

Inhaltsverzeichnis

KERMIT und die zentralen Prüfungen: Eng verwandt und doch verschieden	2
Das Konzept.....	2
Die Aufgabenentwicklung	4
Die Durchführung	5
Die Auswertung.....	5
Die Ergebnisse	6
Fazit: Eng verwandt und doch verschieden.....	8

KERMIT und die zentralen Prüfungen: Eng verwandt und doch verschieden

(Langfassung des gleichnamigen Artikels in „Hamburg macht Schule“, Heft 1/2016, S. 44-45)

In einigen Hamburger Schulen wird vonseiten der Lehrkräfte und Schulleitungen angemerkt, dass sich KERMIT-Ergebnisse und die Resultate aus den Hamburger zentralen Prüfungen (z. B. zum mittleren Schulabschluss, MSA) nicht immer entsprechen. Kolleginnen und Kollegen des Instituts für Bildungsmonitoring und Qualitätsentwicklung haben dieses Phänomen zusammen mit den Fachreferentinnen und -referenten im Amt für Bildung sowie an der Aufgabenentwicklung beteiligten Lehrerinnen und Lehrern untersucht. Dabei zeigt sich, dass beide Formen der schulischen Leistungsmessung statistisch zum Teil eng zusammenhängen. Warum Abweichungen bei einzelnen Schülerinnen und Schülern, bei Klassen und sogar auf Ebene der Schule dennoch möglich sind, zeigt der folgende Vergleich zwischen KERMIT und den zentralen Prüfungen. Er gliedert sich in die Bereiche Konzept, Aufgabenentwicklung, Durchführung und Auswertung. Im Abschnitt Ergebnisse werden die Abweichungen beispielhaft für die KERMIT 9-Erhebung des Jahres 2014 und die schriftlichen Überprüfungen in Klasse 10 an Gymnasien aus dem Jahr 2015 beschrieben.

Das Konzept

Das Instrumentarium „Kompetenzen ermitteln“ (KERMIT) wird seit knapp vier Jahren in Hamburg flächendeckend eingesetzt, um die Leistungen der Schülerinnen und Schüler in den Bereichen Deutsch Leseverstehen, Mathematik, Naturwissenschaften sowie in Englisch Hör- und Leseverstehen mit standardisierten Fachleistungstests zu erfassen. Zu den *zentralen Prüfungen* zählen nicht nur die *Abschlussprüfungen* zum ersten allgemeinbildenden und mittleren Schulabschluss (ESA und MSA) sowie zum Abitur, sondern auch die so genannte *schriftliche Überprüfung* im Jahrgang 10 an den Gymnasien. Es handelt sich hierbei um eine zentral gestellte Klassenarbeit in den Fächern Deutsch, Mathematik sowie der ersten Fremdsprache. In mindestens zwei der Fächer wird die schriftliche Überprüfung durch eine mündliche Prüfung ergänzt. Die Gesamtnoten gehen mit 30 Prozent in die Zeugnisnote des jeweiligen Faches ein.

Sowohl KERMIT als auch die zentralen Abschlussprüfungen und schriftlichen Überprüfungen beziehen sich auf die Hamburger Bildungspläne und damit auf die nationalen Bildungsstandards. Diese werden seit dem Jahr 2003 von der Kultusministerkonferenz (KMK) entwickelt und geben Lehrziele für die bundesdeutsche Bildung vor. Trotz dieser gemeinsamen Basis gibt es Unterschiede in der Zielsetzung und den Inhalten beider Verfahren sowie in der Vorbereitung der Schülerinnen und Schüler.

KERMIT und die zentralen Prüfungen verfolgen zwei unterschiedliche **Ziele**: Mit den zentralen Prüfungen werden die Leistungen der Schülerinnen und Schüler festgestellt, um jedem Einzelnen ein Zertifikat mit hoher Relevanz für den weiteren Bildungsweg auszustellen. Demgegenüber stellt

KERMIT Informationen bereit, die es den Lehrkräften ermöglichen, ihren Unterricht weiterzuentwickeln. Auch die Schulentwicklung soll durch KERMIT auf eine solide Datenbasis gestellt werden. Die Schülerinnen und Schüler können ihr Ergebnis bei Interesse zwar erfahren, dieser Service ist aber vor allem als Anerkennung und zur Motivation der Testteilnehmerinnen und Teilnehmer gedacht. Für eine verlässliche Individualbeurteilung werden die wiederholten Ergebnisse und Beobachtungen aus den Klassenarbeiten sowie der täglichen Mitarbeit der Schülerinnen und Schüler herangezogen.

Ein weiterer Unterschied zwischen KERMIT und den zentralen Prüfungen liegt in den **Inhaltsbereichen** der jeweiligen Leistungsmessung. KERMIT konzentriert sich meist auf Teilbereiche eines Schulfachs, die als Voraussetzung für erfolgreiche Lernprozesse von großer Bedeutung sind. Dazu gehört im Fach Deutsch das Leseverstehen und im Fach Englisch zusätzlich das Hörverstehen.

Zentrale Prüfungen gehen darüber hinaus. Insbesondere im Fach Deutsch und der ersten Fremdsprache ist die Produktion von Texten ein zentrales Bildungsziel und deshalb ein wesentlicher Bestandteil in den Prüfungen. Sicher ist davon auszugehen, dass das Produzieren anspruchsvoller Texte eine andere Herausforderung darstellt als das sinnhafte Erschließen von gegebenen Texten. Im Fach Mathematik sind solche inhaltlichen Abweichungen ebenfalls vorhanden, wenngleich in geringerem Ausmaß und aus anderen Gründen: Das Lösen geometrischer Probleme unter Nutzung trigonometrischer Zusammenhänge ist bspw. in der neunten Klasse während der KERMIT-Testung noch keine zu beherrschende Fähigkeit, im Jahr darauf in der zehnten hingegen schon (siehe Tabelle 1).

	KERMIT 9	Schriftliche Überprüfung 10
Mathematik		
Leitidee Zahl	X	X
Leitidee Messen	X	X
Leitidee Raum und Form	X	X
Leitidee funktionaler Zusammenhang	X	X
Leitidee Daten und Zufall	X	X
Themen ab dem 2. Halbjahr der 9. Klasse	-	X
Deutsch		
Leseverstehen	X	X
Textproduktion	-	X
Rechtschreiben	-	X
Englisch		
Leseverstehen	X	X
Textproduktion	-	X
Rechtschreiben	-	X

Tabelle 1: Beispielhafte Gegenüberstellung der Inhaltsbereiche zwischen KERMIT 9 und den Schriftlichen Überprüfungen im Jahrgang 10

Zudem unterscheiden sich KERMIT und die zentralen Prüfungen hinsichtlich der **Vorbereitung**. Die zentralen Abschlussprüfungen und schriftlichen Überprüfungen stellen wichtige Etappen in der Bildungskarriere der Schülerinnen und Schüler dar, die entsprechend im Vorfeld intensiv durch den Unterricht vorbereitet werden. Die Regelungen für die zentralen schriftlichen Prüfungsaufgaben und die Hinweise zu den schriftlichen Überprüfungen geben Schwerpunktthemen für die Prüfungsaufgaben vor, die explizit Themen des Unterrichts sind. KERMIT dagegen bedarf keiner speziellen inhaltlichen Vorbereitung, außer dass die Schülerinnen und Schüler mit den Aufgabenformaten vertraut sein sollten. Das Wissen von Lehrstoff (z. B. Fakten und Formeln) spielt eine untergeordnete Rolle. Weil nicht sicher ist, wann in einem Zeitraum von bis zu zwei Jahren welche Inhalte genau behandelt wurden, können so möglichst gleiche Bedingungen für alle Schülerinnen und Schüler geschaffen werden. Nicht selten werden in KERMIT-Aufgaben deshalb Fakten und Zusammenhänge erläutert, bevor die Schülerinnen und Schüler diese auf neue Sachverhalte anwenden müssen.

Nach Einschätzung der Aufgabenentwicklerinnen und -entwickler beider Verfahren müssen die Schülerinnen und Schüler für die zentralen Abschlussprüfungen und schriftlichen Überprüfungen tendenziell mehr leisten als für KERMIT. Allerdings bearbeiten sie KERMIT aus dem Stegreif und werden auf die Schwerpunktthemen der zentralen Prüfungen im Unterricht intensiv vorbereitet.

Die Aufgabenentwicklung

Die Entwicklung der Aufgaben unterliegt sowohl bei KERMIT als auch bei den zentralen Prüfungen **hohen Qualitätsstandards**. Immer sind erfahrene Lehrkräfte die Urheber der Aufgaben. Nach der Entwicklung durchlaufen die Aufgaben einen Qualitätssicherungsprozess, der insbesondere die Begutachtung und Überarbeitung durch andere Fachkolleginnen und -kollegen umfasst.

Unterschiede zwischen KERMIT und den zentralen Prüfungen und Überprüfungen existieren im Aufgabenformat. Während bei KERMIT Multiple-Choice-Aufgaben, Zuordnungsaufgaben und Kurzantworten dominieren, werden bei den zentralen Prüfungen auch längere Texte oder Lösungswege gefordert. Außerdem werden die Teilaufgaben in den zentralen Prüfungen meist einheitlich inhaltlich eingebettet, während sich bei KERMIT zumindest in den Fächern Mathematik und Naturwissenschaften fast jede Aufgabe einem neuen Thema widmet.

KERMIT Aufgaben werden vor der eigentlichen Testung an repräsentativen Stichproben erprobt. Diese so genannte **Pilotierung** liefert wichtige Erkenntnisse über den Schwierigkeitsgrad der Aufgaben. Auch ermöglichen Pilotierungen eine Antwort auf die Frage, ob alle Aufgaben jeweils Facetten dergleichen Kompetenz erfassen und ob sie sich somit in die bestehenden KERMIT-

Testaufgaben einreihen lassen. Im Gegensatz profitieren die Entwicklerinnen und Entwickler der zentralen Prüfungen nicht vom Nutzen dieser Aufgabenpilotierungen, denn Prüfungsaufgaben sind streng geheim und dürfen nicht im Vorfeld in die Öffentlichkeit gelangen.

Die Durchführung

Der Zeitpunkt der Durchführung unterscheidet sich zwischen KERMIT und den zentralen Prüfungen. Beispielsweise wird KERMIT zu Beginn des zweiten Halbjahres der 9. Jahrgangsstufe durchgeführt, die schriftlichen Überprüfungen 10 und die Prüfung zum MSA etwa ein Jahr später in der zehnten Klasse. Dass die Ergebnisse zwischen KERMIT und den zentralen Prüfungen variieren, hängt demnach auch mit unterschiedlichen Lernentwicklungen und Prüfungsvorbereitungen in diesem zusätzlichen Lernjahr zusammen.

Ein wichtiger Unterschied in der Durchführung beider Verfahren besteht in der **Motivation** der Schülerinnen und Schüler: Weil KERMIT für die getesteten Schülerinnen und Schüler ohne persönliche Konsequenzen bleibt, liegt die Anstrengungsbereitschaft im Mittel vermutlich unterhalb der Motivation in Prüfungssituationen. Da dies für alle Schülerinnen und Schüler im Mittel gleichermaßen gilt, wird die Aussagekraft von KERMIT für die Schul- und Unterrichtsentwicklung nicht geschmälert. Es kann aber Unterschiede im Abschneiden zwischen KERMIT und den zentralen Prüfungen bzw. Überprüfungen erklären, nämlich dann, wenn die Schülerinnen und Schüler einer Schule bei KERMIT überproportional stark oder schwach motiviert sind, und in den zentralen Prüfungen wie alle Schülerinnen und Schüler ihre maximale Leistungsfähigkeit abrufen. Auf der anderen Seite sind auch Phänomene wie Prüfungsangst oder Prüfungsversagen bei den zentralen Prüfungen nicht auszuschließen, die bei KERMIT eine untergeordnete Rolle spielen dürften.

Ebenso von Relevanz ist, dass die zentralen Abschlussprüfungen und schriftlichen Überprüfungen wesentlich weniger **Zeit** zur Aufgabenlösung einräumen, als es bei KERMIT der Fall ist. In den Prüfungen berichten die Schülerinnen und Schüler nicht selten von Zeitdruck. Bei KERMIT werden dagegen fehlende Antworten *am Ende* des Tests als „nicht vorgelegt“ betrachtet, um den Aspekt der Bearbeitungsgeschwindigkeit herauszurechnen. Anders als bei KERMIT muss man in den zentralen Prüfungen also nicht nur gut, sondern auch ziemlich schnell sein.

Die Auswertung

Die Auswertung verläuft bei KERMIT zentral durch geschulte Personen, die die Aufgaben ausschließlich in richtig bzw. falsch unterscheiden (das sogenannte Codieren). Da die Codiererinnen und Codierer in den meisten Testbereichen fachfremd sind, gibt es dazu eindeutige Codieranweisungen. Bei ca. 10 Prozent aller Testhefte wird eine unabhängige Zweitcodierung

durchgeführt. Ab einer Abweichung von mehr als 3 Prozent wird die Codiererin / der Codierer nachgeschult und korrigiert anschließend ihre / seine bisher eingegebenen Testhefte. Die Abweichungen bewegen sich dadurch im Bereich zwischen 0,5 und 1,1 Prozent. Die so erzeugten Daten werden auf der Grundlage des sogenannten Raschmodells skaliert. Dadurch wird sichergestellt, dass alle Aufgaben Facetten der gleichen Kompetenz erfassen, und dass es keine „Tricks“ gibt, die die Aufgabenlösung ohne fachliche Kompetenzen erlauben (z. B. geschicktes Ausschließen von Antwortoptionen).

Bei den zentralen Prüfungen erfolgt die Bewertung durch die Fachlehrkräfte. Dadurch können auch Punkte für richtige Teillösungen bzw. gute Lösungsansätze vergeben werden. Benotet wird anhand einer detaillierten Anweisung, die auch die Notenfestlegung einheitlich beschreibt. Eine Kontrolle der Ergebnisse erfolgt über die kollegiale Begutachtung der Erstbewertungen durch weitere Fachlehrkräfte.

Die Ergebnisse

Trotz aller bisher erläuterten Unterschiede weisen die Ergebnisse aus KERMIT und den zentralen Prüfungen einen substantiellen Zusammenhang auf. Dieser wird in einer so genannten Korrelation ausgedrückt und beträgt z. B. zwischen KERMIT 9 und der schriftlichen Überprüfung im Fach Mathematik 0,6. Läge dieser Wert bei 0, bestünde kein Zusammenhang, läge er bei 1, bestünde ein perfekter Zusammenhang. Im letzten Fall könnte man auf eines der beiden Verfahren verzichten, denn die Noten der zentralen Prüfungen ließen sich ohne Unsicherheit aus den KERMIT Ergebnissen berechnen (und umgekehrt). Umgangssprachlich bedeutet eine Korrelation von 0,6, dass Schulen, die in der neunten Klasse bei KERMIT im Verhältnis zu anderen Schulen gut abschneiden, im Großen und Ganzen in der schriftlichen Überprüfung der 10. Klasse ebenfalls erfolgreich sind. Grafisch lässt sich das Ergebnis des Vergleichs an einer Punktwolke erkennen, die tendenziell von links oben nach rechts unten verläuft (siehe Abbildung 1).

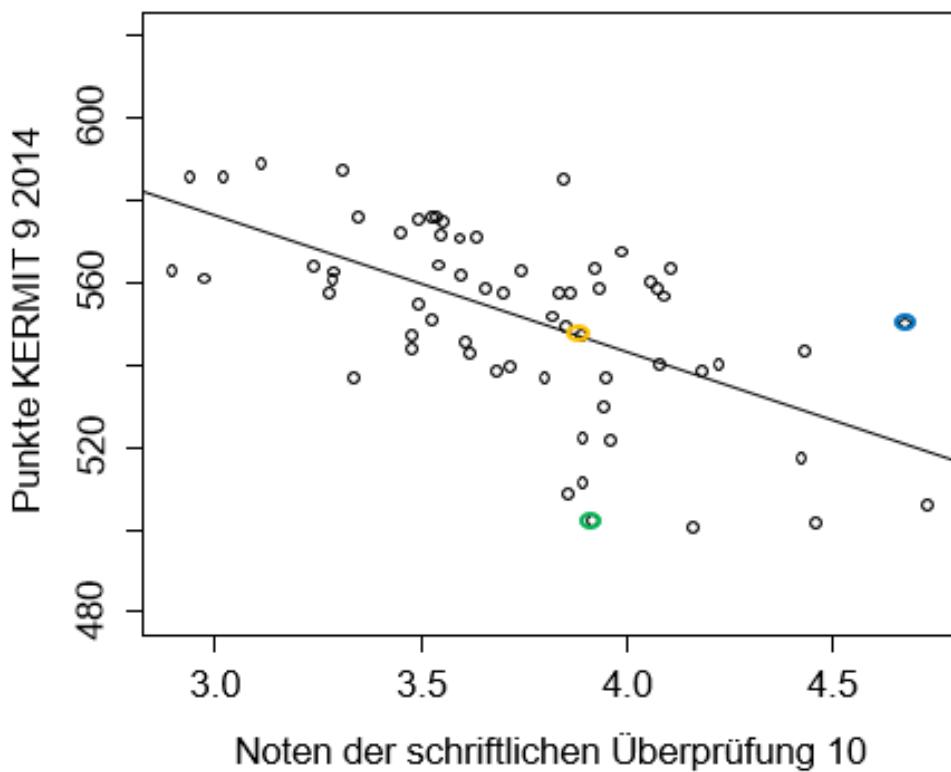


Abbildung 1: Der Zusammenhang zwischen KERMIT 9 und der schriftlichen Überprüfung in Klasse 10 am Beispiel Mathematik. Dargestellt sind die Mittelwerte von 63 Gymnasien¹.

Es ist jedoch nicht auszuschließen, dass die unterschiedlichen Skalen beider Verfahren unterschiedlich wahrgenommen werden, selbst wenn sie das Gleiche aussagen. Beispielsweise liegt der Erwartungswert einer Schule, die im Schnitt 552 Punkte bei KERMIT erzielt hat, bei einer Note von 3,7 in der schriftlichen Überprüfung (etwa der orangefarbene Kringel, siehe Abbildung). Während 552 KERMIT Punkte eventuell „erfolgreich“ klingen, da sie mehr als 50 Punkte über dem Durchschnitt der Hamburger Schülerschaft liegen, suggeriert eine knappe Vier in der schriftlichen Überprüfung ein Problem. Tatsächlich liegen die Ergebnisse der Schule aber jeweils knapp über dem Durchschnitt der übrigen Gymnasien und sagen insofern dasselbe aus.

Darüber hinaus gibt es in Einzelfällen tatsächliche Abweichungen von den Erwartungswerten. Eine Schule, die bei KERMIT mit etwa 550 Punkten im Durchschnitt der Gymnasien liegt, hat mit einem Notendurchschnitt von mehr als 4,5 in den schriftlichen Überprüfungen ein deutlich unterdurchschnittliches Ergebnis erzielt (blaue Markierung). Dies gilt mit umgekehrten Vorzeichen z. B. für die grün markierte Schule in der Abbildung.

¹ Die negative Steigung der Geraden resultiert aus der negativen Polung von Schulnoten. Anders als bei KERMIT-Punkten gilt bei Noten „je kleiner desto besser“.

Die Ursachen für derartige Abweichungen können in allen oben beschriebenen Unterschieden zwischen KERMIT und den zentralen Prüfungen zu finden sein. Meist wird es eine Kombination aus verschiedenen Faktoren sein, die in unterschiedlichen Schulen unterschiedlich stark auf die Ergebnisse der beiden Verfahren wirken und so das Gefühl einer Diskrepanz hervorrufen.

Fazit: Eng verwandt und doch verschieden

KERMIT und die zentralen Abschlussprüfungen bzw. schriftlichen Überprüfungen sind inhaltlich eng verwandt. Beide Verfahren beziehen sich auf die Hamburger Bildungspläne und damit auf die Bildungsstandards der KMK. Sie verfolgen jedoch gänzlich unterschiedliche Ziele. KERMIT liefert vornehmlich Informationen für Lehrkräfte zur Verbesserung von Schule und Unterricht. Die zentralen Prüfungen weisen Bildungszertifikate zu, die für die weitere schulische, berufliche und soziale Laufbahn der Schülerinnen und Schüler von höchster Relevanz sind. In der Folge ergeben sich zahlreiche Unterschiede in der Konzeption, der Durchführung, der Auswertung und – wie vor diesem Hintergrund zu erwarten – im Ergebnis. Dass KERMIT und die zentralen Prüfungen im Kern dennoch die gleichen Kompetenzen erfassen, zeigen die statistischen Analysen. Insofern erfüllen beide Erhebungen ihren jeweiligen Zweck.

Gleichwohl wird derzeit von den Verantwortlichen beider Verfahren geprüft, inwieweit man diese noch besser aufeinander beziehen kann. Alle Verantwortlichen arbeiten daran, dass KERMIT seine Signal- und Steuerwirkung für die Lehrkräfte und Schulen weiter verbessert. So können mehr Schülerinnen und Schüler durch passgenauen Unterricht diejenigen Ziele erreichen, die später in den zentralen Abschlussprüfungen und schriftlichen Überprüfungen verlangt werden. In einem ersten Schritt wurden regelmäßige gemeinsame Treffen zwischen sämtlichen Beteiligten vereinbart, in denen an der Verbesserung der jeweiligen Aussagekraft von KERMIT und den zentralen Abschlussprüfungen und schriftlichen Überprüfungen gearbeitet wird.

Dr. Frank Musekamp (BQ 22-2)
frank.musekamp@ifbq.hamburg.de

Dr. Claudia Hildenbrand (BQ 21-6)
claudia.hildenbrand@ifbq.hamburg.de

Gitta Taube (BQ-F 1)
gitta.taube@ifbq.hamburg.de

Stefan Klitsche (BQ 22-3)
stefan.klitsche@ifbq.hamburg.de

Freie und Hansestadt Hamburg
Institut für Bildungsmonitoring und Qualitätsentwicklung – IfBQ
Beltgens Garten 25, 20537 Hamburg
Telefon: +49(0)40-42 88 51-495
www.hamburg.de/ifbq