Departement
Bildung, Kultur und Sport
Generalsekretariat
Bachstrasse 15
5001 Aarau
www.ag.ch/bks

BKS

BILDUNG

Handreichung zum Umgang mit Ergebnissen von Leistungstests

Leistungen messen und beurteilen

KANTON AARGAU
Departement
Bildung,
Kultur und Sport

KBL • CEA

Vorwort

Leistungen messen und beurteilen ist seit je ein zentrales Thema der Schule. Was wissen und können Schülerinnen und Schüler? Wo liegen ihre Stärken, wo ihre Schwächen? Haben sie die Lernziele erreicht? Welche Potenziale sind vorhanden, welche Defizite bestehen? Diese Fragen müssen immer wieder gestellt werden und sie erfordern klare Antworten. Denn sie betreffen die Schulqualität und entscheiden über Zukunftschancen.

Traditionellerweise wurden diese Fragen mit Hilfe von Klassenprüfungen, Noten und Zeugnissen beantwortet. Im Laufe der Zeit kamen weitere Instrumente wie Lernjournals, Beobachtungsbogen, Wortberichte und viele andere hinzu. In den letzten Jahren fanden verstärkt auch Leistungstests Eingang in die Schule. Leistungstests gab es zwar schon lange vor PISA; so wurden beispielsweise in der Schweiz bereits Mitte des 19. Jahrhunderts die Schreib- und Rechenkenntnisse der jungen Männer mit Rekrutenprüfungen erhoben, um die nachhaltige Leistung der Schule festzustellen. Eher neu ist aber, dass Leistungstests zur direkten Unterstützung von Lern- und Lehrprozessen eingesetzt werden: Sie liefern eine Grundlage, welche die Lehrpersonen zur Förderung ihrer Schülerinnen und Schüler sowie zur Weiterentwicklung ihres Unterrichts nutzen können. Dabei ersetzen die Tests und ihre Ergebnisse die unterrichtsinterne Leistungsmessung und -beurteilung nicht. Vielmehr stellen sie eine Ergänzung dar und eröffnen neue Perspektiven. Dieser Ansatz ist wegweisend.

Der Umgang mit Ergebnissen von Leistungstests ist allerdings nicht selbstverständlich. Mit der vorliegenden Handreichung wird deshalb eine Hilfestellung dazu geboten, wie von der externen Leistungsmessung und -beurteilung zur Optimierung der Unterrichtspraxis zu kommen ist. Die Handreichung basiert auf Erfahrungen mit dem freiwilligen Leistungstest Check 5 im Kanton Aargau. Sie ist aber nicht auf einen bestimmten Test oder auf eine bestimmte Schulstufe ausgerichtet, sondern beinhaltet Informationen, die für Lehrpersonen aller Schulstufen nützlich sein können. Ich wünsche Ihnen eine gewinnbringende Verwendung der Handreichung in der Schulpraxis!

Landammann Rainer Huber
Vorsteher des Departements Bildung, Kultur und Sport

Inhalt

Einleitung **4**

Leistungen messen und beurteilen **5**

Funktionen der Leistungsmessung 5
Kriterien eines guten Leistungstests 6
Grenzen der Leistungsmessung 6

Fünf Schritte zur Verbesserung der Unterrichtsqualität **8**

1. Schritt: Leistungen bestimmen und Tests entwickeln 8
2. Schritt: Leistungen messen und beurteilen 8
3. Schritt: Testergebnisse analysieren und interpretieren 9
4. Schritt: Ziele setzen und Massnahmen ergreifen 9
5. Schritt: Massnahmen umsetzen und deren Wirkung überprüfen 11

Testergebnisse analysieren und interpretieren **12**

Testergebnisse auf verschiedenen Ebenen des Bildungssystems 12
Standortbestimmung für die Schülerinnen und Schüler 12
Zum Umgang mit Testergebnissen einzelner Schülerinnen und Schüler 14
Standortbestimmung für die Klassen 15
Erfolgreicher Unterricht dank Testergebnissen? 18

Literatur **19**

Herausgeber: Departement Bildung, Kultur und Sport des Kantons Aargau (BKS)
Autor/-in: Urs Moser und Sarah Tresch, Kompetenzzentrum für Bildungsevaluation und Leistungsmessung an der Universität Zürich (KBL)

Einleitung

«Die persönlichen Stärken und Schwächen der Schülerinnen und Schüler von einer externen Seite beurteilen zu lassen, interessiert mich, um eine umfassendere individuelle Förderung anstreben zu können.»

«Check 5 war eine Chance, über den eigenen Unterricht nachzudenken. Für dieses Instrument bin ich sehr dankbar! Solche Angebote sind meines Erachtens dann von grossem Nutzen, wenn die Teilnahme auf freiwilliger Basis beruht!»

Lehrpersonen befinden sich in einem kontinuierlichen Interaktionsprozess mit ihren Schülerinnen und Schülern: Sie unterrichten, fördern, beobachten, prüfen und beurteilen. Dieser Prozess kann von einer Beurteilung der Leistungen der Schülerinnen und Schüler von aussen nur profitieren, denn die Kenntnis des Leistungsstandes ist für ein förderorientiertes Handeln im Unterricht unentbehrlich. Es ist deshalb auch für die Mehrheit der Lehrpersonen unbestritten, dass Leistungsmessung ein wirkungsvolles Instrument zur Qualitätssicherung sein kann; vorausgesetzt, die damit verbundenen Ziele und Konsequenzen sind transparent.

Mit Check 5 ermöglicht das Departement Bildung, Kultur und Sport des Kantons Aargau den Lehrpersonen der 5. Klassen, die Leistungen ihrer Schülerinnen und Schüler messen und beurteilen zu lassen. Die Nutzung des Angebots ist freiwillig. Wer das Angebot nutzt, ist allerdings dazu verpflichtet, die Ergebnisse der Schülerinnen und Schüler in den Leistungstests zu analysieren, daraus Lernziele für die Förderung abzuleiten und entsprechende Massnahmen im Unterricht umzusetzen. Check 5 ist erst dann abgeschlossen, wenn die Lehrpersonen die Wirkungen der Massnahmen im Unterricht überprüft und schriftlich festgehalten haben.

Die Leistungsmessung ist bei Check 5 Bestandteil der Qualitätssicherung im Unterricht. Die Lehrpersonen erhalten eine externe Beurteilung der Leistungen ihrer Schülerinnen und Schüler, die sie für die Förderung im Unterricht nutzen sollen. Dieses Vorgehen ist für die meisten Lehrpersonen neu und ungewohnt. Mit der vorliegenden Handreichung «Leistungen messen und beurteilen» erhalten die Lehrpersonen eine Anleitung zum Umgang mit den Ergebnissen von Leistungstests. Es wird aufgezeigt, (1) wozu die Leistungsmessung im Bildungswesen eingesetzt wird, (2) wie die Leistungsmessung zur Optimierung der Unterrichtsqualität genutzt werden kann und (3) worauf bei der Analyse und Interpretation der Ergebnisse von Leistungstests zu achten ist.

Die Handreichung zum Umgang mit den Ergebnissen von Leistungstests basiert auf den Erfahrungen der ersten Durchführung von Check 5 im Schuljahr 2004/2005. Aus dieser Durchführung stammen auch die Eindrücke von Lehrpersonen, die in der Handreichung in Form von Zitaten enthalten sind. Die Stimmen aus der Praxis zeigen, dass die externe Leistungsmessung im Dienste der zuverlässigen Schülerbeurteilung und zur Optimierung der Unterrichtsqualität in Zukunft zu den grundlegenden Instrumenten der Schule gehört.

Leistungen messen und beurteilen

Funktionen der Leistungsmessung

In den letzten Jahren wurde die Schule zunehmend mit Fragen der Evaluation und der Leistungsmessung konfrontiert.[1] Die Ergebnisse der Leistungsmessung werden für unterschiedliche Zwecke genutzt. Es lassen sich drei Funktionen unterscheiden:

– Transparenz und Rechenschaftslegung

– Qualitätssicherung und Qualitätsentwicklung

– Beurteilung und Selektion

Transparenz und Rechenschaftslegung
Die Ergebnisse beim internationalen Schulleistungsvergleich PISA haben für die Schweiz aufgezeigt, dass ein relativ hoher Anteil von Jugendlichen am Ende der obligatorischen Schulzeit nicht über jene Kompetenzen in Lesen und Mathematik verfügt, die für einen reibungslosen Übergang in den Arbeitsmarkt notwendig sind.[2] Dieses beunruhigende Ergebnis hat dazu geführt, dass in Zukunft Transparenz über die Kompetenzen der Schülerinnen und Schüler gefordert wird. Daher hat die Schweizerische Konferenz der kantonalen Erziehungsdirektoren beschlossen, nationale Bildungsstandards zu definieren. Bildungsstandards beschreiben, welche Kompetenzen die Schülerinnen und Schüler erreichen sollten.[3] Mit Leistungstests wird festgestellt, wie gut die Bildungsstandards erreicht werden. Dadurch wird gegenüber der Öffentlichkeit Rechenschaft über den Erfolg des Bildungssystems abgelegt.

«Ich bin der Meinung, dass auf Grund des immer grösseren Drucks auf die Primarschule von Seiten der Eltern und der abnehmenden Schulen Leistungstests notwendig sind.»

Die Beurteilung der Leistungen der Schülerinnen und Schüler im Sinne der Rechenschaftslegung ist auch für die einzelnen Schulen zunehmend von Bedeutung. Mit dem Ziel der Weiterentwicklung der Schul- und Unterrichtsqualität werden im Kanton Aargau die Schulen in ihrer Verantwortung, aber auch in ihrer Rechenschaftspflicht gestärkt. Schulen müssen in Zukunft vermehrt über ihre Qualität Auskunft geben, was die Transparenz über die Kompetenzen ihrer Schülerinnen und Schüler einschliesst. Die Leistungsmessung bildet deshalb einen elementaren Bestandteil des Evaluationsverfahrens von professionell geleiteten Schulen.

Qualitätssicherung und Qualitätsentwicklung
Das Kerngeschäft der Schule ist der Unterricht. Welche Ergebnisse die Schülerinnen und Schüler bei einer Leistungsmessung erreichen, hängt auch zu einem grossen Teil vom Handeln der Lehrpersonen im Unterricht ab. Eine wichtige Massnahme zur Qualitätssicherung und Qualitätsentwicklung besteht deshalb darin, den Unterricht vermehrt auf seine Wirkung hin zu überprüfen und das Handeln im Unterricht zu reflektieren. Die Teilnahme an einer Leistungsmessung führt für die Lehrpersonen zu einer unabhängigen Standortbestimmung ihrer Schülerinnen und Schüler beziehungsweise ihrer Klasse und liefert dadurch die Grundlage für die Reflexion über das eigene Handeln im Unterricht.

«Ich habe ein grundsätzliches Interesse an Möglichkeiten zur Verbesserung der Schul- und Unterrichtsqualität. Ich nehme an Check 5 teil, um meine blinden Flecken zu erkennen, das heisst, die Unterrichtsbereiche zu eruieren, in denen meine Schülerinnen und Schüler gut sind, faktisch aber nicht gut sind.»

Allerdings stehen den Lehrpersonen zurzeit für eine Standortbestimmung noch kaum Leistungstests zur Verfügung, die eine objektive, klassenübergreifende Beurteilung der Leistungen der Schülerinnen und Schüler zulassen. Das Angebot für die Lehrpersonen der 5. Klassen des Kantons Aargau, mit Check 5 ausgewählte Leistungen der Schülerinnen und Schüler messen und beurteilen zu lassen, entspricht daher einem echten Bedürfnis. Check 5 dient insbesondere der Unterrichtsentwicklung und liefert die Grundlage für eine individuelle Förderung der Schülerinnen und Schüler.

Beurteilung und Selektion
Die Beurteilung der Schülerinnen und Schüler und die damit verbundenen Selektionsentscheide gehören zu den kontrovers diskutierten Themen der Schule; vor allem dann, wenn das Thema auf Grund von Noten diskutiert wird. Im Gegensatz zur Beurteilung im Schulalltag (ob mit Noten oder mit Worten), bei der sich Lehrpersonen in der Regel am Klassenverband orientieren, bieten Leistungstests die Möglichkeit, die eigene Beurteilungspraxis zu überprüfen und auf Grund der Leistungen einer grösseren Vergleichsgruppe allenfalls

«Die Kinder gewöhnen sich an meinen Prüfungsstil, meine Tests sind mit der jeweiligen Unterrichtssituation eng verbunden. Ich fände es sinnvoll, wenn eine Fremdbeurteilung als Ergänzung eingesetzt werden könnte.»

anzupassen. Die Ergebnisse in Leistungstests können zudem Selektionsentscheide unterstützen, allerdings nicht ersetzen. Selektionsentscheide sollten immer unter Einbezug verschiedenster Datenquellen und Erfahrungen zustande kommen.

Noten sagen wenig aus über die Kompetenzen der Schülerinnen und Schüler. Dies hat mittlerweile einen grossen Teil der Unternehmen dazu bewogen, bei der Lehrlingsauswahl Eignungstests einzusetzen, um sich ein objektives Bild über die Kompetenzen der Schülerinnen und Schüler zu verschaffen. Im Kanton Aargau ist geplant, ein Volksschulabschlusszertifikat einzuführen. Teil dieses Volksschulabschlusses bildet eine Beschreibung der Kompetenzen der Schülerinnen und Schüler. Dazu werden Leistungstests eingesetzt.

Kriterien eines guten Leistungstests

Leistungen messen bedeutet, sie mit einem Test zu erfassen. Im Gegensatz zu Prüfungen, wie sie Lehrpersonen im Unterricht regelmässig einsetzen, werden Leistungstests auf der Basis einer Messtheorie entwickelt.[4] Gemäss dieser Theorie müssen Tests verschiedenen Kriterien genügen. Die drei wichtigsten sind im Folgenden kurz erläutert.

– *Tests sollen objektiv messen.* Damit ist gemeint, dass das Testergebnis einzig von der Kompetenz einer Schülerin oder eines Schülers abhängt, nicht aber von der Person, die den Test durchführt oder auswertet. Aus diesem Grund werden Leistungstests nach standardisierten Anleitungen durchgeführt. Beispielsweise wird klar festgehalten, welche Hilfestellungen den Schülerinnen und Schülern beim Ausfüllen der Tests gegeben werden und wie viel Zeit für die Bearbeitung des Tests zur Verfügung steht. Auch die Korrektur der Tests erfolgt nach eindeutigen Kriterien. Vor allem Aufgaben in Form von offenen Fragen oder Aufsätzen stellen hohe Anforderungen an die korrigierende Person. Unter Beizug von klaren Kriterien lassen sich aber auch diese schriftlichen oder mündlichen Beiträge objektiv korrigieren.

– *Tests sollen genau messen.* Damit ist gemeint, dass das Testergebnis möglichst präzis und fehlerlos zustande kommt. Das Testergebnis soll beispielsweise nicht von Ort oder Zeit der Durchführung abhängen. Eine genaue Messung setzt zudem voraus, dass der Test eine genügend grosse Anzahl von Aufgaben enthält. Es wäre unfair, sich auf Grund von sehr wenigen Aufgaben ein Bild über die Kompetenzen einer Schülerin oder eines Schülers zu machen. Die Aufgaben sollten zudem das gesamte Fähigkeitsspektrum umfassen, so dass zuverlässige Aussagen über die Kompetenzen von besonders guten sowie von eher schwachen Schülerinnen und Schülern möglich werden.

– *Tests sollen valide messen.* Damit ist gemeint, dass der Test genau das misst, was er zu messen vorgibt. Die einzelnen Aufgaben des Tests müssen der zu messenden Kompetenz entsprechen und diese auch zuverlässig repräsentieren. Für Check 5 bedeutet dies beispielsweise, dass die Aufgaben dem Lehrplan und den Lehrmitteln des Kantons Aargau entsprechen müssen.

Die Testergebnisse der Schülerinnen und Schüler können nur dann als verlässliche Information interpretiert werden, wenn mit der Leistungsmessung die erwähnten Kriterien eingehalten werden. Weil Testergebnisse – wie auch die Ergebnisse anderer Beurteilungsinstrumente – immer mit einem Messfehler behaftet sind, ist es von Vorteil, die Beurteilung der Schülerinnen und Schüler auf der Basis von verschiedenen Datenquellen vorzunehmen. Ein einzelnes Ergebnis darf nie abschliessend interpretiert werden. Weisen allerdings mehrere Ergebnisse in die gleiche Richtung, dann wird eine zuverlässige Interpretation möglich.

Grenzen der Leistungsmessung

Obwohl das Interesse der Lehrpersonen an Check 5 beweist, dass die Nutzung von Leistungstests zur Verbesserung der Unterrichtsqualität einem Bedürfnis entspricht, werden immer wieder auch Vorbehalte gegenüber Leistungstests angebracht:

– «Schulen verfolgen verschiedene Ziele, die mit Leistungstests nie angemessen berücksichtigt werden können.»

– «Leistungstests orientieren sich an einfach messbarem (trägem) Wissen, höhere Verstehensleistungen können nicht erfasst werden, weshalb diese im Unterricht vernachlässigt werden.»

– «Leistungstests führen zu einer Verschlechterung der Unterrichtsqualität, weil Lehrpersonen ihren Unterricht nach den Tests ausrichten (‹teaching to the test›[4]).»

– «Leistungstests führen bei den Schülerinnen und Schülern zu unerwünschten Nebenwirkungen wie Stress und Angst; die Arbeitsatmosphäre verschlechtert sich.»

– «Der Vergleich von Ergebnissen in Leistungstests zwischen Klassen ist nicht fair, weil sich die unterschiedlichen Lernvoraussetzungen nie zu hundert Prozent ausgleichen lassen.»

Sämtliche Vorbehalte sind zum Teil berechtigt und ernst zu nehmen. Es ist beispielsweise unbestritten, dass sich die Leistungsmessung nicht nur auf einfach messbare Leistungen beschränken darf. Die in der Schule vermittelte Leistung ist ohne Zweifel mehr als das, was gemessen werden kann. Dass auch schlecht messbare Leistungen von Bedeutung sind, versteht sich von selbst. Auf die Messung des Messbaren zu verzichten, um das schwer Messbare zu schützen, wäre aber ein pädagogisch unkluges Vorgehen. Denn die Vorbehalte treffen nicht nur für Tests, sondern generell für Prüfungen im Unterricht zu. Tests unterscheiden sich von Prüfungen in der Schule nur insofern, als sie nach bestimmten Regeln entwickelt werden und den erwähnten Kriterien zu genügen haben.

Für einen wirkungsvollen Umgang mit Leistungstests gilt es, den Stellenwert der Leistungsmessung richtig einzuschätzen und die Ergebnisse verantwortungsvoll zu nutzen. Leistungsmessungen führen zu Transparenz über ausgewählte Kompetenzen der Schülerinnen und Schüler. Die Ergebnisse können als Anlass zur Förderung sowie zur Optimierung des Unterrichts genutzt werden. Sie dienen zudem der Orientierung der Schülerinnen und Schüler, der Eltern, der abnehmenden Schulen, der Behörden und der Öffentlichkeit.

«Ich war sehr skeptisch und bin nun vom Ergebnis der Auswertungsart (Tabellen, Grafiken) beeindruckt. Nicht alle Tests sind unnütz.»

«Zuerst war es nicht einfach für die Schülerinnen und Schüler, über ihre eigenen Stärken und Schwächen zu sprechen. Mit der Zeit gelang es aber immer besser.»

Fünf Schritte zur Verbesserung der Unterrichtsqualität

Die Durchführung von Leistungstests gehört auch im Bildungsbereich zum festen Bestandteil eines wirkungsorientierten Qualitätsmanagements. Die Rückmeldung von Ergebnissen aus Leistungsuntersuchungen alleine führt aber kaum zur gewünschten Verbesserung von Schule und Unterricht. Leistungstests müssen deshalb in ein Evaluationsverfahren eingebaut werden, wie dies bei Check 5 der Fall ist. Ein solches Evaluationsverfahren beinhaltet im Wesentlichen fünf Schritte:[5]

– Leistungen bestimmen und Tests entwickeln

– Leistungen messen und beurteilen

– Testergebnisse analysieren und interpretieren

– Ziele setzen und Massnahmen ergreifen

– Massnahmen umsetzen und deren Wirkung überprüfen

Lehrpersonen können bei allen fünf Schritten in irgendeiner Form beteiligt sein. Die Schritte drei, vier und fünf hingegen sind ausschliesslich durch die Lehrpersonen zu vollziehen.

1. Schritt: Leistungen bestimmen und Tests entwickeln

Damit Leistungstests für Lehrpersonen zu handlungsrelevanten Informationen führen, müssen sie einen Bezug zu Lehrplan, Lehrmittel und Unterricht haben. In einem ersten Schritt wird deshalb festgelegt, welche Leistungen geprüft und wie die Leistungen beurteilt werden. Entsprechend dem umfassenden Auftrag der Schule werden in der Regel sowohl fachliche als auch fachübergreifende Leistungen gemessen und dazu Tests entwickelt.

Bei der Entwicklung eines Leistungstests werden mit Vorteil erfahrene Lehrpersonen sowie Fachdidaktikerinnen und Fachdidaktiker beigezogen. Bevor die Aufgaben zu einem definitiven Test zusammengestellt werden, müssen sie zudem bei rund 150 Schülerinnen und Schülern erprobt werden. So kann festgestellt werden, ob die Aufgaben verständlich sind, ob ihr Schwierigkeitsgrad den Erwartungen entspricht, wie viel Zeit die Bearbeitung der Aufgaben in Anspruch nimmt und ob mit den Tests die Leistungen zuverlässig gemessen werden können. Diese Arbeitsschritte werden von sozialwissenschaftlichen Fachpersonen ausgeführt.

Check 5 setzt fachliche Leistungstests ein, die eine Beurteilung verschiedener Kompetenzen in Deutsch und Mathematik ermöglichen. Mit fachübergreifenden Tests werden das individuelle und das kooperative Problemlösen gemessen; zwei Kompetenzen, die in allen Fächern von Bedeutung sein können. Zudem wird mit einem Fragebogen erfasst, wie die Schülerinnen und Schüler ihre Fähigkeiten im selbstregulierten Lernen einschätzen.

2. Schritt: Leistungen messen und beurteilen

Die Leistungen der Schülerinnen und Schüler werden vorwiegend mit schriftlichen Tests gemessen. Der Grund dafür ist einfach: Schriftliche Tests lassen sich im Klassenverband durchführen und beanspruchen relativ wenig Unterrichtszeit. Mündliche Tests hingegen sind aufwändig und verlangen Assessments mit Kleingruppen oder mit einem Kind allein. Zudem ist eine Standardisierung der Durchführung und der Auswertung des Leistungstests meist nur mit grossem Aufwand zu erreichen.

Je nach Ziel der Leistungsmessung werden die Tests in den Schulklassen von externen Personen oder von den Lehrpersonen selbst durchgeführt. Für die Durchführung der PISA-Tests werden beispielsweise geschulte Testleiterinnen und Testleiter eingesetzt. Bei Check 5 führen die Lehrpersonen die Tests selbst durch. Bei der Durchführung der Tests sind in jedem Fall die standardisierten Anleitungen zu befolgen.

Die Ergebnisse in den Leistungstests werden den Lehrpersonen in schriftlicher Form zugestellt. Sie geben einen Überblick über den Leistungsstand der Schülerinnen und Schüler in den ausgewählten Fachbereichen (Standortbestimmung für die Schülerinnen und Schüler). Der Mittelwert der Ergebnisse aller Schülerinnen und Schüler einer Klasse gibt einen Überblick über den Leistungsstand der Klasse (Standortbestimmung für die Klasse).

Die Beurteilung der Ergebnisse setzt Vergleiche voraus. Die Vergleiche können

– sozial ausgerichtet werden, indem die Ergebnisse der Schülerinnen und Schüler beziehungsweise der Klasse mit den Ergebnissen einer relevanten Gruppe verglichen werden;

– sachlich ausgerichtet werden, indem die Ergebnisse mit konkreten Umschreibungen von Kompetenzen (Kompetenzniveaus) verglichen werden;

– individuell ausgerichtet werden, indem durch mehrmaliges Testen der Lernfortschritt der Schülerinnen und Schüler beziehungsweise der Klasse verfolgt wird.

In Zukunft soll in der Schweiz dem sachlich ausgerichteten Vergleich mehr Beachtung geschenkt werden, wozu Bildungsstandards entwickelt werden. Die Leistungen der Schülerinnen und Schüler sollen mit Mindeststandards verglichen werden. Ziel ist es, dass die Mindeststandards von allen Schülerinnen und Schülern der Volksschule erreicht werden.

«Ich möchte zusätzlich zum schulhausinternen Vergleich auch einen Einblick gewinnen, wie meine Klasse im kantonalen Vergleich dasteht.»

3. Schritt: Testergebnisse analysieren und interpretieren

Der Umgang mit Testergebnissen verlangt eine professionelle Berufsauffassung, verbunden mit einer diagnostischen Neugier. Von der Lehrperson wird erwartet, dass sie sich auch bei unerfreulichen Nachrichten mit der notwendigen Offenheit und Objektivität darum bemüht, die Testergebnisse zu erklären. Verantwortungsbewusstes Handeln im Dienste jedes einzelnen Kindes bedeutet, dass das Analysieren der Testergebnisse in Kenntnis sämtlicher relevanten Informationen erfolgt. Dazu gehören zum einen gesichertes Wissen über die Wirksamkeit von Unterricht[6], zum andern aber auch praktische Erfahrungen und die genaue Kenntnis der Lehr-Lern-Bedingungen im Unterricht:

Welche Ziele wurden verfolgt? Welche Kompetenzen wurden erwartet? Welche Lehr-Lern-Formen wurden eingesetzt? Welche Schwierigkeiten sind aufgetreten? Wo liegen meine Stärken, wo liegen meine Schwächen? Welche Bedeutung hat die soziale Zusammensetzung der Klasse für die Ergebnisse in den Leistungstests? Die Testergebnisse zeigen keinen direkten Weg zur Verbesserung auf, sondern bieten einen Anlass für Fragen über mögliche Optimierungsprozesse.[7] Es ist die Aufgabe der Lehrperson, Stärken und Schwächen der Schülerinnen und Schüler zu erkennen.

Ob die Interpretation der Ergebnisse zuverlässig erfolgt, hängt auch von der Offenheit der Lehrperson ab. Wer sich für ein unerfreuliches Ergebnis nicht oder nur am Rande verantwortlich fühlt, braucht sein eigenes Handeln auch nicht weiter zu reflektieren, weil die Gründe für die Testergebnisse dann ausserhalb des eigenen Einflussbereichs liegen. Damit Testergebnisse für die Unterrichtsentwicklung genutzt werden können, wird eine kritische Reflexion über das Zustandekommen der Ergebnisse vorausgesetzt.

Die pädagogische Empfehlung, die Ursachen für die Testergebnisse auch bei sich selbst zu suchen, schliesst natürlich nicht aus, dass bei realistischer Sichtweise auch ausserhalb der eigenen Handlungsmöglichkeiten nach Ursachen gesucht wird.

Die Analyse und Interpretation der Testergebnisse ist ein wichtiger Schritt zur Optimierung der Unterrichtsqualität. Dabei kann sich die Diskussion im Schulhausteam, aber auch mit aussen stehenden Fachpersonen als sehr hilfreich erweisen.

«Ich bin lange über den Auswertungstabellen und den Abbildungen gesessen und habe viel herauslesen können. In den meisten Fällen wurde ich bestätigt, es gab aber auch zwei, drei Überraschungen.»

4. Schritt: Ziele setzen und Massnahmen ergreifen

Im Anschluss an eine differenzierte Analyse und Interpretation der Testergebnisse erfolgt die Reflexion über das Handeln im Unterricht. Ziel ist es, Möglichkeiten zur Optimierung des Lehr-Lern-Prozesses festzulegen. Lassen sich aus den Testergebnissen fachliche Förderbereiche für die ganze Klasse oder für einzelne Schülerinnen und Schüler ableiten? Welche Lernziele sollen angestrebt werden? Mit welchen Massnahmen im Unterricht können die Ziele erreicht werden?

«Ich habe mir Gedanken zu meinem Deutschunterricht gemacht und werde ihn nun neu strukturieren und ich werde die Ziele und die Zielerreichung besser den Kindern kommunizieren.»

Auf der Grundlage der Ergebnisse im Check 5 hat rund ein Drittel der Lehrpersonen ein oder zwei konkrete Massnahmen abgeleitet, ein weiteres Drittel hat drei Massnahmen und das letzte Drittel vier oder mehr Massnahmen ergriffen.

Auch bei diesem Schritt ist es von Vorteil, sich mit Kolleginnen und Kollegen auszutauschen oder andere Fachpersonen beizuziehen. Die Reflexion über Unterricht im Team erweitert die Optik im Hinblick auf konkrete Ziele und Massnahmen.

«Das Sachrechnen habe ich zu wenig beachtet. Deshalb sind die Unterschiede innerhalb der Klasse (Streuung) besonders gross. Das Sachrechnen muss geübt werden und die darin enthaltenen mathematischen Probleme müssen für die Schülerinnen und Schüler spielerisch entdeckt werden können!»

Strategisches Vorgehen beim Lösen von Sachrechenaufgaben

Eine Lehrperson hatte mit ihrer Klasse in Check 5 im Sachrechnen unterdurchschnittlich abgeschnitten. Die Lehrperson setzte sich deshalb zum Ziel, dass ihre Schülerinnen und Schüler beim Lösen einer Sachrechenaufgabe in Zukunft strategisch vorgehen. Die Schülerinnen und Schüler sollen vermehrt darauf achten, das in der Aufgabe enthaltene Problem zu erkennen und zu formulieren. Erst danach soll die Aufgabe gelöst werden. Das Ziel soll erreicht werden, indem das strategische Vorgehen beim Lösen von Sachrechenaufgaben im Unterricht geübt wird und verschiedene Strategien miteinander verglichen werden.

Die Lehrperson formulierte drei Zielsetzungen:

– Die Schülerinnen und Schüler gewinnen dank strategischem Vorgehen beim Lösen einer Sachrechenaufgabe an Sicherheit.
– Die Schülerinnen und Schüler wenden beim Lösen einer Sachrechenaufgabe bewusst eine Strategie an.
– Die Schülerinnen und Schüler sind fähig, das in der Aufgabe enthaltene Problem zu erkennen und zu formulieren.

Zur Zielerreichung wurden verschiedene Massnahmen im Unterricht ergriffen:

– Die Schülerinnen und Schüler lernen verschiedene Strategien zur Lösung von Sachrechenaufgaben kennen und vergleichen diese miteinander.
– Die Schülerinnen und Schüler entwickeln zur Lösung von Sachrechenaufgaben gemeinsam Strategien und wenden diese bewusst beim Üben an.
– Die Schülerinnen und Schüler erfinden eigene Sachrechenaufgaben.

Im Mathematikunterricht wurde das strategische Vorgehen beim Lösen von Sachrechenaufgaben ein Semester lang wöchentlich während einer Lektion geübt.

«Die Ergebnisse in der Orthographie liegen unter dem Mittelwert aller Klassen. Ich habe viele fremdsprachige Kinder in meiner Klasse. Doch auch das mangelnde Sprachtraining hat zu diesem Ergebnis geführt.»

Alltagsbegriffe korrekt schreiben

Eine Lehrperson hatte mit ihrer Klasse im Vergleich zu den anderen Klassen in Deutsch, insbesondere im Teilbereich Orthographie, deutlich schlechter abgeschnitten. Die Ergebnisse der einzelnen Schülerinnen und Schüler zeigen ihr, dass vor allem fremdsprachige Kinder Mühe hatten, einfache Alltagsbegriffe richtig zu verstehen und zu schreiben. Die Lehrperson führte dieses Ergebnis aber auch auf den Unterricht zurück, den sie zu wenig auf die in Check 5 geprüften Kompetenzen ausgerichtet hatte.

Die Lehrperson formulierte folgende Zielsetzung:

– Die Schülerinnen und Schüler verbessern sich in der Orthographie. Sie lernen Begriffe aus dem Alltag verstehen und richtig schreiben.

«Gut war, dass ich die Massnahmen vorher formulieren musste, sonst hätte ich mich vielleicht nicht daran gehalten.»

Zur Zielerreichung führte sie ein systematisches Wortschatztraining mit folgenden Übungsteilen durch:

– Die Schülerinnen und Schüler lesen vermehrt kurze Texte aus dem Alltag sowie Zeitungstexte und schreiben anschliessend die unbekannten oder schwierigen Begriffe heraus.
– Die Schülerinnen und Schüler suchen für die Alltagsbegriffe mit Hilfe des Dudens Synonyme.
– Mundartbegriffe aus dem Alltag werden ins Hochdeutsche übersetzt und schriftlich festgehalten.
– Jede Woche werden Alltagsbegriffe überprüft.

Das Wortschatztraining wurde in verschiedenen Fächern während acht Monaten durchgeführt.

Konkrete und überprüfbare Ziele helfen, den Lehr-Lern-Prozess effizient zu planen, Massnahmen entsprechend festzulegen und den Lernerfolg nach dem Umsetzen der Massnahmen zu überprüfen. Es hat sich als erfolgreich erwiesen, alle am Lehr-Lern-Prozess Beteiligten über die Ziele zu informieren, vor allem die Schülerinnen und Schüler, aber auch die Eltern. Transparenz sowie schriftliches Festhalten der Ziele und der Massnahmen erhöhen die Verbindlichkeit und dadurch die Wahrscheinlichkeit, dass die Optimierung des Unterrichts eintrifft.

«In Form von schriftlichen Verträgen mit den Schülerinnen und Schülern hielt ich die Ziele und Massnahmen fest. Wir schauten sie alle zwei Wochen wieder an.»

5. Schritt: Massnahmen umsetzen und deren Wirkung überprüfen

Von der Umsetzung konkreter Massnahmen im Unterricht wird eine Verbesserung der Unterrichtsqualität erwartet, die sich in den Leistungen der Schülerinnen und Schüler niederschlägt. Um Gewissheit darüber zu erhalten, ob diese Verbesserung eingetroffen ist, müssen die Leistungen der Schülerinnen und Schüler erneut gemessen und beurteilt werden. Um die Wirkung der Massnahmen für den individuellen Lernfortschritt zu überprüfen, kann der gleiche Test noch einmal eingesetzt werden. Steht kein Test zur Verfügung, dann werden dazu die üblichen Beobachtungs- und Beurteilungsinstrumente eingesetzt.

Strategisches Vorgehen beim Lösen von Sachrechenaufgaben

Die Lehrperson, die sich zum Ziel setzte, Strategien für den Umgang mit Sachrechenaufgaben zu vermitteln, forderte die Schülerinnen und Schüler auf, diese Strategien beim Lösen von Aufgaben regelmässig zu verbalisieren. Die Wirkung der Massnahme überprüfte die Lehrperson, indem die Schülerinnen und Schüler bei den Mathematikprüfungen ihre Strategien aufschreiben mussten. Durch den Einsatz von Aufgaben mit vergleichbaren Anforderungen während eines halben Jahres konnte bei nahezu allen Schülerinnen und Schülern eine Leistungssteigerung nachgewiesen werden.

«Die Massnahmen im Bereich Orthographie waren für die Klasse und für mich sehr wichtig. Wir haben seit Check 5 zirka 800 so genannte ‹Lernwörter im Alltag› gelernt. Diese Wörter werden in einem speziellen Verfahren jede Woche ausgewählt und in Form einer Wörterprüfung geschrieben. Alle haben sich in der Rechtschreibung merklich verbessert!»

Alltagsbegriffe korrekt schreiben

Die Lehrperson, die sich zum Ziel setzte, dass ihre Schülerinnen und Schüler Begriffe aus dem Alltag besser verstehen und richtig schreiben, überprüfte den Lernerfolg wöchentlich. Insgesamt wurden 800 neue Wörter festgehalten, deren Schreibweise regelmässig geübt und der Erfolg immer wieder überprüft. Die bewusste Auseinandersetzung mit unverstandenen Wörtern (sichten, nachschlagen, aufschreiben, anwenden, prüfen) führte zu einer objektiv nachweisbaren Verbesserung des Wortschatzes und der Orthographie aller Schülerinnen und Schüler. Die Schülerinnen und Schüler lernten zudem, den Duden zu nutzen und effizient einzusetzen.

Für die Überprüfung der Wirkung der Massnahmen im Unterricht lohnt es sich, wenn die Lernziele für die Klasse, aber auch für die einzelnen Schülerinnen und Schüler schriftlich festgehalten werden. Je konkreter die Lernziele formuliert werden, desto einfacher wird es für die Lehrperson sein, sie zu überprüfen. Auch für die Erfolgskontrolle ist es sinnvoll, mit anderen Lehrpersonen zusammenzuarbeiten. Prüfungen können ausgetauscht oder im Team entwickelt werden. Die Aussensicht lässt sich auch innerhalb einer Schule organisieren und ohne grossen Aufwand erfolgreich nutzen.

Testergebnisse analysieren und interpretieren

Testergebnisse auf verschiedenen Ebenen des Bildungssystems

Leistungstests führen zu Erkenntnissen, die auf verschiedenen Ebenen des Bildungssystems genutzt werden können.

– Zunächst führen sie für die Schülerinnen und Schüler zu einer unabhängigen Beurteilung der Kompetenzen. Diese Standortbestimmung dient den Lehrpersonen zur Ausrichtung der individuellen Förderung.

– Durch den Vergleich der Klassenergebnisse mit den Ergebnissen anderer Klassen erhalten die Lehrpersonen eine Fremdbeurteilung des Leistungsstandes ihrer Klasse. Diese Standortbestimmung dient den Lehrpersonen zur Optimierung des Unterrichts.

– Sofern die Leistungstests in mehreren Klassen einer Schule durchgeführt werden, lassen sich die Ergebnisse auch zwischen den Klassen innerhalb einer Schule vergleichen. Diese Standortbestimmung dient der Schule zur Optimierung der Schulqualität; vor allem dann, wenn die Ergebnisse im Kollegium analysiert werden und der Erfolg der getroffenen Massnahmen gemeinsam überprüft wird.

– Falls die Teilnahme an der Leistungsmessung nicht freiwillig erfolgt, sondern auf Grund der Zugehörigkeit zu einer wissenschaftlich gebildeten Stichprobe, dann können die Ergebnisse auch national oder international sinnvoll verglichen werden. Diese Standortbestimmung kann für Anpassungen des Bildungssystems genutzt werden.

Standortbestimmung für die Schülerinnen und Schüler

Testergebnisse als Anteil richtig gelöster Aufgaben

Bei vielen Leistungstests werden die Testergebnisse in Form von einfachen Kennzahlen wie dem Anteil richtig gelöster Aufgaben in Prozenten dargestellt. Diese Kennzahlen haben den Vorteil, dass sie allgemein verständlich sind, auch für die Schülerinnen und Schüler. Der Nachteil liegt darin, dass nicht unmittelbar einsichtig ist, ob beispielsweise 30 Prozent richtig gelöster Aufgaben als gut oder als schlecht zu beurteilen sind. Die Beurteilung der Leistungen auf Grund von einfachen Kennzahlen verlangt einen sozial ausgerichteten Vergleich: Wenn nur ein kleiner Teil der Schülerinnen und Schüler mehr als 30 Prozent der Aufgaben lösen kann, dann wird das Ergebnis eher positiv beurteilt, wenn nahezu alle Schülerinnen und Schüler mehr als 30 Prozent der Aufgaben lösen können, dann wird das Ergebnis eher negativ beurteilt. Die Ergebnisse eines Schülers oder einer Schülerin müssen im Vergleich zu den Ergebnissen der Schülerinnen und Schüler innerhalb der Klasse oder zu den Ergebnissen der Schülerinnen und Schüler aller Klassen analysiert und interpretiert werden.

Testergebnisse als standardisierte Werte

Leistungstests sind meist so konstruiert, dass sich die Verteilung der Testergebnisse einer Normalverteilungskurve annähert. Das bedeutet, dass die mittleren Werte häufig, die extrem hohen oder tiefen Werte selten vorkommen. Die Normalverteilungskurve hat die Form einer Glockenkurve. Ist die Glockenkurve eher flach und breit, dann ist die Streuung der Testergebnisse um den Mittelwert gross. Ist die Glockenkurve eher steil und schmal, dann ist die Streuung der Testergebnisse um den Mittelwert eher gering.

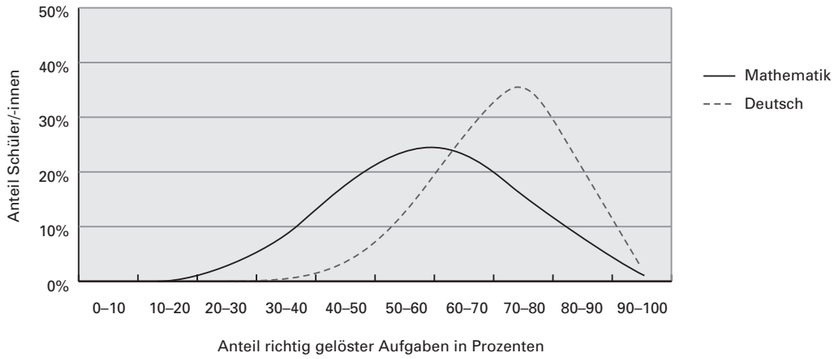

Abbildung 1:
Verteilungskurven der Leistungen in Mathematik und Deutsch

Abbildung 1 zeigt die Verteilung der Fachleistungen der Schülerinnen und Schüler in Check 5. Die Leistungsunterschiede zwischen den Schülerinnen und Schülern sind zu Beginn der 5. Klasse gross. Während die leistungsstärksten 10 Prozent der Schülerinnen und Schüler zwischen 78 und 97 Prozent der Mathematikaufgaben richtig lösen, werden von den schwächsten 10 Prozent zwischen 26 und 39 Prozent der Aufgaben richtig gelöst. In Deutsch sind die Unterschiede etwas geringer. Die besten 10 Prozent der Schülerinnen und Schüler lösen zwischen 85 und 97 Prozent der Aufgaben richtig, die schwächsten 10 Prozent lösen zwischen 43 und 57 Prozent der Aufgaben richtig.

Die Normalverteilungskurven lassen sich anhand des Mittelwertes und der Standardabweichung beschreiben. Der Mittelwert beziehungsweise das arithmetische Mittel entspricht der Summe aller Testergebnisse dividiert durch die Anzahl Testergebnisse. Im Mathematiktest beträgt der Mittelwert 59 Prozent richtig gelöster Aufgaben, im Deutschtest beträgt er 71 Prozent richtig gelöster Aufgaben. Die Standardabweichung ist ein Mass für die Streuung der Testergebnisse um den Mittelwert und wird aus den Abweichungen der einzelnen Testergebnisse vom Mittelwert berechnet (sie entspricht der Wurzel der Summe der quadrierten Abweichungen). Im Mathematiktest beträgt die Standardabweichung 15 Prozent richtig gelöster Aufgaben, im Deutschtest beträgt sie 11 Prozent richtig gelöster Aufgaben. Dass die Streuung der Testergebnisse in der Mathematik grösser ist als in Deutsch, wird aus der Breite der beiden Kurven ersichtlich.

Zu den Eigenschaften der Normalverteilungskurve gehört, dass rund 68 Prozent der Ergebnisse innerhalb von plus/minus einer Standardabweichung liegen. Rund 95 Prozent der Ergebnisse liegen innerhalb von plus/minus zwei Standardabweichungen und mehr als 99 Prozent der Ergebnisse liegen innerhalb von plus/minus drei Standardabweichungen.

Werden die Ergebnisse in zwei verschiedenen Tests anhand des Anteils richtig gelöster Aufgaben miteinander verglichen, dann sind sie kaum sinnvoll zu interpretieren. Wie aus Abbildung 1 hervorgeht, sind 50 Prozent richtig gelöster Aufgaben im Mathematiktest nicht gleich zu beurteilen wie 50 Prozent richtig gelöster Aufgaben im Deutschtest. Damit die Ergebnisse in verschiedenen Tests verglichen werden können, müssen sie in die standardisierte Normalverteilung umgerechnet werden. Die Standardnormalverteilung wird durch den Mittelwert = 0 und die Standardabweichung = 1 definiert. Mittelwert und Standardabweichungen der standardisierten Normalverteilung lassen sich allerdings beliebig transformieren, wie dies beispielsweise für die Darstellung der Ergebnisse im PISA-Test gemacht wird.

Der Mittelwert beträgt beim PISA-Test 500 Punkte, die Standardabweichung 100 Punkte. Rund 68 Prozent der Testergebnisse liegen zwischen 400 und 600 Punkten, rund 98 Prozent zwischen 300 und 700 Punkten und nahezu alle Testergebnisse zwischen 200 und 800 Punkten. Bei der PISA-Studie im Jahr 2003 erreichte die Schweiz in der Mathematik 527 Punkte, im Lesen 499 Punkte. Im internationalen Vergleich erreichten die Schweizer Jugendlichen folglich in der Mathematik bessere Testergebnisse als im Lesen.

Testergebnisse als Kompetenzbeschreibungen
Den beiden bisher vorgestellten Kennzahlen – Testergebnisse als Anteil richtig gelöster Aufgaben und Testergebnisse in standardisierter Form – ist eines gemeinsam: Sie erlauben nur beschränkt Aussagen darüber, was die Schülerinnen und Schüler wissen und können. Die Beurteilung der Testergebnisse erfolgt primär über den sozialen Vergleich. Es fehlt eine konkrete Umschreibung der Kompetenzen, die einem bestimmten Anteil richtig gelöster Aufgaben oder einem standardisierten Wert entspricht.

Aus diesem Grund werden bei Leistungsmessungen die Testergebnisse je länger, je mehr zusätzlich in Form von Kompetenzniveaus dargestellt. Ein Kompetenzniveau entspricht einer konkreten Umschreibung dessen, was die Schülerinnen und Schüler auf dem entsprechenden Niveau wissen und können. Damit die Testergebnisse in dieser Form dargestellt werden können, muss ein Test entsprechend einem Modell entwickelt werden, das den hierarchischen Aufbau der Kompetenz anhand von Kompetenzniveaus beschreibt. Der Test muss zudem so umfangreich sein, dass die einzelnen Kompetenzniveaus durch eine genügende Anzahl Aufgaben repräsentiert werden. Jede Aufgabe lässt sich einem Niveau zuordnen und erfordert unterschiedliche Kompetenzen.

Aufgabe zu Niveau I:	7800	7852	7904	7956	
Aufgabe zu Niveau II:	–	–	–	–	–
Aufgabe zu Niveau III:	3	5	8	13	21
Aufgabe zu Niveau IV:	1	2	6	24	

Abbildung 2: Mathematikaufgabe nach Kompetenzniveau

Abbildung 2 zeigt Aufgaben aus dem Mathematiktest von Check 5 zum Thema «Reihen fortsetzen». Das Fortsetzen der Zahlen- und Musterfolgen verlangt unterschiedliche Kompetenzen. Die Aufgabe zu Kompetenzniveau I entspricht einer arithmetischen Zahlenfolge. Das Lösen der Aufgabe verlangt die Addition einer Konstante (52). Die Aufgabe zu Niveau II entspricht einer Musterfolge. Die Schwierigkeit besteht im Gegensatz zur einfachen arithmetischen Zahlenfolge darin, dass die ersten acht Elemente als ein Muster erkannt werden müssen. Die Aufgabe zu Kompetenzniveau III entspricht einer Fibonaccifolge. Die Fibonaccifolge ist eine Folge von Zahlen, bei der jede Zahl aus der Summe der zwei vorhergehenden Zahlen besteht (3 + 5 = 8, 5 + 8 = 13, 8 + 13 = 21). Beim Aufgabenbeispiel zu Kompetenzniveau IV handelt es sich um eine komplexere Zahlenfolge. Das Lösen der Aufgabe verlangt die Multiplikation der Zahl mit der Stelle der folgenden Zahl in der Reihe (1 x 2 = 2, 2 x 3 = 6, 6 x 4 = 24).

Auch die Ergebnisse im PISA-Test werden nach Kompetenzniveaus dargestellt. Im Lesen werden beispielsweise fünf Kompetenzniveaus unterschieden. Erreicht ein Schüler Kompetenzniveau I, dann verfügt er bestenfalls über rudimentäre Lesekompetenzen. Das heisst, er kann aus einem Text einzelne Informationen ermitteln, es gelingt ihm aber nur ungenügend, das Lesen für das Lernen zu nutzen. Erreicht eine Schülerin Kompetenzniveau V, dann ist sie in der Lage, anspruchsvolle Texte mit nicht vertrautem Inhalt zu verstehen und im Text die relevanten Informationen zu erkennen.

Die Darstellung der Testergebnisse in Form von Kompetenzniveaus hat den Vorteil, dass sich die Lehrperson eine konkrete Vorstellung darüber machen kann, was die Schülerinnen und Schüler wissen und können und was sie noch erreichen sollten. Das Testergebnis hat einen Bezug zum Lehr-Lern-Prozess im Unterricht und ist für die Förderung der Schülerinnen und Schüler direkt nutzbar.

Zum Umgang mit Testergebnissen einzelner Schülerinnen und Schüler

«Vielleicht eine Ausrede: Meine Schülerinnen und Schüler waren extrem nervös auf den Check 5, allen meinen Beruhigungsversuchen zum Trotz.»

Das Ergebnis einer Schülerin oder eines Schüler in einem Test entspricht einer Momentaufnahme des Leistungsstandes in einem ausgewählten fachlichen oder fachübergreifenden Bereich. Obwohl die Leistungen mit Tests objektiv und zuverlässig gemessen werden können, besteht die Gefahr, dass ein Testergebnis den Leistungsstand einer Schülerin oder eines Schülers aus irgendwelchen Gründen ungenau abbildet. Beispielsweise könnte ein Testergebnis durch die Tagesform beeinflusst werden. Ein einzelnes Testergebnis sollte deshalb nicht überbewertet werden, vor allem dann nicht, wenn es auf Grund von Ergebnissen in anderen Tests oder anderer Erfahrungsquellen als wenig plausibel erscheint.

Der Leistungsstand einer Schülerin oder eines Schülers kann vor allem dann zuverlässig beurteilt werden, wenn Informationen aus mehreren Datenquellen in das Urteil einfliessen. Für die Berechnung des Klassenmittelwertes hingegen fallen Ungenauigkeiten beim Zustandekommen des Testergebnisses einer Schülerin oder eines Schülers weit weniger ins Gewicht, zum Teil gleichen sie sich innerhalb der Klasse sogar aus. Der Klassenmittelwert entspricht deshalb einer relativ zuverlässigen Kennzahl.

Standortbestimmung für die Klassen

Testergebnisse als Klassenmittelwerte
Die einfachste Information über den Leistungsstand der Klasse ergibt sich aus dem Mittelwert beziehungsweise aus der Summe der Testergebnisse durch die Anzahl Schülerinnen und Schüler der Klasse (Klassenmittelwert). Die Darstellung des Mittelwerts der eigenen Klasse im Vergleich zu anderen Klassenmittelwerten sowie im Vergleich zum Mittelwert aller am Test beteiligten Klassen (Gesamtmittelwert) führt zu einer ersten Standortbestimmung und einem allgemeinen Eindruck über die durchschnittliche Leistungsfähigkeit der Schülerinnen und Schüler.

«Die Klasse ist ehrgeizig und wollte beim Test gut abschneiden. Sie merkte, dass beim kooperativen Problemlösen von Check 5 auch Nachgeben nötig war. Eine Woche vorher hatte nämlich alles ganz anders ausgesehen: In Zweier- und Dreiergruppen sollten sie Schoggitaler verkaufen. Dazu mussten sie vier bis sechs gemeinsame Verkaufsstunden finden und keines der Kinder wollte Kompromisse eingehen.»

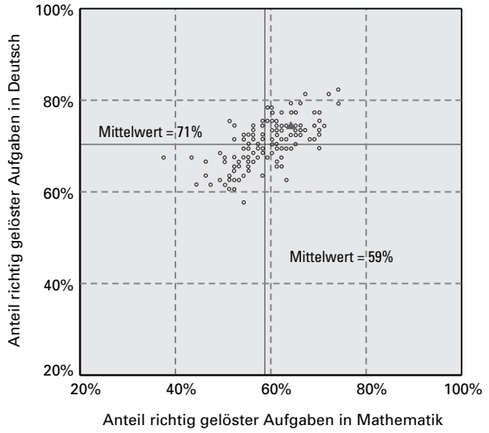

Abbildung 3:
Klassenmittelwerte in den fachlichen Leistungstests von Check 5

Abbildung 3 zeigt die Klassenmittelwerte aller beteiligten Klassen. Das schwarze Dreieck entspricht dem Mittelwert jener Klasse, für die eine Klassenrückmeldung mit den Testergebnissen hergestellt wurde. Die Position des Dreiecks beziehungsweise der Punkte ergibt sich aus den Prozentanteilen richtig gelöster Aufgaben in Mathematik und Deutsch pro Klasse (Klassenmittelwerte in Mathematik und Deutsch). Der Gesamtmittelwert beträgt in der Mathematik 59 Prozent richtig gelöster Aufgaben, in Deutsch 71 Prozent.

Die Abweichung eines Klassenmittelwertes vom Gesamtmittelwert lässt sich – entsprechend den in der Wissenschaft verwendeten Konventionen – als gering, mittel oder gross beurteilen. Die Position des schwarzen Dreieckes in Abbildung 3 zeigt, dass der Klassenmittelwert sowohl in der Mathematik als auch in Deutsch über dem Mittelwert liegt (mittlerer Unterschied im Vergleich zum Gesamtmittelwert). Für die Lehrperson der Klasse, die in der Abbildung 3 durch das Dreieck repräsentiert wird, bedeutet dies vorerst, dass sie die Ziele des Lehrplans im Vergleich zu anderen Klassen eher besser erreicht, dass es aber auch Klassen gibt, deren Ergebnis sowohl in der Mathematik als auch in Deutsch mehr als 10 Prozent höher liegt.

Bei Check 5 waren es 7 Prozent der Klassen, bei denen die Abweichung des Klassenmittelwertes vom Gesamtmittelwert als gross bezeichnet werden konnte, bei rund 23 Prozent konnte sie als mittel, bei rund 70 Prozent als gering bezeichnet werden.

«Da ich meine Klasse schwächer eingeschätzt habe, bin ich erleichtert über die Ergebnisse meiner Klasse im Vergleich mit den anderen Klassen. Als Lehrperson, die zum ersten Mal eine 5. Klasse unterrichtet, fühle ich mich in meiner Arbeit bestärkt.»

Weil die Testergebnisse in einem Leistungsbereich meist zusätzlich nach Fachbereichen aufgeschlüsselt werden können, wird ersichtlich, ob die allgemeine Beurteilung für alle Fachbereiche Gültigkeit hat oder ob sich die Ergebnisse nach Fachbereich unterscheiden. Bei Check 5 wurden in der Mathematik die Ergebnisse für die Fachbereiche «Arithmetik», «Problemlösen» und «Sachrechnen» aufgeschlüsselt. Der Deutschtest führte zu Ergebnissen für die Fachbereiche «Textverständnis», «Texte verfassen», «Sprachbetrachtung» und «Orthographie». Weichen die Ergebnisse zwischen den Fachbereichen deutlich voneinander ab, dann gibt dies für die Lehrperson einen Anhaltspunkt für die fachliche Ausrichtung im Unterricht. Die Ursache für das vergleichsweise schwache Abschneiden in einzelnen Fachbereichen könnte darin liegen, dass bestimmte Fachbereiche im Unterricht vernachlässigt oder ungenügend vermittelt wurden. Die Ursache für das vergleichsweise gute Abschneiden in einzelnen Fachbereichen könnte darin liegen, dass die didaktische Vermittlung besonders geschickt und dadurch das Interesse der Schülerinnen und Schüler im Unterricht besonders hoch war.

Ein vergleichsweise hoher Klassenmittelwert zeigt, dass die Klasse die Ziele des Lehrplans in Bezug auf die gemessenen Leistungen vergleichsweise gut erreicht hat. Ein vergleichsweise tiefer Klassenmittelwert zeigt, dass die Ziele des Lehrplans vergleichsweise schlecht erreicht wurden und Anstrengungen unternommen werden müssen, damit sich die Schülerinnen und Schüler verbessern können. Klassenmittelwerte führen für die Lehrpersonen somit zu einer Orientierung darüber, wie gut ihre Klasse im Vergleich zu anderen ist und in welchen Fachbereichen allenfalls Anpassungen notwendig sind, die alle Schülerinnen und Schüler betreffen. Worüber der Klassenmittelwert kaum Auskunft gibt, ist hingegen die Qualität des Unterrichts.

Der Klassenmittelwert kann nicht einfach als das Ergebnis des Lehr-Lern-Prozesses im Unterricht betrachtet werden. Leistungsunterschiede zwischen Schulklassen lassen sich zu einem grossen Teil auch durch die soziale und kulturelle Herkunft der Kinder erklären. In der Schweiz unterscheiden sich die Einzugsgebiete der Schulen zum Teil beträchtlich in Bezug auf die soziale und kulturelle Herkunft der Schülerinnen und Schüler. Die Lernvoraussetzungen sind deshalb nicht in jeder Klasse gleich, was sich nicht nur im Unterricht, sondern auch in den Testergebnissen der Klassen niederschlägt.

In Klassen mit Schülerinnen und Schülern aus vorwiegend sozial privilegierten Verhältnissen wird es für die Lehrperson einfacher sein, einen Klassenmittelwert zu erreichen, der über dem Gesamtmittelwert liegt, als in Klassen mit Schülerinnen und Schülern aus vorwiegend sozial benachteiligten Verhältnissen. Aus diesem Grund wird der Vergleich der Klassen auch häufig mit «statistisch kontrollierten Mittelwerten» durchgeführt. Das heisst, die Bedeutung der Lernvoraussetzungen für das Testergebnis wird bei der Berechnung des Klassenmittelwerts berücksichtigt.[8]

Testergebnisse unter Berücksichtigung ungleicher Lernvoraussetzungen
Damit die Testergebnisse unter Berücksichtigung der unterschiedlichen Lernvoraussetzungen dargestellt werden können, wird mit einem rechnerischen Verfahren festgestellt, welche Bedeutung die Lernvoraussetzungen der Schülerinnen und Schüler für das Testergebnis haben. Die Vorteile der privilegierten Lernvoraussetzungen für das Zustandekommen der Leistungen werden genauso wie die Nachteile der ungünstigen Lernvoraussetzungen bestimmt und bei der Berechnung des Klassenmittelwerts in Rechnung gestellt. Das heisst, der Anteil richtig gelöster Aufgaben einer Klasse wird um den Einfluss, den die Lernvoraussetzungen der Klasse auf die Leistungen haben, korrigiert.

«Bei vielen herrscht eine allgemeine Spracharmut. Viele Kinder sind ausländischer Herkunft. Sie können auch die eigene Sprache nicht.»

Durch dieses Verfahren werden die Ergebnisse von Klassen mit privilegierten Lernvoraussetzungen reduziert, jene von Klassen mit ungünstigen Lernvoraussetzungen angehoben. Ergebnisse von Klassen, deren Schülerinnen und Schüler beispielsweise überdurchschnittlich häufig aus sozioökonomisch benachteiligten Verhältnissen stammen, werden um den Einflussfaktor der sozialen Herkunft auf die Schulleistungen «nach oben» korrigiert. Umgekehrt werden Ergebnisse bei Klassen, deren Schülerinnen und Schüler praktisch ausschliesslich aus privilegierten Familien stammen, um den Einflussfaktor der sozialen Herkunft auf die Schulleistungen «nach unten» korrigiert. Bei Klassen mit durchschnittlichen Lernvoraussetzungen ändern sich durch die Berücksichtigung der Lernvoraussetzungen die Ergebnisse kaum. Die Korrektur der Ergebnisse einer Klasse entspricht in ihrer Grösse immer dem Ausmass des Vorteils beziehungsweise des Nachteils, der durch die Lernvoraussetzungen entsteht.

Ist die Bedeutung der Lernvoraussetzungen für die gemessenen Leistungen gross, dann werden die Testergebnisse mit Vorteil auch unter Berücksichtigung der Lernvoraussetzungen berechnet. Der korrigierte Klassenmittelwert nähert sich jenem Ergebnis, das auf das Lehr-Lern-Angebot und das Handeln im Unterricht zurückzuführen ist. Ein vergleichsweise hoher Klassenmittelwert zeigt, dass der Unterricht unabhängig von den Lernvoraussetzungen der Schülerinnen und Schüler wirksam war. Ein vergleichsweise tiefer Klassenmittelwert kann nicht mehr durch die Herkunft der Schülerinnen und Schüler erklärt werden, sondern durch Merkmale des Unterrichts.

Sowohl der Klassenmittelwert als auch der nach Lernvoraussetzungen korrigierte Klassenmittelwert sind für die Lehrpersonen von Bedeutung. Der Klassenmittelwert informiert über den faktischen Leistungsstand der Klasse. Dieser kann trotz erfolgreichem Unterricht relativ tief sein, weil die Lernvoraussetzungen ungünstig sind. Die Lehrperson sollte dann trotz erfolgreichem Unterricht Massnahmen treffen, damit die Ergebnisse ihrer Klasse besser werden. Der nach Lernvoraussetzungen korrigierte Klassenmittelwert informiert über die Wirkung des Unterrichts. Diese kann trotz hohem Klassenmittelwert relativ gering sein, weil die Lernvoraussetzungen sehr günstig sind. Die Lehrperson sollte dann vor allem ihren Unterricht überdenken.

Streuung der Ergebnisse innerhalb der Klasse
Ähnliche Klassenmittelwerte können auf Grund von sehr unterschiedlichen Ergebnissen der Schülerinnen und Schüler innerhalb der Klasse zustande kommen.

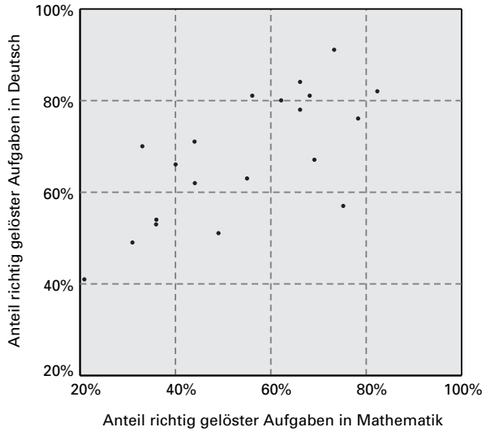

Abbildung 4:
Ergebnisse der Schülerinnen und Schüler in Mathematik und Deutsch bei Check 5

Abbildung 4 zeigt die Ergebnisse der einzelnen Schülerinnen und Schüler einer Klasse in Mathematik und Deutsch. Jeder Punkt steht für eine Schülerin oder für einen Schüler und ergibt sich aus den Prozentanteilen richtig gelöster Aufgaben in Mathematik und Deutsch.

Bei der Klasse handelt es sich um eine in Bezug auf die fachlichen Leistungen heterogene Klasse. Die Testergebnisse streuen innerhalb der Klasse stark. Ein Schüler konnte in der Mathematik nur gerade 20 Prozent der Aufgaben richtig lösen, ein anderer löste rund 90 Prozent der Aufgaben richtig. Für die Lehrperson bedeutet dies, dass sie ihren Unterricht individualisieren und unterschiedlichen Leistungsniveaus anpassen muss. Insgesamt zeigt sich innerhalb der Klasse auch ein relativ enger Zusammenhang zwischen den Mathematik- und den Deutschleistungen. Viele Schülerinnen und Schüler erreichen in beiden Leistungstests entweder eher gute oder eher schlechte Ergebnisse. Es gibt aber auch Schülerinnen und Schüler, für die dieser Sachverhalt nicht zutrifft. Für eine differenzierte Beurteilung der einzelnen Leistungen müssen die Ergebnisse zusätzlich nach den Fachbereichen analysiert werden.

«Es hat in dieser Klasse viele schwache Rechnerinnen, deren Interessen nicht im mathematischen Bereich liegen. Da mein Unterricht dadurch stark auf der Grundlage Fertigkeit und Üben aufbaut, finden dies wohl die starken Rechner oft etwas mühsam. Die Spannbreite der Leistungen ist so breit, dass es schwierig ist, einen interessanten Unterricht für alle zu bieten.»

In eher homogenen Klassen liegen die Ergebnisse der Schülerinnen und Schüler relativ nahe beim Klassenmittelwert (geringe Streuung der Ergebnisse innerhalb der Klasse). In heterogenen Klassen liegen die Ergebnisse der Schülerinnen und Schüler eher weit auseinander (grosse Streuung der Ergebnisse innerhalb der Klasse). Wie ein Mittelwert zustande kommt, ist für die didaktisch-methodische Nutzung der Testergebnisse ebenso relevant wie der Mittelwert selbst. Die Lehrperson erfährt, welche Schülerinnen und Schüler Lücken oder besondere Stärken haben. Die Streuung zeigt auch, ob allenfalls eine stärkere Berücksichtigung individueller Lernvoraussetzungen notwendig wäre. Ein hoher Klassenmittelwert und eine geringe Streuung der Ergebnisse sind für den Unterricht ideal, weil die Lehrperson davon ausgehen kann, dass sie bei einem angepassten Anspruchsniveau allen Kindern der Klasse mehr oder weniger gerecht wird. Ein durchschnittlicher Klassenmittelwert und eine grosse Streuung der Ergebnisse verlangen von der Lehrperson hingegen, dass sie durch einen individualisierenden Unterricht den unterschiedlich leistungsstarken Schülerinnen und Schülern gerecht wird.[9]

Erfolgreicher Unterricht dank Testergebnissen?

Testergebnisse orientieren Lehrpersonen über den Leistungsstand ihrer Schülerinnen und Schüler. Doch wozu können diese Ergebnisse genutzt werden? Wie sollte auf Grund von Mittelwerten und Prozentangaben der Unterricht verbessert werden? Was machen Lehrpersonen, deren Klasse unterdurchschnittliche Leistungen erreicht? Solche Fragen lassen sich nicht allgemein beantworten, sondern nur in Kenntnis des Kontextes, in dem Unterricht stattfindet. Testergebnisse führen auch nicht zu neuen Anforderungen für die Lehrpersonen. Sie bieten ihnen einzig eine Unterstützung in ihrem verantwortungsbewussten Handeln im Dienste der Schülerinnen und Schüler. Denn Leistungen zu messen und zu beurteilen, Schülerinnen und Schüler individuell zu fördern und Unterricht zu verbessern sowie über das eigene Handeln zu reflektieren, das alles bildet die Grundlage für die Optimierung der Schul- und Unterrichtsqualität.

Literatur

[1] Oelkers, J. (2003). Wie man Schule entwickelt. Eine bildungspolitische Analyse nach PISA. Weinheim: Beltz.

[2] Organisation for Economic Cooperation and Development (2001). Lernen für das Leben. Erste Ergebnisse der internationalen Schulleistungsstudie PISA 2000. Paris: OECD Publications.
Organisation for Economic Cooperation and Development (2004). Learning for Tomorrow's World. First Results from PISA 2003. Paris: OECD Publications.

[3] http://www.edk.ch/d/EDK/Geschaefte/framesets/mainHarmoS_d.html

[4] Heller, K. A. & Hany, E. A. (2002). Standardisierte Schulleistungsmessung. In F. E. Weinert (Hrsg.), Leistungsmessungen in Schulen (S. 87–101). Weinheim und Basel: Beltz.

[5] Moser, U. & Tresch, S. (2003). Best Practice in der Schule – von erfolgreichen Lehrerinnen und Lehrern lernen. Buchs: Lehrmittelverlag des Kantons Aargau.

[6] Helmke, A. (2003). Unterrichtsqualität erfassen, bewerten, verbessern. Seelze: Kallmeyersche Verlagsbuchhandlung.

[7] Watermann, R. & Stanat, P. (2004). Schulrückmeldungen in PISA 2000. Sozialnorm- und kriteriumsorientierte Rückmeldeverfahren. In B. Kohler & F.-W. Schrader (Hrsg.), Ergebnisrückmeldung und Rezeption. Empirische Pädagogik, 18(1), 40–61.

[8] Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), Psychologie des Unterrichts und der Schule, Enzyklopädie der Psychologie, Serie Pädagogische Psychologie, Band 3, (S. 71–176). Göttingen, Bern, Toronto, Seattle: Hogrefe.

[9] Lienert, G. A. & von Eye, A. (1994). Erziehungswissenschaftliche Statistik. Eine elementare Einführung für pädagogische Berufe. Weinheim und Basel: Beltz.