

Inhaltsverzeichnis

1.	Einleitung.....	2
2.	KERMIT: ein System zur Unterstützung der Qualitätsentwicklung der Schulen und zur Generierung von Steuerungswissen	4
3.	Sicherstellung einer hohen Aufgabenqualität bei KERMIT	9
3.1	Testentwicklung in der Aufgabenentwicklergruppe Deutsch.....	11
3.2	Testentwicklung in der Aufgabenentwicklergruppe Mathematik.....	14
3.3	Testentwicklung in der Aufgabenentwicklergruppe Englisch.....	17
	Kompetenzbereich Hörverstehen	17
	Textmapping als Methode zur Erstellung von Testaufgaben	18
3.4	Testentwicklung in der Aufgabenentwicklergruppe Naturwissenschaften.....	20
4.	Die Darstellung der Ergebnisse in den KERMIT-Rückmeldungen	23
5.	Von der Ergebnisreflektion hin zu einer systematischen Schul- und Unterrichtsentwicklung	30
6.	Diskussion und Ausblick.....	34
7.	Literatur.....	38

KERMIT – Kompetenzen ermitteln

Markus Lücke, Franziska Thonke, Britta Pohlmann, Helga Hoffmann, Reinhard Golecki, Johannes Rosendahl, Margarete Benzing und Jan Poerschke

1. Einleitung

Sowohl bundesweit als auch innerhalb der Bundesländer gibt es vielfältige Projekte und Ansätze zur datengestützten Schul- und Unterrichtsentwicklung (z. B. MARKUS, Helmke & Jäger, 2002; QUASUM, Peek, 2004 oder QUISS, Brackhahn & Brockmeyer, 2004). Der Auftrag der Schulen, moderne Schul- und Unterrichtsentwicklung umzusetzen, machte einen Paradigmenwechsel notwendig, der sich von einer Input- oder Ressourcenorientierung hin zu einer Outcome-Orientierung bewegt, in der die zu erreichen Ziele fokussiert werden. Nicht die Ausbildung der Lehrkräfte, die Bereitstellung von Ressourcen und ein guter Lehrplan stehen im Vordergrund, sondern das Ziel, „kompetente“ Schülerinnen und Schüler auszubilden, ausgehend von verbindlichen Standards. Dies bedeutet eine größere Schulautonomie, aber gleichzeitig auch eine höhere Eigenverantwortlichkeit der Schule. Dieser Perspektivwechsel stellt hohe Anforderungen an alle Beteiligten des Bildungssystems. Spätestens seit PISA 2000 und der daraus entstandenen Diskussion zur Bildung in Deutschland, wurde eine empirische Wende eingeläutet, in der sich die Schulpolitik und die pädagogische Praxis stärker an sozialwissenschaftlich gesicherten Befunden orientiert (Lange, 2008). In der Kultusministerkonferenz (KMK) einigten sich die Länder auf eine Gesamtstrategie zum Bildungsmonitoring, die vier Eckpunkte enthält (KMK, 2006): Neben der Teilnahme an internationalen Schulleistungsuntersuchungen werden auch innerhalb Deutschlands Verfahren zur Beobachtung und Weiterentwicklung der Bildungsprozesse entwickelt und regelmäßig durchgeführt. Auf der Grundlage dieser Gesamtstrategie wird vom Institut zur Qualitätsentwicklung im

Bildungswesen (IQB) in regelmäßigen Abständen ein Ländervergleich in den Kernfächern durchgeführt. Eine gemeinsame Bildungsberichterstattung von Bund und Ländern ist ebenfalls ein Teil der Gesamtstrategie. Als vierter Eckpunkt der Gesamtstrategie wurden in (fast) allen Bundesländern¹ verbindliche Vergleichsarbeiten (VERA) in den Jahrgangsstufen 3 und 8 zur Unterstützung der datengestützten Schul- und Unterrichtsentwicklung eingeführt. Diese Vergleichsarbeiten, die in den meisten Bundesländern unter dem Titel „VERA“, in Hamburg bisher unter dem Titel Lernstandserhebungen und nun unter dem Namen KERMIT 3 und 8 laufen, werden seit 2007/08 durchgeführt.

Dennoch gibt es sowohl von wissenschaftlich-methodischer Seite (z. B. Kühle & van Ackeren, 2012; Berkemeyer & van Holt, 2012) als auch von bildungspolitischer Seite Kritik (z. B. Hundt, 2009), die sich u. a. darauf bezieht, dass das Potenzial von Schulleistungstests im Rahmen von VERA noch nicht ausgeschöpft wird. So wird kritisiert, dass die Vergleichsarbeiten nur den Lernstand der Klassen wiedergeben und nicht zu belastbaren Längsschnittdaten führen, die zur individuellen Diagnose der Leistungsentwicklung der Klassen und Schülerinnen und Schüler herangezogen werden können (Berkemeyer & van Holt, 2012). Dazu müssen zusätzlich soziale und kategoriale Bezugsnormen zur Verfügung gestellt werden, anhand derer die Schulleitungen und Lehrkräfte die Ergebnisse ihrer Klassen interpretieren können. Häufig sind die Ergebnisdarstellungen für die Adressaten auch nicht nachvollziehbar, so fehlen Ihnen konkrete Handlungsempfehlungen, wie sie durch den Umgang mit den Ergebnissen Anregungen für die weitere Schul- und Unterrichtsentwicklung erhalten (Koch, 2011). Diese Kritik versucht das Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ) Hamburg, das für die Durchführung und Auswertung von KERMIT 3 und 8 verantwortlich ist, aufzugreifen. Durch eine wiederholte Ermittlung der Kompetenzen im Verlauf des Bildungswegs wird die Dokumentation der Lernentwicklung der Schülerinnen und Schüler durch „echte Längsschnitte“ ermöglicht, und kann zusätzlich mit sozialen

¹ Baden-Württemberg beteiligt sich nur an VERA 3 und nicht an VERA 8.

und kriterialen Bezugsnormen verglichen werden, die sich auf die Hamburger Schulen und den von der Behörde für Schule und Berufsbildung (BSB) verabschiedeten Bildungsplänen beziehen.

2. KERMIT: ein System zur Unterstützung der Qualitätsentwicklung der Schulen und zur Generierung von Steuerungswissen

Hamburg hat mit LAU (Aspekte der Lernausgangslage und der Lernentwicklung, Lehmann, Peek & Gänßfuß, 2002; BSB, 2011) und KESS (Kompetenzen und Einstellungen von Schülerinnen und Schülern; Bos, Gröhlich & Pietsch, 2007; Bos & Gröhlich; 2010; Vieluf, Ivanov & Nikolova, 2012) bereits eine lange Tradition in der systemischen Durchführung und Auswertung von Schulleistungsstudien. Beide Studien lieferten und liefern wichtiges Steuerungswissen für das Hamburger Bildungssystem und darüber hinaus auch für die Lehr-Lern-Forschung.

Auf Grundlage dieser Erhebungen wurden in Hamburg die Lernausgangslagen-Tests in den Jahrgangsstufen 5 und 7 (LeA 5 und 7) als ein Instrument zur Dokumentation der Leistungsentwicklung in den Schulen entwickelt. Zu Beginn der weiterführenden Schule (Jahrgangsstufe 5) konnten die Schulen LeA 5 einsetzen, um Informationen über den Leistungsstand und die Leistungsheterogenität in ihren Klassen zu bekommen. Bei LeA 7 wurde dann darüber hinaus die Leistungsentwicklung (ausgehend von der Jahrgangsstufe 5) dokumentiert. Dies erlaubt den Schulen eine Überprüfung, inwieweit ihre Schülerinnen und Schüler in den ersten beiden Jahren der weiterführenden Schule vom Unterricht profitieren konnten. Dieses Instrument war ausschließlich für den schulinternen Gebrauch zur datengestützten Schul- und Unterrichtsentwicklung gedacht und wurde nicht zum Systemmonitoring genutzt. Die Ergebnisse wurden nur den Schulen zur Verfügung gestellt und diese hatten die Interpretationshoheit über die Daten. Der Zuspruch in

den Hamburger Schulen zu diesem Instrument war groß, so dass fast 90 Prozent aller weiterführenden Schulen sich bis 2011 freiwillig an LeA 5 beteiligt haben, und nahezu alle dieser Schulen auch an LeA 7.

Aus den Erfahrungen mit diesen Längsschnittstudien und den Vergleichsarbeiten sowie geleitet von der Akzeptanz der Hamburger Schulen und Lehrkräfte hat die Behörde für Schule und Berufsbildung (BSB) als eine der Maßnahmen zur Qualitätsentwicklung an Hamburger Schulen im Rahmen ihrer Qualitätsoffensive 2012 das umfassende Kompetenzermittlungssystem KERMIT (Kompetenzen ermitteln) an den Schulen eingeführt. Seit Beginn des Schuljahrs 2012/13 werden nun an den Hamburger Schulen ausgewählte Kompetenzen der Schülerinnen und Schüler in zentralen Fachbereichen regelhaft mit Hilfe standardisierter Tests ermittelt. KERMIT wird jährlich in sechs Jahrgangsstufen (2, 3, 5, 7, 8 und 9) durchgeführt und integriert die bundesweiten Vergleichsarbeiten (VERA bzw. Lernstand) sowie die in Hamburg seit Jahren etablierten Erhebungen zur Lernausgangslage (s. Abb. 1).

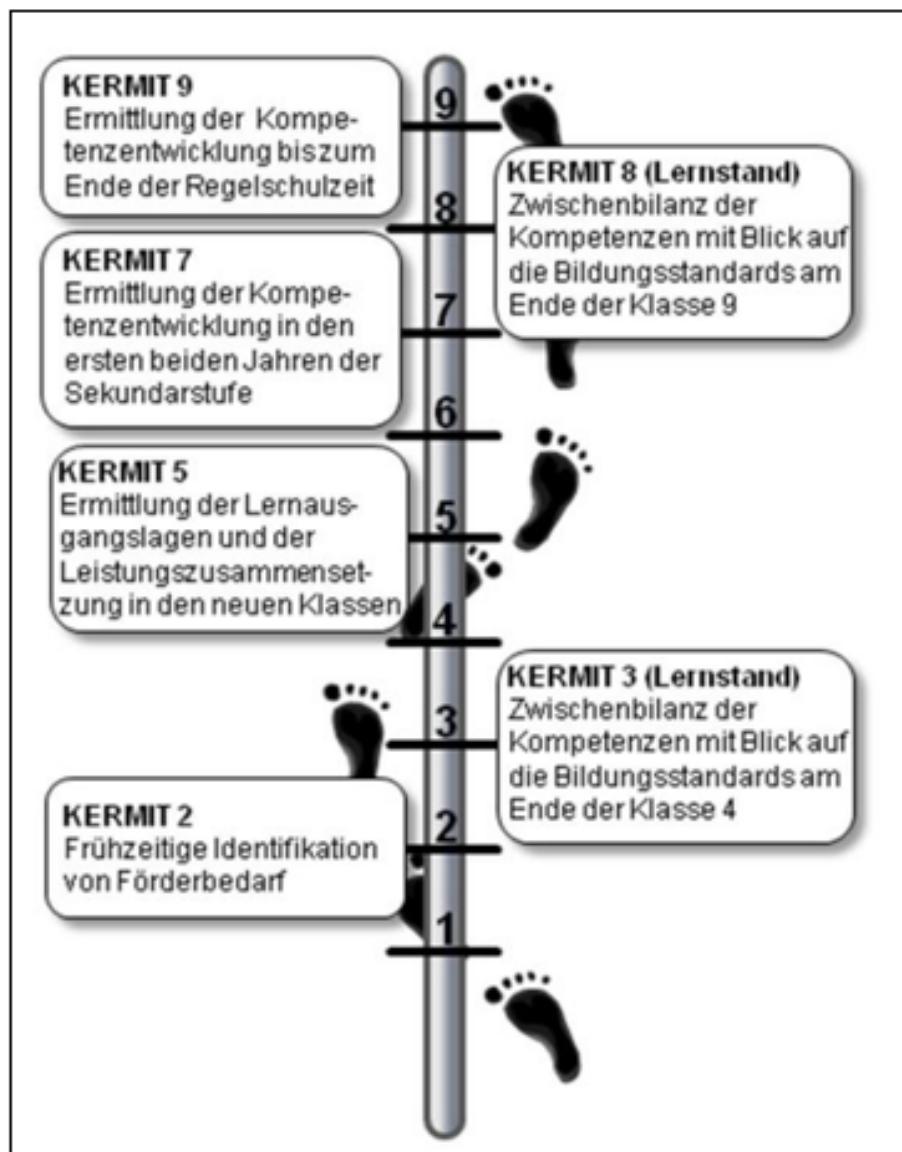


Abbildung 1: Chronologie der KERMIT-Erhebungen mit ihren wichtigsten Zielen

Mit den KERMIT-Erhebungen werden zentrale Leistungsindikatoren erfasst, die ebenfalls in den nationalen und internationalen Schulleistungsuntersuchungen (z. B.

PISA, IQB-Ländervergleich) zur Untersuchung der Leistungsfähigkeit von Bildungssystemen herangezogen werden. Zum Einsatz kommen standardisierte, normierte und aufeinander abgestimmte Schulleistungstests. Auf diese Weise werden objektive diagnostische Informationen gewonnen, mit denen die aktuellen schulischen Leistungen der Hamburger Schülerinnen und Schüler sowie die Entwicklung ihrer Leistungen im Verlauf der Schulzeit zuverlässig abgebildet werden. Lehrkräfte und Schulleitungen können die Ergebnisse für eine evidenzbasierte Unterrichts- und Schulentwicklung nutzen, indem sie die Schülerleistungen vor dem Hintergrund von individuellen, sozialen und kriterialen Bezugsnormen reflektieren. Da die KERMIT-Erhebungen in den Jahrgangsstufen 5, 7 und 9 durch externe Testleiterinnen und Testleiter administriert werden, lassen sich gerade aus diesen Ergebnissen besonders aussagekräftige Erkenntnisse über das Hamburger Bildungssystem und die Leistungsentwicklung an den Schulen gewinnen. Dadurch ist es zum einen möglich, ein laufendes Systemmonitoring zentraler kognitiver Kompetenzen der Hamburger Schülerinnen und Schüler zu realisieren. Zum anderen können Effekte von Steuerungsmaßnahmen und Schulversuchen auf relevante Leistungskriterien zeitnah untersucht werden. Damit eröffnet sich eine Vielzahl an Auswertungsperspektiven, insbesondere wenn ab dem Schuljahr 2014/2015 die Lernentwicklungen der Schülerinnen und Schüler über mehrere Jahre nachvollziehbar werden und evtl. eine Verknüpfung der KERMIT-Ergebnisse mit weiteren Daten der Bildungs- und Schulstatistik möglich wird.

Alle KERMIT-Erhebungen orientieren sich an den bundesweit verbindlichen Bildungsstandards und dienen der Schul- und Unterrichtsentwicklung. Die Bildungsstandards sind für zentrale Schnittstellen im Bildungsweg formuliert: Sie liegen für das Ende der Grundschulzeit (Ende der 4. Klassen) sowie für das Ende der Sekundarstufe I vor.² Wie bisher bei den Lernstandserhebungen wird auch zukünftig bei KERMIT 3 und 8 geprüft, welche Kompetenzen die Schülerinnen und Schüler

² Seit 2012 liegen auch Bildungsstandards für das Abitur vor, diese spielen im Rahmen von KERMIT aber keine Rolle.

im Hinblick auf die nationalen Bildungsstandards in Deutsch, Mathematik und in der ersten Fremdsprache (nur in Jahrgangsstufe 8) erreicht haben. Diese Erhebungen werden auch wie bisher von den Lehrkräften durchgeführt und die Schulen erhalten dazu ausführliches Begleitmaterial für die weitere Unterrichtsentwicklung (siehe Tab. 1). Bei den KERMIT-Erhebungen, die in Hamburg für die Jahrgänge 2, 5, 7, und 9 exklusiv entwickelt werden, werden die Aufgaben verstärkt mit den Hamburger Bildungsplänen für die Fächer Deutsch, Mathematik, Englisch und Naturwissenschaften abgeglichen, um die curriculare Validität dieser Tests und damit die Aussagekraft über die Leistungsentwicklung Hamburger Schülerinnen und Schüler zu erhöhen. Diese Erhebungen werden darüber hinaus auch durch externe, geschulte Testleiterinnen und Testleiter vorgenommen, um die Durchführungsobjektivität zu erhöhen.

Tabelle 1: Testkomponenten und Durchführungsbedingungen der KERMIT-Erhebungen

Zeitpunkt	Testinhalte	Durchführung	Referenzrahmen
Jg. 2 (2. Hj.)	• Deutsch		Hamburger Bildungspläne
Jg. 3 (2. Hj.) =Lernstand/VERA	• Mathematik	Lehrkräfte	Nationale Bildungsstandards
Jg. 5 (Anfang)	• Deutsch • Mathematik	Externe Testleiter/innen	Hamburger Bildungspläne
Jg. 7 (Anfang)	• Englisch • Naturwissenschaften		
Jg. 8 (2. Hj.) =Lernstand/VERA	• Deutsch • Mathematik • 1. Fremdsprache	Lehrkräfte	Nationale Bildungsstandards
Jg. 9 (2. Hj.)	• Deutsch • Mathematik • Englisch • Naturwissenschaften	Externe Testleiter/innen	Hamburger Bildungspläne

Somit stellt KERMIT einen Ansatz dar, die unterstützende Funktion von VERA/Lernstand bei der datengestützten Schul- und Unterrichtsentwicklung durch die Ermittlung des Lernstands in den Jahrgangsstufen 3 und 8 mit den Möglichkeiten der Erfassung der Leistungsentwicklung in ein einheitliches System zu integrieren und darüber hinaus durch systemische Analysen auch Erkenntnisse für die Bildungsadministration zu generieren. Damit all diese Funktionen von KERMIT auch erfüllt werden können, ist eine Qualitätssicherung notwendig. Neben einer sorgfältigen Durchführung und Auswertung der Erhebungen nach wissenschaftlichen Standards sowie der didaktischen Aufbereitung der Ergebnisrückmeldung und der Unterstützung der Schulen bei der datengestützten Schul- und Unterrichtsentwicklung nehmen vor allem in der Phase der Aufgabenentwicklung wissenschaftliche Grundsätze eine zentrale Stellung ein.

3. Sicherstellung einer hohen Aufgabenqualität bei KERMIT

Im Rahmen der KERMIT-Erhebungen werden Aufgaben in einem aufwändigen Verfahren entwickelt, um die Qualität des Instruments zu sichern. Dieses Verfahren entspricht dem Vorgehen, wie es auch im Rahmen anderer großer Schulleistungsstudien Standard ist (Hornberg, Bos, Buddeberg et al., et al., 2006; Frey, Carstensen, Walter et al., 2008; Köller, Knigge & Tesch, 2010; Rupp, Vock, Harsch & Köller, 2008).

Aufgaben, die im Rahmen eines Testverfahrens eingesetzt werden („Testaufgaben“), müssen anderen Ansprüchen genügen als Aufgaben, die den Schülerinnen und Schülern im Unterricht gestellt werden („Lernaufgaben“). Neben den klassischen Bedingungen Objektivität, Reliabilität und Validität gehört dazu auch z. B. auch die die Beachtung der Aufgabenfairness, so dass es nicht zu einer Bevorzugung eines

Geschlechts oder von Personen mit bestimmten kulturellen Hintergründen bei einzelnen Aufgaben kommt.

Die Aufgaben für KERMIT 2, 5, 7 und 9 werden in Hamburg am IfBQ in verschiedenen Aufgabenentwicklergruppen erarbeitet. Im Rahmen der Aufgabenentwicklung arbeiten zwei Gruppen im Bereich Grundschule (Deutsch und Mathematik) sowie vier Gruppen im Bereich weiterführende Schule (Deutsch, Mathematik, Englisch und Naturwissenschaften). Grundlage für die Aufgabenentwicklung sind dabei die nationalen Bildungsstandards, die Hamburger Bildungspläne sowie theoretische Grundlagen aus der jeweiligen Fachdidaktik. Dadurch wird gewährleistet, dass sich die Aufgaben auch auf das beziehen, was sie messen sollen (Sicherung der Validität). Die Aufgabenentwicklergruppen setzen sich aus Hamburger Lehrkräften und Psychologen zusammen und werden fachdidaktisch von Wissenschaftlerinnen und Wissenschaftlern verschiedener Universitäten begleitet. Im Vorfeld werden die betreffenden Lehrkräfte hinsichtlich der theoretischen Grundlagen der Aufgabenentwicklung geschult. Bevor die neu entwickelten Aufgaben im Testverfahren eingesetzt werden, werden sie in einem aufwändigen Verfahren geprüft. In einem ersten Schritt werden die Aufgaben im Rahmen einer Präpilotierung in kleineren Stichproben eingesetzt, um zu kontrollieren, ob die Aufgabenstellungen für die Schülerinnen und Schüler verständlich sind, die Aufgaben eindeutig gelöst werden können und keine Abhängigkeiten zwischen den Aufgaben vorhanden sind. Die universitäre Begleitung kommentiert parallel dazu die Aufgaben hinsichtlich ihrer fachdidaktischen Eignung. Gegebenenfalls werden die Aufgaben aufgrund der Ergebnisse aus der Präpilotierung optimiert.

Im Anschluss an die Präpilotierung werden die Aufgaben an einer großen Schülerinnen- und Schülerpopulation normiert. Sinnvollerweise werden in diesem Schritt ebenfalls bereits etablierte Aufgaben aus dem gleichen Testbereich eingesetzt, um die neuen Aufgaben an das bewährte Testverfahren zu justieren. Für die statistische Auswertung wird mit Hilfe des eindimensionalen Rasch-Modells

dann überprüft, ob sich alle Testaufgaben eines Kompetenzbereichs hinsichtlich ihrer Itemparameter auf eine Skala abbilden lassen. Für jede Aufgabe liegen somit wichtige Gütekriterien, wie die Lösungshäufigkeit und die Trennschärfe vor, die eine Entscheidung erlauben, ob diese Aufgabe im Rahmen eines Testverfahrens eingesetzt werden kann. Eindeutige Korrekturanweisungen und externe Testleitungen sichern darüber hinaus die Auswertungs- und Durchführungsobjektivität des Verfahrens.

Nachfolgend wird die Arbeit der jeweiligen Aufgabenentwicklergruppen vorgestellt.

3.1 Testentwicklung in der Aufgabenentwicklergruppe Deutsch

Neben den Kompetenzbereichen Sprache und Sprachverständnis sowie Zuhören werden in der Aufgabenentwicklergruppe Deutsch vor allem Aufgaben zum Leseverstehen entwickelt. Der Fähigkeit sinnentnehmend Texte lesen zu können, kommt eine Schlüsselposition zu. Sie ist insofern bedeutsam, als dass sie Grundlage für die Teilnahme am gesellschaftlichen Leben darstellt (Bos, Lankes, Prenzel & Schwippert, 2003; Bremerich-Vos & Böhme, 2009). Aus diesem Grund werden im Zuge von KERMIT regelhaft Lesekompetenzen erfasst, die dann im Längsschnitt abgebildet werden. So kann für jede Schülerinnen und jeden Schüler ihre bzw. seine Entwicklung im Bereich Lesen aufgezeigt werden.

Grundlage der Aufgabenentwicklung im Bereich Deutsch Leseverstehen bildet ein Stufenmodell, das auf dem Kompetenzstufenmodell des IQB aufbaut und die verschiedenen Hierarchie-Ebenen beim Lesen beschreibt (Tab. 2; Zabka, 2008):

Tabelle 2: Stufenmodell zur Aufgabenbeurteilung von Aufgaben im Bereich
Deutsch-Leseverstehen (Zabka, 2008)

<i>Aspekt</i>	<i>Fragen zu textseitigen Anforderungen an ein gutes Verstehen (Beispiele):</i>
1 manifeste Informationen verstehen	<ul style="list-style-type: none"> • Welche manifesten Informationen könnten inhaltlich falsch oder ungenau verstanden werden? (Verwechslung mit anderen Wörtern und nicht gemeinten Wortbedeutungen) • Welche manifesten Informationen und Informationsverknüpfungen könnten aus syntaktischen Gründen falsch oder ungenau verstanden werden? (mögliche Verkennung der Funktion von Satzgliedern, Konnektoren, Pro-Formen) • Welche manifesten Informationsverknüpfungen könnten überlagert werden durch konkurrierende Informationen und (text- oder aufgabenseitig provozierte) Schlussfolgerungen?
2 Informationen lokal erschließen	<ul style="list-style-type: none"> • An welchen Stellen ist das Erschließen von Informationen oder Informationszusammenhängen notwendig, aber schwierig? (Unauffälligkeit der Informationslücke, Uneindeutigkeit der zu erschließenden Information) • An welchen Stellen muss man erschließen, wie eine Aussage semantisch oder pragmatisch gemeint ist? (mehrere mögliche Bedeutungen, Sprechakt) • An welchen Stellen muss man erschließen, dass es eine Differenz zwischen der Aussage und dem Gemeinten gibt? (rhetorische Indirektheit: z.B. Ironie, Übertreibung, Vergleich, Metapher) • An welchen Stellen muss man erschließen, dass Wahrnehmungen, Äußerungen, Gedanken usw. keine unmittelbaren Text- oder Verfasser-Aussagen sind, sondern zu einer im Text relativierten Position gehören? (perspektivische Indirektheit: z.B. Zitate, Figurenrede, Figurengedanken)
3 globale Zusammenhänge verstehen	<ul style="list-style-type: none"> • Welche Informations-, Instruktions-, Argumentations- oder Handlungszusammenhänge müssen die Leser verstehen, um zu einem kohärenten Mindest-Verstehen des Textinhalts zu gelangen? (Textweltmodell) • Welche dieser Zusammenhänge könnten schwer zu verstehen sein, z. B.

	<ul style="list-style-type: none"> ○ weil die Elemente im Text zwar explizit verbunden, aber weit verstreut sind; ○ weil es andere Informationen gibt, die wichtige Zusammenhänge überlagern; ○ weil der Zusammenhang nur durch eine Schlussfolgerung hergestellt werden kann?
4 konzeptuell bündeln	<p>Welche übergeordneten Vorstellungen oder Begriffe zur Zusammenfassung und Interpretation sollen gebildet werden? Worauf sollen sie sich beziehen? Auf...</p> <ul style="list-style-type: none"> ● Orte, Dinge, Umstände usw. der dargestellten äußeren Wirklichkeit; ● Taten, Gefühle, Einstellungen, Ziele, Eigenschaften von Personen; ● Absichten, Thesen, Meinungen, Sprechakte des Verfassers; ● Themen, Hauptaussagen, Funktionen von Textabschnitten; ● Thema, Hauptaussage, Zweck, Sinn des gesamten Textes.
5 Textgestaltung interpretieren	<ul style="list-style-type: none"> ● Welche rhetorischen und stilistischen Texteigenschaften sollen wahrgenommen und zum Ausweis eines sprachbewussten Textverständens interpretiert werden? ● Welche Superstrukturen, insbesondere Eigenschaften und Funktionen von Textsorten, sollen erkannt und für die Interpretation genutzt werden? ● In Bezug auf welche Texteigenschaften ist dies Bedingung oder Bereicherung des Verstehens?

Anhand dieser Hierarchie-Ebenen kann ein Text von den Aufgabenentwicklerinnen und -entwicklern bearbeitet werden. Zunächst wird geprüft, ob sich zu dem vorliegenden Text Aufgaben zu allen Aspekten formulieren lassen. Darüber hinaus muss der Text weiteren Kriterien genügen. Dazu gehört seine Textlänge (hinreichend, aber nicht zu lang), seine Motivierungsqualität und das Thema des Textes. Dieses soll nicht nur bestimmte Schülergruppen ansprechen, sondern für viele Schülerinnen und Schüler gleichermaßen interessant sein. Nach der Textauswahl werden anhand des obigen Schemas Aufgaben entwickelt. Diese

werden dann wie in Abschnitt 3 dargestellt fachdidaktisch und psychometrisch geprüft und ggf. weiterentwickelt bzw. überarbeitet.

3.2 Testentwicklung in der Aufgabenentwicklergruppe Mathematik

Die wesentlichen Grundlagen für die fachspezifische Testentwicklung im Bereich Mathematik sind auch im Fach Mathematik die nationalen Bildungsstandards, die Hamburger Bildungspläne sowie die vom IQB entwickelten Testaufgaben und die aus ihrem Einsatz empirisch gewonnenen Kompetenzstufen. Als relevant für alle Ebenen des mathematischen Arbeitens, für eine erfolgreiche Nutzung und Aneignung von Mathematik, werden in den Bildungsstandards Mathematik zunächst allgemeine mathematische Kompetenzen genannt und an Beispielen konkretisiert. Einen kurzen Ausschnitt zeigt Tabelle 3.

Tabelle 3: Liste der allgemeinen mathematischen Kompetenzen mit Beispielen aus den nationalen Bildungsstandards Mathematik

Allgemeine mathematische Kompetenzen³	Beispiele
(K 1) Mathematisch argumentieren	Lösungswege beschreiben und begründen
(K 2) Probleme mathematisch lösen	geeignete Hilfsmittel und Strategien zum Problemlösen auswählen
(K 3) Mathematisch modellieren	Situationen in mathematische Begriffe und Strukturen übersetzen
(K 4) Mathematische Darstellungen verwenden	Beziehungen zwischen Darstellungsformen erkennen
(K 5) Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen	mit Gleichungen, Tabellen und Diagrammen arbeiten
(K 6) Kommunizieren	Äußerungen von anderen und Texte zu mathematischen Inhalten verstehen

Die inhaltsbezogenen mathematischen Kompetenzen der Bildungsstandards werden jeweils ausgewählten Leitideen zugeordnet, die sich an grundlegenden mathematischen Konzepten orientieren. Einige Beispiele sind in der Tabelle 4 abgebildet.

³ Im Primarbereich Problemlösen, Kommunizieren, Argumentieren, Modellieren, Darstellen.

Tabelle 4: Inhaltsbezogene mathematische Kompetenzen (Leitideen) mit Beispielen

Leitideen⁴	Beispiele: Die Schülerinnen und Schüler
(L 1) Leitidee Zahl	verwenden Prozent- und Zinsrechnung sachgerecht.
(L 2) Leitidee Messen	berechnen Volumen und Oberflächeninhalt von Prisma, Pyramide und Zylinder.
(L 3) Leitidee Raum und Form	wenden Sätze der ebenen Geometrie bei Konstruktionen und Berechnungen an.
(L 4) Leitidee Funktionaler Zusammenhang	verwenden die Sinusfunktion zur Beschreibung von periodischen Vorgängen.
(L 5) Leitidee Daten und Zufall	werten graphische Darstellungen und Tabellen von statistischen Erhebungen aus.

Weiter werden drei Anforderungsbereiche für die allgemeinen mathematischen Kompetenzen benannt. Der Anforderungsbereich I umfasst die Wiedergabe und direkte Anwendung von grundlegenden Begriffen, Sätzen und Verfahren in einem abgegrenzten Gebiet und einem wiederholenden Zusammenhang. Charakteristisch für den Anforderungsbereich II ist die Bearbeitung bekannter Sachverhalte durch die Verknüpfung von Kenntnissen und Fähigkeiten, die auf verschiedenen Gebieten erworben wurden. Der Anforderungsbereich III zielt auf eigene Lösungen, Begründungen und Folgerungen beim Bearbeiten komplexer Sachverhalte.

Damit gilt als Maßgabe für die Entwicklung von an den Bildungsstandards Mathematik orientierten Kompetenztests in KERMIT, bei der Entwicklung von

⁴ Im Primarbereich gibt es eine etwas andere Bezeichnung der Leitideen: L1 – Zahlen und Operationen; L2 – Größen und Messen; L3 – Raum und Form; L4 – Muster und Strukturen; L5 – Daten, Häufigkeit und Wahrscheinlichkeit

Testaufgaben alle allgemeinen mathematischen Kompetenzen und alle Leitideen etwa gleichmäßig zu berücksichtigen. Gleichzeitig werden in der Aufgabenentwicklergruppe die neuen Aufgaben regelmäßig mit den ausdifferenzierten allgemeinen und inhaltsbezogenen mathematischen Kompetenzen aus den Hamburger Bildungsplänen abgeglichen und bestimmt, in welchen Jahrgangsstufen der Einsatz der Testaufgaben sinnvoll ist.

3.3 Testentwicklung in der Aufgabenentwicklergruppe Englisch

In der Aufgabenentwicklergruppe Englisch werden Aufgaben für die Jahrgangsstufen 5, 7 und 9 entwickelt, die vornehmlich den Kompetenzbereichen Hörverstehen und Leseverstehen zugeordnet sind. Auch die Aufgabenentwicklergruppe Englisch orientiert sich an den nationalen Bildungsstandards und am Gemeinsamen europäischen Referenzrahmen. Die nationalen Bildungsstandards in den Fremdsprachen sind an die im Gemeinsamen europäischen Referenzrahmen (GeR) beschriebenen Niveaus gebunden.

Der GeR beschreibt Kompetenzen von der elementaren Sprachverwendung (A1 bis A2: HSA) über selbstständige Sprachverwendung (B1 bis B2: MSA) bis zur kompetenten Sprachverwendung(C1 bis C2: Abitur). Für die einzelnen KERMIT-Erhebungen werden somit unterschiedliche Niveaustufen erwartet, die von den Schülerinnen und Schülern erreicht werden sollen.

Beispielhaft soll im Folgenden die Aufgabenentwicklung für den Kompetenzbereich Hörverstehen dargestellt werden.

Kompetenzbereich Hörverstehen
Ein Teilbereich der in den Bildungsstandards ausgewiesenen Kompetenzen ist das Hörverstehen. Die Alltagskommunikation besteht zu 45 Prozent aus Hören, aus diesem Grund gilt das Hörverstehen als wichtigste Kompetenz. Sie ist unabdingbar

für die Entwicklung anderer Kompetenzen, insbesondere des Sprechens. Wie schwierig ein Hörtext ist, wird von verschiedenen Eigenschaften beeinflusst (z. B. Grotjahn & Tesch, 2010, S. 133):

- Textlänge
- Anzahl der Sprecher und Sprechgeschwindigkeit
- Art und Lautstärke der Hintergrundgeräusche (unterstützend/störend)
- Verwendete Sprache/Aussprache
- Vertrautheit mit dem Thema
- Textstruktur (einfacher, klarer Aufbau vs. Komplizierte, schwer durchschaubare Gliederung)

Zusätzlich spielt im Rahmen der Kompetenzmessung eine Rolle, welches Hörziel angestrebt wird und in Folge welcher Hörstil eingesetzt wird. Unterschieden wird zwischen globalem Hörverstehen, selektivem und detailliertem Hörverstehen.

Textmapping als Methode zur Erstellung von Testaufgaben

Das Mapping von Texten folgt der Arbeitsweise internationaler

Testentwickler/experten. Die Methode dient der Entwicklung von Testaufgaben.

Ziel ist es, vor dem Erstellen von Aufgaben die am häufigsten wahrgenommenen Inhalte eines Textes oder Tapescripts zu erfassen (Lesen oder Hören). Der kleinste gemeinsame Nenner dieser wahrgenommenen Inhalte bildet die Grundlage für die zu stellenden Aufgaben. Diese Methode erhöht die Validität der Aufgaben, indem sicher gestellt wird, dass entscheidende Inhalte der Texte als Grundlage für die Aufgaben herangezogen werden. Vorher muss aufgrund des Hörstils vor dem Mapping festgelegt werden, ob man sich auf globales, detailliertes oder selektives Hören konzentriert. Mapping und die Erstellung von Testaufgaben erfordern Teamarbeit, die im Sinne von Entwickeln, Erproben und Überarbeiten hilfreich sind.

Dieser Prozess fördert deutlich den kritischen Umgang mit fertigen Aufgabenangeboten. Das Mapping muss in der Aufgabenentwicklergruppe geübt werden (s. Abb. 2).

Julia's school

Listen to Julia talking about her school in America. While listening answer the questions below in 1 to 5 words or numbers. There is an example at the beginning (0).

You will hear the recording twice.
 At the end of the first recording you will have a pause of 10 seconds.
 At the end of the second recording you will have 30 seconds to complete your answers.
 You now have 15 seconds to look at the task.



0 What does Julia think about her school? It's very nice.

1 What does Julia say about the school building?

2 Where can students go in their break?

3 Name two things students can buy in the cafeteria
 a)
 b)

4 What are the teachers like?

5 What do the teachers sometimes do with the students?

Abbildung 2: Beispielaufgabe aus dem Bereich

Mappingergebnisse, Aufgabe Julia's School:

(Konsens des Aufgabenentwicklerteams)

- is very nice
- big building
- nice park

- cafeteria is yummy
- hamburger is delicious
- teachers are very nice
- has a good time at school
- a lot of homework

Auf dieser Grundlage wurden die einzelnen Items der Aufgabe Julia's School entwickelt. Diese werden dann wie in Abschnitt 3 dargestellt fachdidaktisch und psychometrisch geprüft und ggf. weiterentwickelt bzw. überarbeitet.

3.4 Testentwicklung in der Aufgabenentwicklergruppe Naturwissenschaften

Entsprechend den Hamburger Bildungsplänen umfasst der Testbereich Naturwissenschaften in KERMIT die Fächer Biologie, Physik, Chemie und Informatik. Mit dem Ziel einer hohen curricularen Validität orientieren sich die Aufgaben inhaltlich eng an den in den Rahmenplänen ausgewiesenen Kompetenzbereichen und Themenfeldern. Als Grundlage für die Aufgabenkonstruktion dient das Kompetenzmodell, das im Rahmen des Projekts „Evaluation der Standards in den Naturwissenschaften in der Sekundarstufe (ESNaS)“ entwickelt wurde (vgl. Kauertz, Fischer, Mayer, Sumfleth und Walpuski, 2010; Neumann, Kauertz & Fischer, 2007). Diesem Modell nach werden die Kompetenzen durch drei Dimensionen beschrieben (siehe Abb. 3).

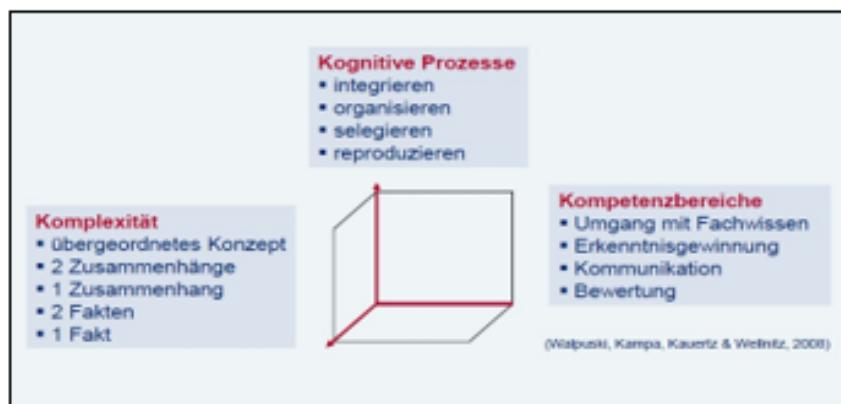


Abbildung 3: Kompetenzmodell von ESNaS

Die erste Dimension ist inhaltlich ausgerichtet und bildet die vier in den nationalen Bildungsstandards vorgeschlagenen Kompetenzbereiche ab: Umgang mit Fachwissen, Erkenntnisgewinnung, Kommunikation und Bewertung. Kompetenzen im Umgang mit Fachwissen beschreiben die Fähigkeit, Phänomene mithilfe von Fachkonzepten zu erklären und konkrete Probleme anhand naturwissenschaftlicher Theorien und Modelle zu lösen. Kompetenzen im Bereich Erkenntnisgewinnung

sind durch das erfolgreiche Anwenden von Methoden und Strategien zur Untersuchung naturwissenschaftlicher Fragestellungen gekennzeichnet. Unter Kommunikation versteht man die adäquate Nutzung naturwissenschaftstypischer Kommunikations- und Darstellungsformen, unter Bewertung schließlich die Fähigkeit, naturwissenschaftliche Informationen und Sachverhalte zu reflektieren und daraus angemessene Entscheidungen abzuleiten.

Die anderen beiden Dimensionen beschreiben schwierigkeitsgenerierende Aufgabenmerkmale: *Komplexität* und *kognitive Prozesse*. Das Merkmal *Komplexität* wird durch die Anzahl der in der Aufgabe zu bearbeitenden Elemente bzw. die Anzahl an erforderlichen Verknüpfungen bestimmt und ist in fünf Stufen unterteilt: (1) Fakt, (2) zwei Fakten, (3) Zusammenhang, (4) zwei Zusammenhänge, (5) übergeordnetes Konzept.

Anhand der Dimension *kognitive Prozesse* soll angegeben werden, welche Denkvorgänge zur Lösung der Aufgabe notwendig sind. Unterschieden werden dabei die Prozesse des Reproduzierens (Identifizieren von Information), des Selegeriens (Auswählen von Information), des Organisierens (Strukturieren von Information) und des Integrierens (Einbinden von Information in die Wissensbasis).

Für die Kompetenzbereiche Umgang mit Fachwissen und Erkenntnisgewinnung gibt es inzwischen langjährige Erfahrungen mit der Konstruktion und dem Einsatz entsprechender Aufgaben (z. B. PISA, 2006; TIMSS: Bos, Bonsen, Kummer et al., 2007). Die Operationalisierung der Kompetenzbereiche Kommunikation und Bewertung ist hingegen noch nicht abgeschlossen. Aus diesem Grund wurden für KERMIT bisher nur Aufgaben für die Bereiche Umgang mit Fachwissen und Erkenntnisgewinnung entwickelt. Erste Hinweise, dass diese beiden Bereiche auch empirisch unterscheidbar sind, liegen mittlerweile vor (Klos, Henke, Kieren, Walpuski und Sumfleth, 2008).

Eine besondere Herausforderung bei der Aufgabenerstellung besteht darin, die Aufgaben kontextuell einzubinden und den Einfluss außerschulisch erworbenen

Vorwissens zu minimieren. Dazu müssen alle relevanten Informationen im Aufgabenstamm vorgegeben sein, ohne das Leseverständnis insbesondere der schwächeren Schülerinnen und Schüler zu überfordern. Zudem sollen die Aufgaben nicht zu viel Zeit in Anspruch nehmen, um dem Kriterium der Testökonomie zu genügen, und sie sollten für alle Schülergruppen gleichermaßen motivierend sein. Das Kompetenzmodell von ESNaS dient dabei der Aufgabenentwicklergruppe als Leitfaden. Durch die im Modell beschriebenen drei Dimensionen ist es möglich, Aufgaben mit unterschiedlichen inhaltlichen und kognitiven Anforderungen zu realisieren. Gleichzeitig werden die neuen Aufgaben mit den in den Hamburger Bildungsplänen beschriebenen inhaltlichen und prozessbezogenen Kompetenzen abgeglichen, so dass unterschiedlich anspruchsvolle Aufgaben entstehen, die gleichzeitig einen repräsentativen Bereich der naturwissenschaftlichen Bildung erfassen können.

4. Die Darstellung der Ergebnisse in den KERMIT-Rückmeldungen

Die Ergebnisse der KERMIT-Erhebungen geben den Leistungsstand von Lerngruppen in zentralen Bereichen der getesteten Fächer wieder und dienen der Anpassung und Weiterentwicklung des Unterrichts an den Schulen. Eine innerschulische Reflexion über die Ergebnisse kann zudem Impulse für die Schul- und Unterrichtsentwicklung geben. Den Schulen werden die KERMIT-Ergebnisse in verschiedenen Versionen zur Verfügung gestellt, um den jeweiligen Adressaten die Nutzung zu erleichtern. Neben einer klassenbezogenen Rückmeldung, die sich vor allem an die jeweiligen Fachlehrkräfte und Klassenlehrkräfte richtet, wird auch eine schulbezogene Rückmeldung ausgegeben, die von der Schulleitung und dem Leitungsteam der Schule (Stufen- und Fachkoordinatorinnen und -koordinatoren) zur weiteren Schulentwicklung herangezogen werden sollte. Darüber hinaus erhalten die Schulen eine Präsentation, in der die Ergebnisse unter Einhaltung des Datenschutzes so aufbereitet werden, dass sie der Schulöffentlichkeit (z. B. auf einem Elternabend) präsentiert werden können.

Die Testergebnisse von KERMIT 2, 5, 7 und 9 werden pro Kompetenzbereich unter Zugrundelegung des eindimensionalen Raschmodells skaliert und analog zu (internationalen Schulleistungsuntersuchen (z. B. PISA, TIMSS, IQB-Ländervergleich) auf einen Mittelwert von 500 Punkten (Standardabweichung: 100 Punkte) transformiert. Somit beträgt der Mittelwert aller Hamburger Schülerinnen und Schüler einer Jahrgangsstufe in jedem Testbereich 500 Punkte. Als Richtwert zur Interpretation der Ergebnisse dient ein Unterschied von 30 Punkten, der auf Gruppenebene (Klasse oder größere Gruppe) als bedeutsam zu interpretieren ist. Differenzen, die kleiner als 30 Punkte sind, werden als nicht bedeutsam bewertet (siehe auch PISA-Konsortium: OECD, 2003). Anders als die ausschließlich in Hamburg eingesetzten Verfahren KERMIT 2, 5, 7 und 9 werden die Erhebungen KERMIT 3 und 8, wie bereits beschrieben, bundesweit durchgeführt. Hier wurden die Testergebnisse so transformiert, dass alle Schülerinnen und Schüler der

deutschlandweiten Normierungsstichprobe im Durchschnitt 500 Punkte erzielten.

Auch hier sind Unterschiede von 30 Punkten als bedeutsam zu interpretieren.

Im Folgenden sollen die einzelnen Ergebnisdarstellungen anhand von Beispielergebnissen vorgestellt werden. Dabei wird erläutert, wie diese Ergebnisse gelesen und interpretiert werden können, und wie daraus Vorschläge zur weiteren Unterrichts- und Schulentwicklung abgeleitet werden können.

In der jahrgangsbezogenen Rückmeldung werden für jeden getesteten Kompetenzbereich zunächst die Mittelwerte der Klasse im Vergleich zu verschiedenen Bezugsgruppen dargestellt (siehe Abb. 4). Die erste Bezugsgruppe, mit denen der Mittelwert der Klasse verglichen werden kann, ist der Mittelwert der gesamten Schule. Dieser Vergleich ermöglicht eine Einschätzung, ob die jeweilige Klasse in dem Fach deutlich von den Werten der anderen Klassen abweicht oder nicht. Des Weiteren wird für jede Schule eine Vergleichsgruppe von sechs Schulen bestimmt, die hinsichtlich der sozialen Zusammensetzung der Schülerschaft möglichst ähnlich ist. Zu diesem Zweck wird auf den Hamburger Sozialindex zurückgegriffen, der für jede Schule die soziale Belastung beschreibt (IfBQ, 2013). Die Gegenüberstellung mit diesen Vergleichsschulen ermöglicht es den Lehrkräften und Schulleitungen einzuschätzen, ob sich die erreichten Leistungswerte der Klasse und der Schule in einem vergleichbaren Rahmen mit den Leistungswerten anderer Schulen befinden, die hinsichtlich verschiedener sozialer Belastungsfaktoren eine ähnliche Schülerschaft besitzen (Fiege, Reuther & Nachtigall, 2011). Zusätzlich werden als Bezugsgruppe die Mittelwerte für alle Schülerinnen und Schüler in den beiden Hamburger Schulformen dargestellt. In den schulbezogenen Ergebnisrückmeldungen wird nicht nur das Ergebnis der einen Klasse im Vergleich zu den Bezugsgruppen dargestellt, sondern auch die Mittelwerte für alle vorhandenen Klassen eines Jahrgangs gezeigt. Diese Darstellung erlaubt es dem Schulleitungsteam auch einen Vergleich der Leistungen innerhalb einer Jahrgangsstufe zu ziehen.

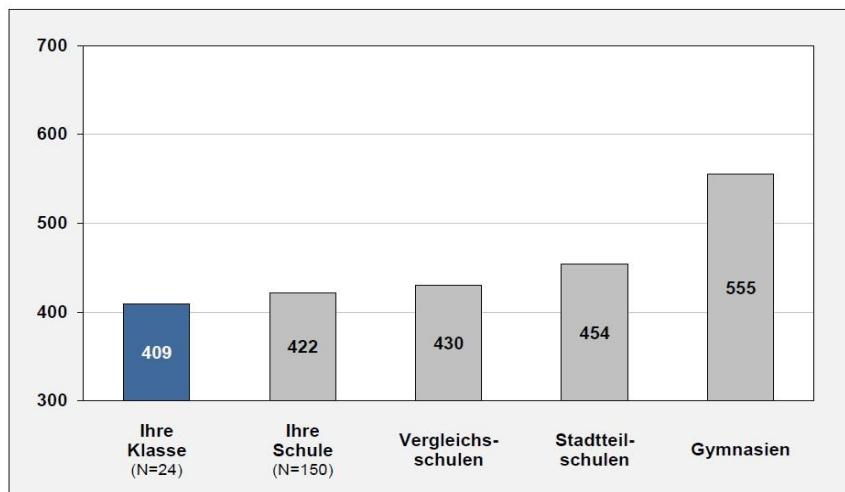


Abbildung 4: Deutsch Leseverstehen - Mittlere Leistungswerte einer fiktiven Stadtteilschule in Hamburg im Vergleich zu verschiedenen Bezugsgruppen

Die Darstellung der Mittelwerte liefert lediglich Informationen zur zentralen Tendenz der Ergebnisse innerhalb einer Klasse bzw. Schule. Nicht abgebildet werden damit die Homogenität bzw. Heterogenität sowie die Spannbreiten der Leistungen in den einzelnen Gruppen. Aus diesem Grunde werden den Schulen zusätzlich Informationen über die Leistungsverteilungen der verschiedenen Gruppen zur Verfügung gestellt (siehe Abb. 5). Dazu wurde zunächst die Gesamtstichprobe aller Hamburger Schülerinnen und Schüler aus einer Testung (eines Jahrgangs) gemäß ihrer Testleistung der Reihe nach geordnet und in vier gleich große Leistungsgruppen unterteilt. Anschließend wird bestimmt, wie viele Punkte eine Schülerin / ein Schüler erreichen muss, um in die jeweilige Leistungsgruppe zu gehören. Die 25 Prozent Schülerinnen und Schüler mit den schwächsten Leistungen bilden den Leistungsbereich I, im Leistungsbereich II folgen dann die Schülerinnen und Schüler mit etwas besseren Leistungen, bis hin zu denjenigen Lernenden, die sich im oberen Leistungsviertel (= Leistungsbereich IV) befinden. In der Ergebnisrückmeldung wird für die einzelne Gruppen dargestellt, wie viele Schülerinnen und Schüler sich in den vier verschiedenen Leistungsvierteln befinden.

So erkennt man in der Abbildung 5, dass in dieser Klasse 59 Prozent der Schülerinnen und Schüler zu den 25 Prozent der leistungsschwächsten Schülerinnen und Schülern Hamburgs gehören, und nur 4 Prozent der Klasse eine vergleichbar gute Leistung erzielt haben wie das oberste Leistungsviertel in Hamburg. Auch hier geben die verschiedenen Bezugsgruppen eine Orientierung für die Lehrkraft, inwieweit die Verteilung der Leistung der Klasse den Erwartungen entspricht und ob beispielsweise die Besetzung der obersten Leistungsgruppe der entsprechenden Schulform oder der Gruppe der Vergleichsschulen entspricht.

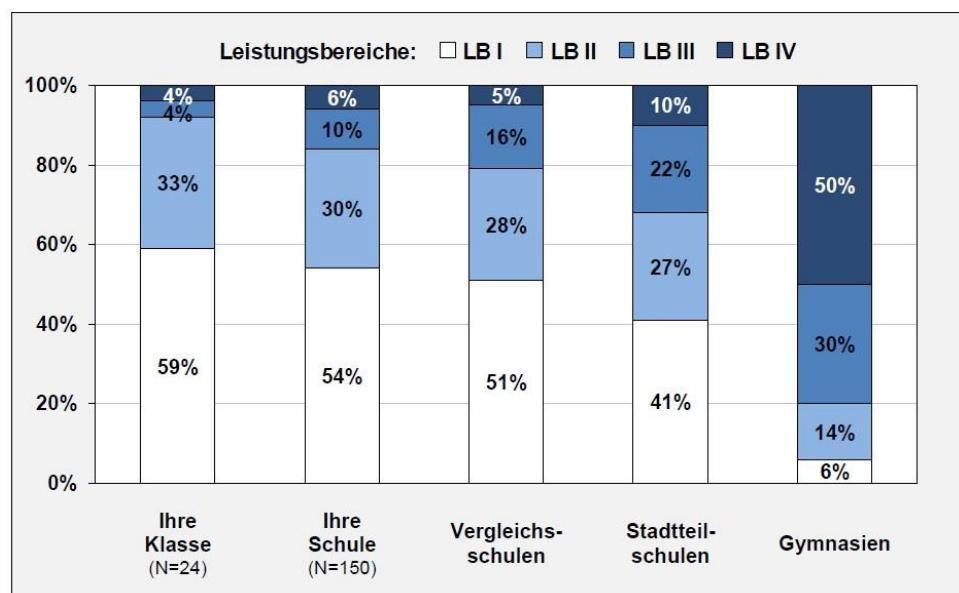


Abbildung 5: Deutsch Leseverstehen - Leistungsverteilungen der Klasse im Vergleich zu den Bezugsgruppen

In der klassenbezogenen Rückmeldung werden zusätzlich die individuellen Leistungsdaten aller Schülerinnen und Schüler einer Klasse für alle getesteten Bereiche abgebildet (S. Abb. 6). Für jede Schülerin oder Schüler, die/der über einen Schülercode für die Lehrkraft identifizierbar ist, werden die Leistungswerte und die Zugehörigkeit zu den oben beschrieben vier Leistungsgruppen dargestellt. Bei den Ergebnissen für KERMIT 5 werden zusätzlich für die beiden Kompetenzbereiche

Deutsch-Leseverstehen und Mathematik ausgegeben, ob die Schülerinnen und Schüler den für das Ende der Grundschulzeit geltenden Mindeststandard oder Regelstan-

Aufgrund der Einpunkt-Messung sind diese individuellen Leistungsdaten mit einem höheren Messfehler behaftet, d. h. sie erlauben keine eindeutigen Aussagen über die erreichten Kompetenzen jeder einzelnen Schülerin / jedes einzelnen Schülers.

Dennoch geben diese Ergebnisse Hinweise auf die individuellen Leistungen der Schülerinnen und Schüler sollten mit der Erfahrung und der Einschätzung der Lehrkraft abgleichen werden. Unerwartete Ergebnisse sollten immer einen Anlass bieten, sich genauer mit der Schülerin oder dem Schüler auseinanderzusetzen und über die eigene Einschätzung zu reflektieren.

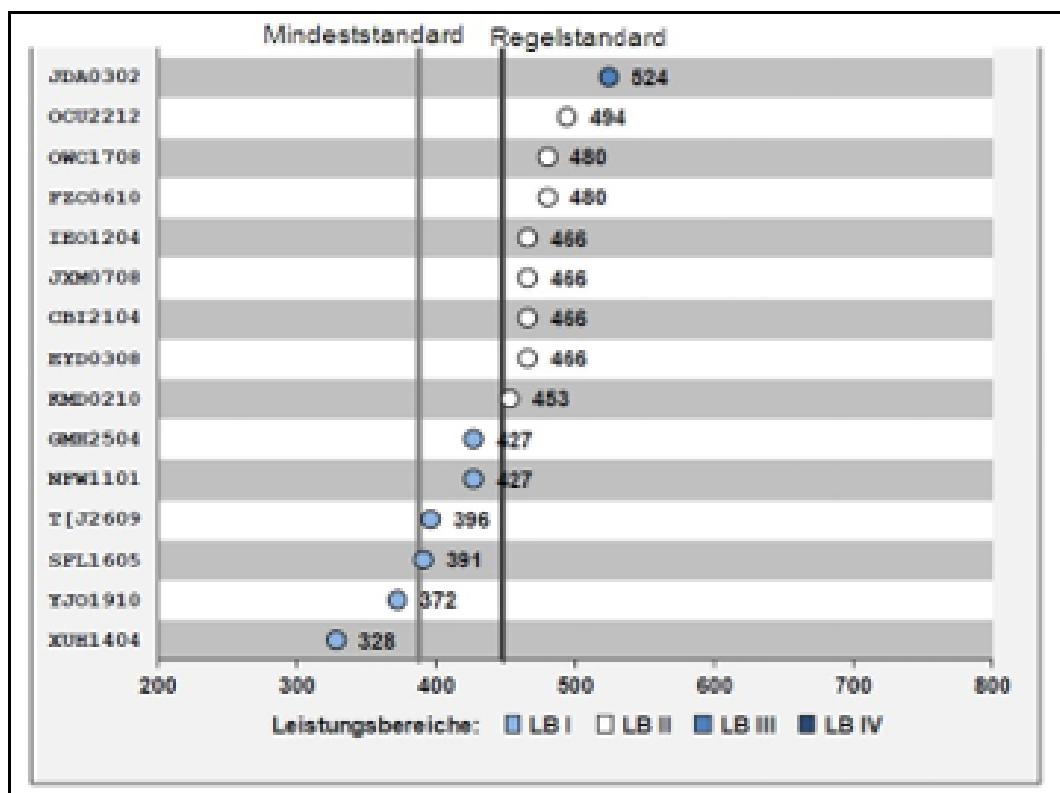


Abbildung 6: Darstellung der individuellen Schülerergebnisse

Während sich die bisherigen Ergebnisdarstellungen nur auf einen momentanen Leistungsstand der Klassen sowie Schülerinnen und Schülern beziehen, bietet KERMIT – wie bereits oben dargestellt – die Möglichkeit, auch die Lernentwicklung zu verfolgen. Dazu werden in weiteren Abbildungen in den Rückmeldungen auch die Leistungsstände der jeweils vorherigen KERMIT-Erhebung abgebildet und der entsprechende Lernzuwachs über die Schuljahre (s. Abbildung 7). Anhand dieser Darstellung können Lehrkräfte und Schulleitungen erkennen, ob die durchschnittliche Leistungsentwicklung mit den Bezugsgruppen vergleichbar ist. Deutliche Abweichungen von dieser Entwicklung lassen vermuten, dass die Klasse nicht optimal (bei negativen Differenzen) bzw. besonders gut (bei positiven Differenzen) gefördert wurde.

Im oberen Beispiel ist zu erkennen, dass die Klasse sowohl im Jahrgang 5 als auch in 7 einen unterdurchschnittlichen Leistungsstand erzielt hat, ihr Leistungszuwachs jedoch vergleichbar mit dem der Bezugsgruppe ist. Die Klasse konnte daher bei schwierigen Voraussetzungen zu Beginn von Jahrgang 5 ähnliche Lernfortschritte erzielen wie sozial vergleichbare Klassen. Gerade die für die Schulen wertvolle Betrachtung von Leistungsentwicklungen im sozialen Vergleich ist nur durch eine Längsschnittstudie möglich.

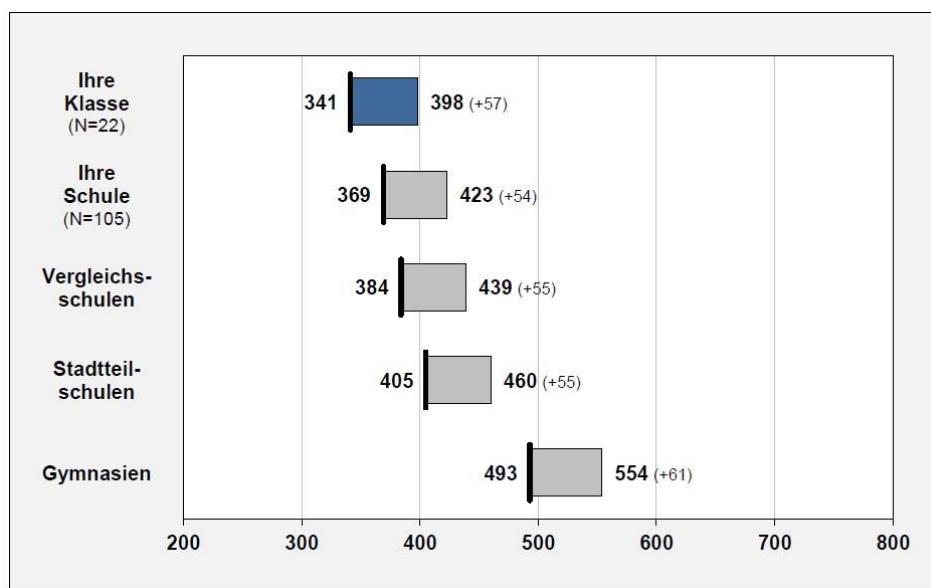


Abbildung 7: : Darstellung der mittleren Leistungsentwicklung von Jahrgang 5 zu

5. Von der Ergebnisreflektion hin zu einer systematischen Schul- und Unterrichtsentwicklung

Nach der umfassenden Analyse der Ergebnisse besteht die innerschulische Herausforderung in der Interpretation dieser Ergebnisse, um daraus adäquate und sinnvolle Maßnahmen abzuleiten. Um Ursachen für die jeweiligen Ergebnisse zu identifizieren, müssen die Lehrkräfte und Leitungsteams die Bedingungsfaktoren des Unterrichts und der Lerngruppen differenziert in den Blick nehmen. Ursachen für unzureichende Ergebnisse sind oft vielschichtig und komplex, eindimensionale Erklärungsversuche und Schuldzuschreibungen sollten deshalb vermieden werden. In der Regel lassen sich Unterschiede im Leistungsvermögen auf zwei große Bereiche zurückführen: Zum einen liegen die Ursachen in den Bedingungen der jeweiligen Lerngruppe und deren Lernvoraussetzungen. Der Fokus liegt in diesem Fall auf der Zusammensetzung der Lerngruppe. Zum anderen können die Ursachen in den unterrichtlichen Bedingungen und schulischen Voraussetzungen liegen. Da Unterricht ein überaus komplexer Vorgang ist, wird es selten gelingen, die Ursachen für schwache oder sehr gute Ergebnisse eindeutig zu identifizieren. In der Regel liegt die Erklärung in einer Kombination verschiedener Aspekte vor Ort. Aber die eingehende Interpretation der Ergebnisse in Verbindung mit den Erfahrungen aus der Schulentwicklung zeigen, dass sich allein durch eine gemeinsame Auseinandersetzung mit den Ergebnissen von Schulleistungstests im Kollegium und der Umsetzung einzelner Maßnahmen erste Verbesserungen erzielen lassen (Metzeld, et al., 2009).

Bei der Reflexion der Ergebnisse ist eine genaue Identifizierung von Stärken und Schwächen eine wesentliche Voraussetzung für die zielgerichtete Ursachenforschung und für die Ableitung erfolgversprechender Maßnahmen zur Unterrichts- und Schulentwicklung. Dabei sollte sich die Analyse der Ergebnisse nicht nur auf Entwicklungsbedürfnisse fokussieren, sondern auch auf Stärken und Kompetenzen, die bei den Schülerinnen und Schülern bereits vorhanden sind. Je

nach Ebene können unterschiedliche Leitfragen die Reflexion der Ergebnisse vertiefen.

Auf der individuellen Ebene der Schülerinnen und Schüler stehen die Leistungen einer einzigen Klasse bzw. der einzelnen Schülerinnen und Schüler dieser Klasse im Zentrum. Auf dieser Ebene reflektiert vor allem die Fachlehrkraft, ob und was in der zukünftigen Unterrichtsplanung verändert werden kann oder soll. Dabei können Fragen gestellt werden, wie z. B.: Decken sich die Einzelergebnisse mit den Erfahrungen der Fachlehrkraft? Wie homogen bzw. heterogen ist die Leistung in der Klasse? Oder auch: Weisen die Leistungsverteilungen der Klasse für die verschiedenen Testbereiche ähnliche bzw. unterschiedliche Muster auf?

Die zweite Ebene stellt die Aufgabenebene dar. Die Rückmeldung bietet vielfältige Reflektionsanlässe zum Vergleich der Fächer, Kompetenzbereiche und Leitideen. In den KERMIT-Erhebungen werden ergänzend zum Gesamtergebnis des jeweiligen Faches auch differenzierte Ergebnisse für Teilbereiche zur Verfügung gestellt. Beispielsweise wird für Mathematik angegeben, wie viele Aufgaben der fünf verschiedenen mathematischen Leitideen die Schülerinnen und Schüler jeweils lösen konnten. Darauf aufbauend können Lehrkräfte überlegen, welche Leitideen im Unterricht noch intensiver bearbeitet werden sollten. Für den Testbereich Naturwissenschaften werden die Gesamtergebnisse zusätzlich nach den Einzelfächern (Biologie vs. Physik/Chemie/Informatik) und den beiden Kompetenzbereichen ‚Umgang mit Fachwissen‘ und ‚Erkenntnisgewinnung‘ ausgewiesen. Bei KERMIT 3 und 8 werden jedes Jahr komplett neue Aufgaben eingesetzt, so dass die Aufgaben nach der Testung ergänzt durch umfangreiche didaktische Materialien den Lehrkräften zur Verfügung gestellt werden können. Daher gibt es für diese beiden KERMIT-Erhebungen in den klassenbezogenen Rückmeldungen eine ausführliche Auflistung der Lösungshäufigkeiten jeder Aufgabe. Mit Hilfe dieser Darstellung und den zur Verfügung gestellten didaktischen Materialien können die Lehrkräfte planen, welche Aufgaben im Unterricht wiederholt bzw. intensiviert werden sollten.

Die schulbezogene Ergebnisrückmeldung schließlich liefert Reflektionsanlässe auf Schulebene. Der Vergleich der Ergebnisse zwischen den verschiedenen Klassen sollte die Schulleitung bzw. das Schulleitungsteam dazu anregen, Maßnahmen für die weitere Schulentwicklung abzuleiten. Wenn es auffällige Kompetenzunterschiede zwischen den Klassen gibt, sollten die Gründe dafür diskutiert werden und gegebenenfalls Möglichkeiten für eine Optimierung der Situation ausgelotet werden. Die Gründe für die Leistungsunterschiede können in der Zusammensetzung der Klassen liegen, wenn z. B. überwiegend eher leistungsstarke oder leistungsschwache Schülerinnen und Schüler einer Klasse zugeordnet wurden. Aber auch Unterrichtsausfall, Lehrerwechsel oder der Unterrichtsansatz einer Lehrkraft kann zu Unterschieden in den Ergebnissen führen. In diesem Fall kann im Kollegium überlegt werden, welche Kooperationsstrukturen aufgebaut werden, damit z. B. das Kollegium noch besser voneinander lernen kann.

Der wichtigste Schritt für die Schul- und Unterrichtsentwicklung ist schließlich die Ableitung konkreter Maßnahmen aus der Analyse der Ergebnisse. Je nachdem, welche Ursachen für das jeweilige Analyseergebnis identifiziert wurden, bieten sich vielfältige Maßnahmen an, die durch einzelne Lehrkräfte, Fachschaften oder die Schulleitung getroffen werden können und müssen. Dazu müssen wiederum alle Ebenen und Akteure involviert werden. So kann die einzelne Lehrkraft beispielsweise ihren Unterricht inhaltlich und methodisch optimieren und auf Fortbildungen Anregungen für einen binnendifferenzierten, kompetenzorientierten Unterricht erhalten. In der Fachschaft sollte der Material- und Ideenaustausch gefördert werden, das schulinterne Curriculum kann abgestimmt und gegenseitige Unterrichtsbesuche angeregt werden. Die Schulleitung schließlich kann entscheiden, ob z. B. Mittel für zusätzliche Fördermaßnahmen bereitgestellt werden, die Zusammensetzung der Lerngruppen verändert werden sollte oder auch schulinterne Fortbildungen initiiert werden sollen. Einige dieser Maßnahmen werden anlassbezogen zu treffen sein, ein großer Teil sollte als selbstverständlicher und

dauerhafter Bestandteil der schulinternen Qualitätsentwicklung in den Schulen implementiert werden.

6. Diskussion und Ausblick

KERMIT ist ein Instrument zur regelhaften Kompetenzermittlung an Hamburger Schulen. Die von den Schülerinnen und Schülern zu bearbeitenden Aufgaben wurden zum einen eigens für KERMIT entwickelt und orientieren sich an den Anforderungen und Zielen der Hamburger Bildungspläne. Dadurch ist die curriculare Validität gewährleistet. Zum andern sind die bundesweiten VERA-Erhebungen in den Jahrgangsstufen 3 und 8 in KERMIT integriert. Durch die Verwendung der VERA-Aufgaben erhalten die Schulen zusätzlich Hinweise darauf, auf welcher Kompetenzstufe sich ihre Schülerinnen und Schüler im Hinblick auf die Anforderungen der nationalen Bildungsstandards für die Fächer Deutsch, Mathematik und Englisch befinden. Schulleitungen und Lehrkräfte, die sich mit den KERMIT-Ergebnissen auseinandersetzen und sie unter verschiedenen Gesichtspunkten analysieren und interpretieren, können all diese Informationen dafür nutzen, ihren Unterricht auf unterschiedlichen Ebenen systematisch weiterzuentwickeln.

Mit KERMIT ist es prinzipiell möglich, den Kompetenzstand der Schülerinnen und Schüler einer Schule / einer Klasse aber auch auf Individualebene in den getesteten Bereichen zu ermitteln und deren Kompetenzentwicklung über mehrere Jahre zu verfolgen. Die Lehrkräfte und Jahrgangsteams erhalten durch KERMIT Hinweise zum individuellen Kompetenzstand der Lernenden ihrer Schule / ihrer Lerngruppen (Klassen) im Hinblick auf die zu erreichenden Ziele und sie erhalten in regelmäßigen Abständen Informationen darüber, inwieweit es ihnen gelungen ist, den jeweiligen Zielvorstellungen näher zu kommen. Dies ist der kritiale Bezug, den KERMIT bietet. Darüber hinaus werden die KERMIT-Ergebnisse auch in einen sozialen Kontext gestellt. Die individuellen Ergebnisse der Schülerinnen und Schüler sind untereinander vergleichbar und geben damit Aufschluss über die Leistungsheterogenität innerhalb einer Lerngruppe (Klasse). Der Kompetenzstand und die Kompetenzentwicklung mehrerer Lerngruppen sind wiederum untereinander

vergleichbar und können in Beziehung gesetzt werden zu den Ergebnissen, die an allen Schulen derselben Schulform erzielt wurden oder auch an Schulen, die eine sozial ähnlich zusammengesetzte Schülerschaft aufweisen. Auch im Spannungsfeld zwischen schulaufsichtlicher Beratung und Kontrolle können die KERMIT-Ergebnisse dazu beitragen, verbindlich Art und Umfang von erforderlichen (unterstützenden) Maßnahmen zu vereinbaren.

KERMIT bietet grundsätzlich auch Steuerungswissen, das nicht nur für Lehrkräfte und Schulleitungen von Interesse ist, sondern in aggregierter Form und unter Einhaltung des Datenschutzes auch anderweitig genutzt werden könnte. Dies wäre insbesondere dann der Fall, wenn die KERMIT-Ergebnisse mit anderen Daten verknüpft bzw. in Beziehung gesetzt würden. Diese Option weckt bei unterschiedlichen Planungsstellen und Entscheidungsträgern Begehrlichkeiten, die zwar verständlich sind, für den ursprünglichen unterrichtlich-pädagogischen Zweck der Erhebung aber auch bedrohlich werden könnten. Diese Gefahr besteht insbesondere dann, wenn die Lehrerinnen und Lehrer befürchten (müssen), dass die Testergebnisse oberflächlich oder falsch interpretiert werden könnten und damit ihr eigentlicher Zweck, die Schul- und Unterrichtsqualität zu verbessern, konterkariert wird.

Von KERMIT betroffene Lehrerinnen und Lehrer nahezu aller Hamburger Schulen reagieren bisher unaufgeregt auf die verbindliche Einführung der Tests und blicken ausreichend selbstbewusst auf ihre Arbeit, um die erreichten Kompetenzen ihrer Schülerinnen und Schüler und damit auch ihre eigene Arbeit mit standardisierten Verfahren „von außen“ messen zu lassen. Zur hohen Testakzeptanz der Hamburger Schulen trägt möglicherweise auch bei, dass in Hamburg – anders als in den übrigen Bundesländern – der mit einer Testung verbundene Arbeitsaufwand für die Schulen Dank eines aufwändigen außerschulischen Unterstützungssystems insgesamt relativ gering ist. Die Schulen erhalten alle Unterlagen, die für die Testung notwendig sind, frei Haus; die bearbeiteten Testhefte werden nach der Testung wieder abgeholt und

extern ausgewertet. Anschließend erhalten die Schulen eine detaillierte Ergebnisrückmeldung. Die Ergebnisse bieten Anlass zur Reflektion der Lernvoraussetzungen innerhalb der Lerngruppe, der eigenen pädagogisch-didaktischen Arbeit, aber auch zur Diskussion im Kollegenkreis und auf schulischer Entscheidungsebene. Damit überwiegt aus schulischer Sicht der potentielle Nutzen eines KERMIT-Tests offensichtlich die mit einer Testdurchführung verbundene „Störung“ des Schulalltags.

Um die Qualität von KERMIT zu sichern und weiter auszubauen, werden sowohl auf methodischer Ebene als auch auf Ebene der Implementation von KERMIT in den Schulen vielfältige Prozesse initiiert und fortgesetzt. Zur Sicherstellung der methodischen Qualität von KERMIT wird kontinuierlich überprüft, ob sich die Kompetenzentwicklung der Schülerinnen und Schüler in den Lerngruppen valide und reliabel abbilden lässt. Um dies zu erreichen werden weiterhin qualitativ hochwertige und an den Bildungsplänen angepasste Aufgaben, die zur Messung der längsschnittlichen Lernentwicklung geeignet sein sollen, entwickelt und pilotiert. Zur Sicherstellung einer hohen Aufgabenqualität werden die Aufgabenentwicklerinnen und –entwickler darüber informiert, wie ihre Aufgaben in den Pilotierungen, aber auch im Testeinsatz abgeschnitten haben. Dieses Feedback kann dann für die weitere Aufgabenentwicklung genutzt werden. Zusätzlich werden regelmäßig die Ergebnisse von online-Befragungen genutzt, um KERMIT hinsichtlich der Vorbereitung und Durchführung, aber auch die Ergebnisdarstellung zu optimieren.

Zur Unterstützung der Implementation von KERMIT in den Schulen werden verschiedene Maßnahmen fortgesetzt, welche die Akzeptanz dieses Instrumentes und damit auch den Nutzen zu erhöhen. So bietet es sich zum einen an, die Verortung der Aufgaben an den Hamburger Bildungsplänen noch stärker herauszustellen. Damit ist es den Lehrkräften möglich, die Leistungen ihrer Schülerinnen und Schüler noch deutlicher in Beziehung zu den Bildungsplänen zu

stellen und Maßnahmen für ihren weiteren Unterricht abzuleiten, die konkret auf die in den Bildungsplänen beschriebenen Ziele bezogen sind. Zum anderen sollen die KERMIT-Erhebungen durch didaktische Hinweise angereichert werden, wie sie bereits jetzt für KERMIT 3 und KERMIT 8 zur Verfügung stehen. Da die Aufgaben bei KERMIT 5, 7 und 9 geschützt sind und nicht veröffentlicht werden, können diese zwar nicht explizit dargestellt werden, dennoch ist es möglich, beispielhaft aufzuzeigen, welche Aufgaben im Rahmen von KERMIT eingesetzt werden und wie die Ergebnisse der Schülerinnen und Schüler zu interpretieren sind. In begleitenden Fortbildungsmaßnahmen und Informationsveranstaltungen werden die verschiedenen Akteure im Hamburger Bildungssystem über die Möglichkeiten und Grenzen von KERMIT unterrichtet und gleichzeitig motiviert, sich mit datengestützter Schul- und Unterrichtsentwicklung auseinanderzusetzen.

Dieser umfangreiche Maßnahmenkatalog ermöglicht es, dass KERMIT auch in Zukunft in den Schulen und darüber hinaus als sinnvolles und praktikables Instrument der datengestützten Schul- und Unterrichtsentwicklung wahrgenommen und benutzt werden kann.

7. Literatur

- Behörde für Schule und Berufsbildung (Hrsg.) (2011). LAU – Aspekte der Lernausgangslage und der Lernentwicklung Klassenstufen 5, 7 und 9, HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Band 8.
- Berkemeyer, N. & van Holt, N. (2012). Leistungsrückmeldungen im Längsschnitt. Erste Erfahrungen mit dem Schüler-Monitoring-System (SMS). In A. Wacker, U. Maier & J. Wissinger (Hrsg.): *Schul- und Unterrichtsreform durch ergebnisorientierte Steuerung*. Wiesbaden: Springer VS (2012) S. 109-130
- Bos, W., Bonsen, M., Kummer, N., Lintorf, K., Frey, K. (Hrsg.) (2009). TIMSS 2007. *Dokumentation der Erhebungsinstrumente zur Trends in International Mathematics and Science Study*. Münster: Waxmann.
- Bos, W. & Gröhlich , G. (Hrsg.) (2010). *KESS 8 – Kompetenzen und Einstellungen von Schülerinnen und Schülern am Ende der Jahrgangsstufe 8*, HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Band 6.
- Bos, W. & Gröhlich, C. & Pietsch, M. (Hrsg.) (2006). *KESS 4 – Lehr- und Lernbedingungen in Hamburger Grundschulen*, HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Band 2.
- Bos, W., Lankes, E.-M., Prenzel, M. & Schwippert, K. (Hrsg.). (2003). *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Brackhahn, B. & Brockmeyer, R. (2004). Qualitätsverbesserung in Schulen und Schulsystemen – QuiSS (Bd. 1: Konzept - Organisation – Ergebnisse). Köln: Luchterhand.
- Bremerich-Vos, A. & Böhme, K. (2009). Lesekompetenzdiagnostik – die Entwicklung eines Kompetenzstufenmodells für den Bereich Lesen. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther

(Hrsg.), *Bildungsstandards Deutsch und Mathematik. Leistungsmessung in der Grundschule* (S. 219-249). Weinheim: Beltz.

Fiege, C., Reuther, F., & Nachtigall, C. (2011). Faire Vergleiche? – Berücksichtigung von Kontextbedingungen des Lernens beim Vergleich von Testergebnissen aus deutschen Vergleichsarbeiten. *Zeitschrift für Bildungsforschung*, 2, 133 – 149.

Frey, A., Carstensen, C. H., Walter, O., Rönnebeck, S. & Gomolka, J. (2008). Methodische Grundlagen des Ländervergleichs. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006 in Deutschland: Die Kompetenzen der Jugendlichen im dritten Ländervergleich* (S. 375–397). Münster: Waxmann.

Grotjahn, R. & Tesch, B. (2010). Messung der Hörverstehenskompetenz im Fach Französisch. In: R. Porsch, B. Tesch & O. Köller (Hrsg.). *Standardbasierte Testentwicklung und Leistungsmessung: Französisch in der Sekundarstufe I*. Münster: Waxmann.

Helmke, A. & Jäger, R. S. (Hrsg.). (2002). *Das Projekt MARKUS. MathematikGesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. (VEP-Aktuell, Bd. 2). Landau: Verlag Empirische Pädagogik.

Hornberg, S., Bos, W., Buddeberg, I., Potthoff, B. & Stubbe, T. C. (2007). Anlage und Durchführung von IGLU 2006. In W. Bos, S. Hornberg, K.-H. Arnold, G. Faust, L. Fried, E.-M. Lankes, K. Schwippert & R. Valtin (Hrsg.), *IGLU 2006. Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (S. 21–79). Münster: Waxmann.

Hundt, S. (2009). „Vom Wiegen wird das Schwein nicht fett!“ Bremens Grundschüler auf dem letzten Platz beim Lesetest IGLU. Zugriff am: 28.03.2013 unter:

[http://www.dielinkebremen.de/index.php?id=16781&no_cache=1&tx_ttnews\[tt_news\]=10035&tx_ttnews\[backPid\]=16779](http://www.dielinkebremen.de/index.php?id=16781&no_cache=1&tx_ttnews[tt_news]=10035&tx_ttnews[backPid]=16779).

Institut für Bildungsmonitoring und Qualitätsentwicklung (IfBQ). (2013).
Methodisches Vorgehen bei der aktualisierung des Sozialindex für Hamburger Schulen.

Zugriff am: 28.03.2013 unter:
<http://www.bildungsmonitoring.hamburg.de/index.php/file/download/1737>.

Kauertz, A., Fischer, H. E., Mayer, J., Sumfleth, E. & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe I. *Zeitschrift für Didaktik der Naturwissenschaften*. 16,135-154

Klos, S., Henke, C., Kieren, C., Walpuski, M., & Sumfleth, E. (2008). Naturwissenschaftliches Experimentieren und chemisches Fachwissen – zwei verschiedene Kompetenzen. *Zeitschrift für Pädagogik* 3, 304-321.

KMK (2006). Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring. Zugriff am: 28.03.2013 unter:
http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2006/2006_06_02Bildungsmonitoring.pdf.

Köller, O., Knigge, M. & Tesch, B. (Hrsg.). (2010). *Sprachliche Kompetenzen im Ländervergleich*. Münster: Waxmann.

Kühle, B. & Ackeren, I. van (2012): Wirkungen externer Evaluationsformen für eine evidenzbasierte Schul- und Unterrichtsentwicklung. In: M. Ratermann & S. StöbeBlossey (Hrsg.): *Governance von Schul- und Elementarbildung - Vergleichende Betrachtungen und Ansätze der Vernetzung*. Wiesbaden: Springer VS Verlag, S. 4562.

Lange, H. (2008). Vom Messen und Handeln: „empirische Wende“ der Bildungspolitik? *Recht der Jugend und des bildungswesens*, 56 (1), 7-15.

R. H. Lehmann, R. Peek, R. Gänsfuß, V. Husfeldt (2002): *Aspekte der Lernausgangslage und der Lernentwicklung – Klassenstufe 9. Ergebnisse einer Längsschnittsuntersuchung in Hamburg*. Hamburg.

- Metzeld, D., Isaac, K., Groß Ophoff, J., Speck-Hamdan, A. & Böhme, K. (2009). Warum sich der Aufwand lohnt - Der Beitrag von Vergleichsarbeiten zum kompetenzorientierten Unterrichten. *Grundschulunterricht*, 56 (2), 4-8.
- Neumann, K., Kauertz, A., Lau, A., Notarp, H. & Fischer, H. E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 125-143.
- OECD. (2003). The PISA 2003 assessment framework mathematics, reading, science and problem solving knowledge and skills. Paris: OECD.
- Peek, R. (2004). Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik (QuaSUM) - Klassenbezogene Ergebnisrückmeldungen und ihre Rezeption in Brandenburger Schulen. *Empirische Pädagogik*, 18(1), 82-114.
- Rupp, A. A., Vock, M., Harsch, C. & Köller, O. (2008). *Developing standardsbased assessment tasks for English as a first foreign language – Context, processes, and outcomes in Germany*. Münster: Waxmann.
- Vieluf, U., Ivanov, S. & Nikolova, R. (Hrsg.) (2012). *KESS 10/11 – Kompetenzen und Einstellungen von Schülerinnen und Schülern an Hamburger Schulen am Ende der Sekundarstufe I und zu Beginn der gymnasialen Oberstufe*, HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Band 10.
- Zabka, T. (2008). Interpretationsverhältnisse entfalten. Vorschläge zur Analyse und Kritik literaturwissenschaftlicher Bedeutungszuweisungen. *Journal of Literary Theory* 2(3), 51-69.