odds ratios and 95% confidence intervals are given. IBM SPSS Statistics, Version 19.0.0 was used for analysis.

## Results

### Sample

67.3% of all participants successfully completed the first part of the medical state examination in the standard period of study (*Success4Sem*), by spring 2011 (*Success7Sem*) altogether 84.6% had succeeded. Table 1 gives an overview of characteristics of the total sample and both sub-samples.

While GPA was significantly related to the overall score of the examination ( $r=0.24$ ;  $p=0.004$ ) and the Ham-Nat ( $r=-0.24$ ;  $p=0.038$ ), it was not significantly correlated to the NatDenk ( $r=-0.11$ ;  $p=0.324$ ). The negative correlation values show that good (low) GPA is associated with good (high) test results.

### Prediction

The results were evaluated in several steps. First, GPA and test results were examined separately in how far they are related to study success. Subsequently they were incorporated in one overall model which also included the interaction between both criteria.

#### Model 1: GPA

A significant correlation between GPA and study success existed in the total-sample (Odds Ratios: for *Success7Sem* OR=1.72;  $p=0.006$  and for *Success4Sem*, OR=1.98;  $p=0.000$ ). An OR of 1.72 signifies that the chances to have passed the examination after seven semesters increase by 1.72-fold with every improvement in the GPA by approximately 0.6 points (1 standard deviation).

If the division into separate test groups is considered, GPA has an effect on *Success7Sem* which is significant for the HAM-Nat group and ranges just over the significance threshold of  $p=0.05$  for the NatDenk group (see Table 2). Figure 1 depicts this effect: 'GPA' as independent variable was dichotomised at the median into good (GPA < 1.6) vs. bad (GPA ≥ 1.6), and the estimated probability for study success was illustrated separately for both groups. For the criterion *Success4Sem* significant correl-

Table 1: Sample description

	Total (N=162)	Sample HAM-Nat (N=77)	NatDenk (N=85)	HAM-Nat vs. NatDenk
	N (%)	N (%)	N (%)	$\chi^2$ (p)
Gender female	108 (66,7 %)	48 (62,3 %)	57 (67,1 %)	0.17 (0.678)
Success4Sem	109 (67,3 %)	57 (75,0 %)	52 (60,5 %)	3.87 (0.049)
Success7Sem	137 (84,6 %)	66 (85,7 %)	71 (83,5 %)	0.10 (0.751)
	M (SD)	M (SD)	M (SD)	T (p)
Age	21.59 (3.64)	20.63 (2.55)	22.43 (4.21)	3.23 (0.001)
Abitur results	1.80 (0.64)	1.70 (0.53)	1.89 (0.71)	1.87 (0.063)
Test results (%)	-	40.44 (12.70)	58.38 (17.12)	-
Physikumsgesamtnote	2.73 (0.75) <sup>a</sup>	2.62 (0.71) <sup>b</sup>	2.83 (0.78) <sup>c</sup>	1.59 (0.115)

<sup>a</sup> N=137, <sup>b</sup> N=66, <sup>c</sup> N=71

Table 2: Logistic regression with the criteria Success7Sem and Success4Sem

Model Predictors	Success7Sem OR (p)		Success4Sem OR (p)	
	HAM-Nat	NatDenk	HAM-Nat	NatDenk
1 GPA	1.84 (0.031)	1.62 (0.078)	1.74 (0.027)	2.30 (0.001)
2 Test results	2.07 (0.027)	2.58 (0.007)	2.36 (0.006)	1.87 (0.009)
3 Test results	2.32 (0.028)	2.21 (0.029)	2.37 (0.009)	1.80 (0.025)
GPA	2.45 (0.010)	1.36 (0.393)	1.92 (0.029)	2.51 (0.002)
Interaction GPA & test	1.80 (0.037)	0.74 (0.370)	1.48 (0.119)	0.72 (0.270)

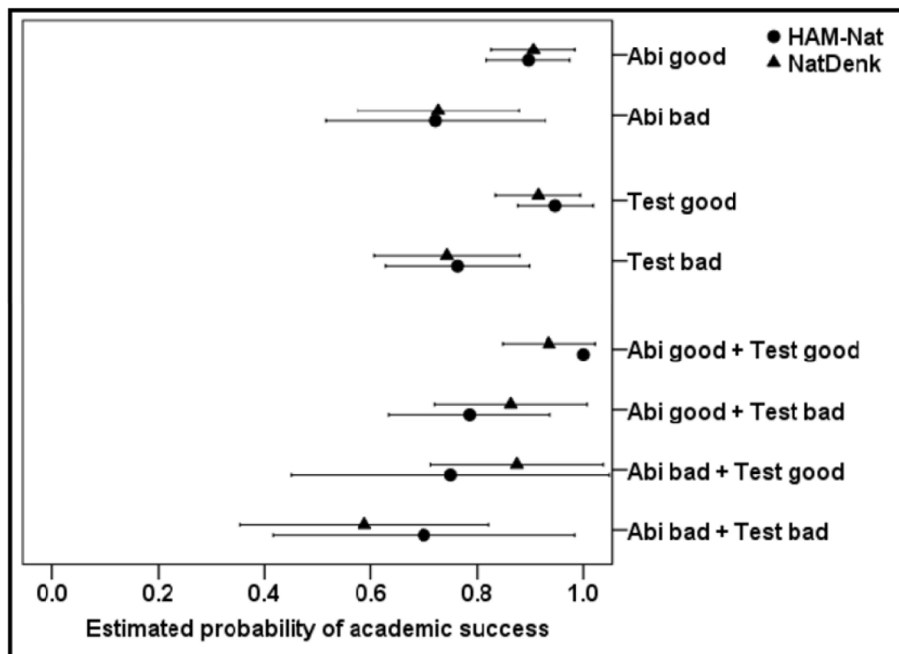


Figure 1: Estimated probability of Success7Sem with relation to GPA and test result. Both predictors were dichotomised at the median (GPA (=Abi) good: <1.6 and test good: >median). The quality of a predictor can be read from the difference between study success probability of good and bad manifestations. For instance, the study success probability of participants with a good GPA is significantly higher for good than for bad HAM-Nat participants. For the NatDenk this difference is considerably smaller.

ation existed between Abitur results and study success in both groups (see Table 2).

**Model 2: Test results (HAM-Nat vs. NatDenk)**

A significant correlation exists between the results of both tests and study success if GPA is not considered (see

Table 2). With regard to the HAM-Nat the probability to successfully pass the examination doubled per standard derivation of the test result (see Figure 1). For the NatDenk the probability increased slightly more by the factor 2.58. No significant difference existed between the ORs (p=0.909).

### Model 3: GPA and test results (HAM-Nat vs. NatDenk) with their interaction term

If the effects of both admission criteria as well as their interaction term are estimated in one model, both main effects and their interaction are significant in the HAM-Nat group, whereas in the NatDenk group only the test result provides a significant contribution to the explained variance (see Table 2). This means that with the HAM-Nat test the predictive efficiency depends on the GPA of the tested individual. Especially in the case of a good GPA, the HAM-Nat provides additional information of study success (see Figure 1). On the other hand, GPA does not provide any further information in addition to the NatDenk.

For the criterion *Success4Sem* the interaction between GPA and HAM-Nat is insignificant, and the GPA loses a little of its predictive power. In the NatDenk group now both predictors contribute significantly to the prediction of study success. However, the influence of the GPA is higher than that of the test results.

### Sensitivity and Specificity

The Receiver Operating Characteristic (ROC) shows the relation between sensitivity and specificity of a test (see Figure 2). ROC-curves can help to illustrate possible effects for different selection quotas.

High sensitivity in our case means to reject as few as possible applicants who can study successfully. The highest sensitivity could be achieved by admittance of all applicants, which corresponds to a selection quota of 100% (thus no selection, right side of the x-axis in Figure 2). All students able to complete the examination would have been given a chance, no one would have been rejected wrongly. However, this also means that everyone not completing the examination would also have been accepted (low specificity, also on the right side of the x-axis).

If our priority is to identify university dropouts, a specific test is required. Laying a stricter selection quota (we move further to the left on the x-axis) results in a reduced sensitivity while specificity increases. Therefore, the accuracy of identifying the ones not passing the examination rises with stricter selection, while at the same time numerous university applicants, who would have passed, would also have been rejected.

A measure of a test's quality is the "area under the curve" (AUC). It shows in how far a test is able to divide two groups from each other (in our case 'successful' vs. 'unsuccessful'). The AUCs for the different models and both tests are depicted in table 3. In both test groups the AUC for Model 1 (GPA alone) does not differ significantly from 0.5 – corresponding to random selection. In contrast, for the tests alone (Model 2), and for the combinations of the tests with GPA (Model 3) the AUC differ significantly from 0.5.

Figure 2 shows that particularly in case of a strict selection quota (high specificity, left on the x-axis) by HAM-Nat

test results, we identify more successful students as compared to selection based on Abitur results only (Model 3 vs. Model 1). The NatDenk shows a higher sensitivity at lower specificity compared to the Abitur results. This situation stands for a selection process with many available study places and few applicants.

84.6% of the total sample had passed the first part of the medical state examination after 7 semesters, after 4 semesters 67.3% succeeded. We extrapolated the influence of different selection quotas (no selection,  $\frac{3}{4}$ ,  $\frac{1}{2}$  or  $\frac{1}{4}$  of the sample) on *Success7Sem*. In both groups the pass rate was similarly high, namely 85.5% for HAM-Nat and 83.7% for NatDenk (see Table 4). In the HAM-Nat group all students would have been successful if only the top 25% judged by test result and GPA would have been accepted. In the NatDenk group this would have produced a success rate of 90.5%. Here again the effect seen in the ROC-curves is highlighted: a strict selection quota would have been useful for the HAM-Nat, whereas a selection quota of 75% would have been most beneficial for the NatDenk.

### Discussion

According to the Hochschul-Informationssystem (HIS) the dropout quota for medical students is, with 5%, relatively small [8]. Nevertheless, due to the fact that the number of applicants clearly exceeds the available places, the potentially best-suited applicants have to be selected. The selection of medical students and dropout quotas increasingly receive attention in debates on the shortage of medical practitioners. Since the completion of the medical studies is a requirement for becoming a qualified doctor, one aim of the selection process is to identify applicants which will successfully complete university. Because dropping out of university is hard to gather numerically, we operationalised the successful completion of studies as "first part of the medical state examination passed after 7 semesters". In our case 15% of the sample did not meet this criterion. This operationalisation was chosen since only a very limited number of students pass after the 7<sup>th</sup> semester. Due to the dichotomy of this parameter a comparison with the correlation between test results and the metric study success parameter "grade in the first part of the medical state examination" given in the introduction cannot be made offhand. The reported odds ratios, however, allow for a comparison of the prediction of the individual parameters.

Both the HAM-Nat and the NatDenk, taken for themselves, have predictive power for the criterion *Success7Sem*. However, both tests do differ if we consider the selection practice as regulated by the 'Hochschulrahmengesetz' (Framework Act for Higher Education in Germany) which regulates to combine test results with GPA. While the HAM-Nat offers additional predictive information particularly on the study success of students with good GPA, the NatDenk allows for better differentiation within the group of applicants with worse GPA. Under the current circum-

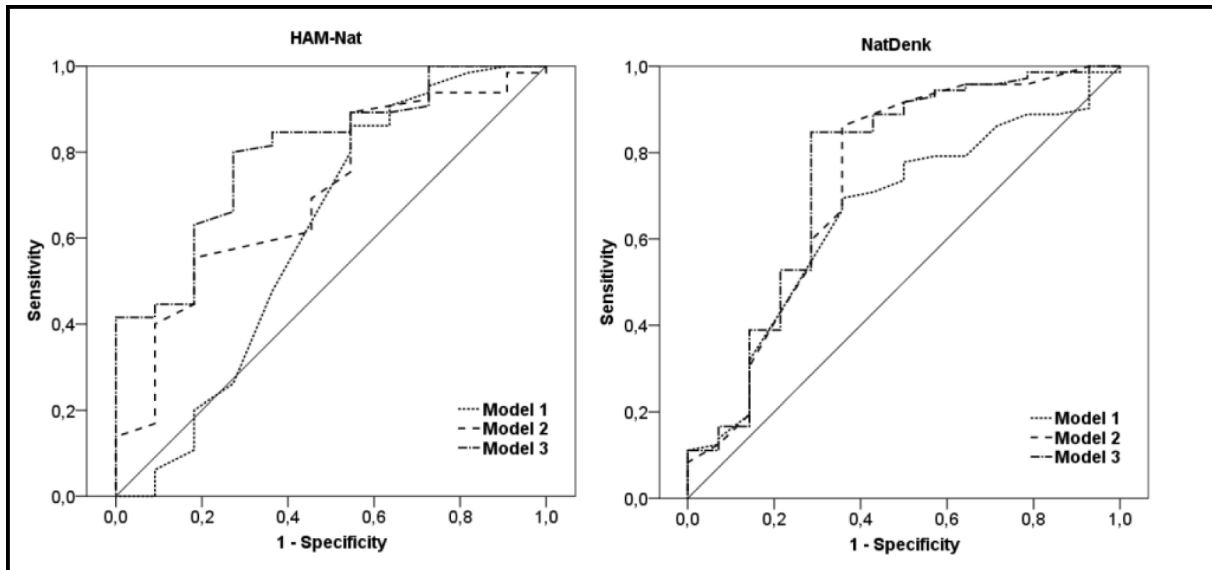


Figure 2: ROC-curves for the criterion *Success7Sem* for HAM-Nat and NatDenk, separated by different regression models (Model 1: GPA alone; Model 2: test result alone; Model 3: GPA, test result, and interaction)

Table 3: Area under the curve (AUC) for the different models of prediction, indicated for HAM-Nat and NatDenk

Model	Predictors	<i>Success7Sem</i> AUC (p)	
		HAM-Nat	NatDenk
1	GPA	0.61 (0.246)	0.66 (0.060)
2	Test	0.70 (0.032)	0.73 (0.006)
3	Test, GPA, interaction	0.77 (0.002)	0.75 (0.003)

Table 4: Extrapolation of the success rate under the influence of different selection quotas.

Selection quota	HAM-Nat and GPA <i>Success7Sem</i>				NatDenk and GPA <i>Success7Sem</i>			
	N	yes	no	%	N	yes	no	%
no selection	76	65	11	85.5	86	72	14	83.7
75%	56	52	4	92.9	64	60	4	93.8
50%	38	36	2	94.7	43	39	4	90.7
25%	19	19	0	100	21	19	2	90.5

stances of medical school selection (many good applicants and a very limited number of available spaces) a selection process combining HAM-Nat and GPA (Model 3) has the highest predictive efficiency. In the group of the top 25% participants in our study (combination of GPA and HAM-Nat), 100% successfully passed the examination in 7 semesters. In combination with the NatDenk results, however, the success quota only improved from non-selected 85.5% to 90.5%.

Transfer of this model to the de facto selection process is not fully possible due to several factors: the small size of the sample, the potential lack of motivation of the participants to endeavour, and the make-up of the sample which included students from groups whose admission e.g. by waiting time or excellent GPA alone can not be influenced by the university. To test the validity of the

results, they will have to be replicated with a new cohort of students.

Aim of the study was to investigate whether the HAM-Nat test shows similar correlations with study success as the test module NatDenk. No assertions of the TMS in its entirety are given. The assumption has been confirmed that HAM-Nat and NatDenk both relate to different constructs. We consider the HAM-Nat results' relative independence of the GPA an advantage. Thereby, we introduced a criterion which, especially in the group of top performers, enables a differentiation and has predictive power for study success, even within this highly selective sample.

## Acknowledgement

The authors thank the dean of the medical school Prof. Dr. U. Koch-Gromus and Dr. B. Andresen for helpful suggestions and discussions as well as their support. We are much obliged to D. Münch-Harrach and C. Kothe for assisting us with the data handling. N. Feddersen was very helpful during the translation process. This study was supported by the "Förderfonds Lehre des Dekanates der Medizinischen Fakultät Hamburg".

## Competing interests

The authors declare that they have no competing interests.

## References

1. Hampe W, Hissbach J, Kadmon M, Kadmon G, Klusmann D, Scheutzel P. Wer wird ein guter Arzt? Verfahren zur Auswahl von Studierenden der Human- und Zahnmedizin. Bundesgesundheitsbl Gesundheitsforsch Gesundheitsschutz. 2009;52(8):821-830. DOI: 10.1007/s00103-009-0905-6
2. Deutscher Bundestag. Hochschulrahmengesetz. BGBl. 2005;l:3835. Zugänglich unter/available from: [http://www.bmbf.de/pub/HRG\\_20050126.pdf](http://www.bmbf.de/pub/HRG_20050126.pdf)
3. Trost G, Flum F, Fay E, Klieme E, Maichle U, Meyer M, Nauels HU. Evaluation des Tests für Medizinische Studiengänge (TMS): Synopse der Ergebnisse. Bonn: ITB; 1998.
4. Trost G. Test für Medizinische Studiengänge (TMS): Studien zur Evaluation, 20. Arbeitsbericht: Institut für Test- und Begabungsforschung. Bonn: ITB; 1996.
5. Trapmann S, Hell B, Weigand S, Schuler H. Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse. Z Padagog Psychol. 2007;21(1):11-27. DOI: 10.1024/1010-0652.21.1.11
6. Hampe W, Klusmann D, Buhk H, Muench-Harrach D, Harendza S. Reduzierbarkeit der Abbrecherquote im Humanmedizinstudium durch das Hamburger Auswahlverfahren für Medizinische Studiengänge - Naturwissenschaftsteil (HAM-Nat). GMS Z Med Ausbild. 2008;25(2):Doc82. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000566.shtml>
7. Hissbach J, Klusmann D, Hampe W. Reliabilität des Hamburger Auswahlverfahrens für Medizinische Studiengänge, Naturwissenschaftsteil (HAM-Nat). GMS Z Med Ausbild. 2011;28(3):Doc44. DOI: 10.3205/zma000756
8. Heublein U, Schmelzer R, Sommer D, Wank J. Die Entwicklung der Schwund- und Studienabbruchquoten an den deutschen Hochschulen. Hannover: HIS Hochschul-Informationssystem; 2008.

### Corresponding author:

Prof. Dr. Wolfgang Hampe  
Universitätsklinikum Hamburg Eppendorf (UKE), Institut für Biochemie und Molekulare Zellbiologie, Martinistraße 52, 20246 Hamburg, Deutschland  
[hampe@uke.uni-hamburg.de](mailto:hampe@uke.uni-hamburg.de)

### Please cite as

Hissbach J, Feddersen L, Sehner S, Hampe W. Eignung von HAM-Nat und TMS-Modul "Medizinisch-naturwissenschaftliches Grundverständnis" für die Studienbewerberauswahl in der Medizin. GMS Z Med Ausbild. 2012;29(5):Doc72. DOI: 10.3205/zma000842, URN: urn:nbn:de:0183-zma0008422

### This article is freely available from

<http://www.egms.de/en/journals/zma/2012-29/zma000842.shtml>

**Received:** 2011-11-30

**Revised:** 2012-06-20

**Accepted:** 2012-06-21

**Published:** 2012-11-15

### Copyright

©2012 Hissbach et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>). You are free: to Share — to copy, distribute and transmit the work, provided the original author and source are credited.