

Fragen zur Zukunft der Leistungsbeurteilung an den Schulen

(Ein Essay)

Es hat sich viel verändert!

Wenn man die Entwicklungen im Bereich der schulischen Leistungsbeurteilung in den letzten 10 bis 20 Jahren überschaut so lässt sich feststellen, dass sich sehr viel verändert hat. Als besonders spektakulär kann gelten, dass die Lehrerinnen und Lehrer Konkurrenz bekommen haben, indem nun auch andere die Leistungen ihrer Schüler beurteilen – meistens mit Tests, die extern entwickelt, aber an den Schulen durchgeführt werden und deren Ergebnisse dann in die öffentliche Diskussion gelangen. Doch auch an den Schulen selbst hat sich vieles verändert: Dort werden heute nicht mehr nur die Ergebnisse von Klassenarbeiten und die mündliche Beteiligung beurteilt, sondern auch Anderes, wie z. B. Lernprozesse und dies geschieht häufig mit dem Ziel, Anhaltspunkte für das nachfolgende Lernen zu finden – im Sinne individueller Förderung. Es ist absehbar, dass die Veränderungen weiter gehen werden. Grund genug genauer hinzuschauen, welche Trends zu erkennen sind und die Frage zu stellen, wohin sich die Leistungsbeurteilung an den Schulen sinnvoller Weise bewegen sollte. Meiner Ansicht nach gibt es dazu bislang in den Bildungsverwaltungen und auch an den Schulen noch keine oder zu wenig klare Vorstellungen, was sich bald als problematisch erweisen könnte. Mit Beuys gesprochen können wir nämlich Folgendes erwarten: „*Die Zukunft, die wir wollen, muss erfunden werden. Sonst bekommen wir eine, die wir nicht wollen.*“

Einige Trends

Im Folgenden werden die Trends kurz genannt (ausführlicher bei Sacher/Winter 2011; Maier 2015; Winter 2018).

Neue Instrumente

- Es werden – wie gesagt – vermehrt standardisierte *Leistungstests* und angewandt. Nicht nur diejenigen zur großflächigen Untersuchung von Schülerleistungen (PISA, IGLU, VERA). Auch andere Tests zur Überprüfung und Diagnose von Lernverläufen (QUOP) und zur Diagnose sowie Förderung (LleA) werden vermehrt in Schulen eingesetzt.
- Schülerleistungen werden heute in großem Ausmaß *sichtbar gemacht und direkt dokumentiert*. Das geschieht z. B. in Portfolios, Lernjournalen, auf Lernplakaten, über Aufführungen, Fotos und Filme. Sie werden vor anderen präsentiert und von ihnen wahrgenommen z. B. von Mitschülern und Eltern. Das demokratisiert die Leistungsbeurteilung.
- In großem Umfang werden heute an Schulen *Beurteilungsraster* und auch Kompetenzraster eingesetzt. Sie sollen dazu dienen, inhaltlich differenzierter zu urteilen, Beurteilungskriterien transparent zu machen und auch Rückmeldung zu erzeugen. Auch Schüler werden aufgefordert ihre Arbeiten anhand von Rastern selbst zu beurteilen. Viele Lehrer nutzen solche Raster allerdings hauptsächlich als Beurteilungshilfe und zur Begründung von Noten.
- In jüngster Zeit werden an vielen Schulen und in manchen Bundesländern *Lern-Entwicklungs-Gespräche* institutionalisiert. Diese finden meist mit Lehrern, Eltern und Schülern statt. Manchmal geht es dabei eher um die Einschätzung und Mitteilung von Lernständen, manchmal eher darum, offen über Entwicklungen und das Wohlbefinden der Schüler zu sprechen sowie Ziele zu vereinbaren. In den

Grundschulen können die Protokolle solcher Gespräche als Ersatz des Halbjahreszeugnisses dienen.

Neue Gegenstände und Anforderungen

Die Gegenstände der Leistungsbeurteilung wurden stark ausgeweitet. Es geht längst nicht mehr nur um den Lernstand in einzelnen Fächern.

- Es werden auch *Lernprozesse* sowie *Entwicklungen* beurteilt und das Sozial- und Arbeitsverhalten.
- *Kompetenzen*, die als Bildungsziele fungieren, sollen eingeschätzt werden.
- Die Beurteilung wird näher an das Lernen gerückt und soll helfen, die nächsten Schritte zu bestimmen. Sie wird in diesem Sinne *formativ*. Sie ist daher vielfältiger und muss genauer zu verschiedenen Unterrichtsformen passen.
- Die Leistungen, die beurteilt werden, sind *nicht mehr hauptsächlich in Klausursituationen erbracht*, auch wenn diese weiterhin eine entscheidende Rolle bei der Benotung spielen. Es werden vermehrt Portfolios, Präsentationen, Referate u. Ä. beurteilt.
- Es besteht verstärkt der *Anspruch, die Schülerinnen und Schüler individuell zu fördern*. Daher soll die Beurteilung diagnostisch sein, Rückmeldung vermitteln und Lernentscheidungen unterstützen.
- Die *Schülerinnen und Schüler sollen lernen, sich selbst zu beurteilen* und entsprechend ihr Lernhandeln steuern.
- Die Beurteilung wird nicht mehr dominant an den Fehlern festgemacht, sie soll vermehrt *auch die Stärken der Arbeiten beschreiben und die Begabungen* der Schüler in den Blick nehmen.

Neue Akteure

- Mit dem Einsatz von Tests tauchen neue Akteure auf (z. B. IQB), die vor allem daran orientiert sind, Leistungen auf objektive Weise – unter Absehung der konkreten Person – und zeitökonomisch zu erfassen. Sie vertreten das Modell psychologischer Testung.
- Verlage und andere kommerzielle Unternehmen engagieren sich vermehrt bei der schulischen Leistungsbeurteilung. Sie verkaufen Lernmaterial, Lernprogramme und entsprechende Prüf- und Auswertungsmaterialien oder online-Tests.
- Schüler und Eltern sind vermehrt mit der Leistungsbeurteilung befasst und auch selbst gefordert, diese vorzunehmen.

Insgesamt beschreiben diese Punkte die Gemengelage, aus der neue Entwicklungen hervorgehen werden. Wie diese aussehen könnten ist unten skizziert.

Der Anachronismus der Notengebung

Die Noten wurden gegen Ende des 19. Jahrhunderts breit an den Schulen eingeführt, um anhand ihrer die Schüler für eine hierarchisch aufgebaute Gesellschaft sortieren zu können. Mit dieser einstufigen Grundgeste prägen sie bis heute die gesamte Leistungsbeurteilung an den Schulen. Die inhaltlich genaue, diagnostische Untersuchung der Leistungen sowie deren Wahrnehmung und Anerkennung wird dadurch behindert. Inzwischen weiß man sehr gut, dass die Noten nur eine grobe Rangfolge der Leistungsfähigkeit innerhalb der jeweiligen Klasse abbilden, nicht aber geeignet sind, die Leistungen der Lernenden verschiedener Klassen, Schulen oder gar Schulsysteme miteinander zu vergleichen. Dafür sind bislang allein Tests geeignet. Außerhalb der eigenen Institution spielen die Schulnoten daher kaum eine Rolle. Große Unternehmen

benutzen ihre eigenen Assessment-Verfahren, wenn es darum geht Bewerber auszuwählen.

Noten sind so gesehen überholt und sogar schädlich. Als Indikatoren für den Leistungsstand an den Schulen sind sie sowohl ungeeignet, als auch – wegen der Tests – völlig entbehrlich. Als Verfahren der Leistungsrückmeldung an Schüler und Eltern mangelt es ihnen an inhaltlicher Aussagekraft. Sie begünstigen bei den Schülern eher eine Status- und Anpassungsorientierung als eine Lernorientierung. Die Motivation, die sie erzeugen, ist vergleichsweise schlecht. Man kann sich daher nur wundern, dass die Notenvergabe noch immer von den Bildungsadministrationen vorgeschrieben ist und eine herausragende Rolle bei der Leistungsbewertung spielt. Die Stärke der Noten liegt m. E. darin, dass sie sehr leicht zu machen und schwer zu bestreiten sind. Vielleicht auch darin, dass sie den realen Leistungsstand in den Klassen zuverlässig verschleiern. Aber das wird zunehmend schwieriger werden.

Sollte man also künftig vor allem auf standardisierte Tests zurückgreifen, wenn es um die schulische Leistungsbeurteilung geht? Dieser Ansicht bin ich nicht. Aber man könnte von den Bildungsadministrationen verlangen, dass sie die Schulen aus der Pflicht zur Notengebung entlassen, damit dort Zeit und Energie frei wird, um nützliche Verfahrensweisen einzusetzen und weiterzuentwickeln. In den Schulen könnten die Lehrerinnen und Lehrer dann vor allem lerndienliches „classroom-assessment“ kultivieren, während sich die Administrationen um das „large-scale-assessment“ kümmern. Ob es dazu kommt ist allerdings fraglich.

Zwei Szenarien

Im Folgenden möchte ich zwei Szenarien entwerfen und überspitze dabei einige Aspekte, um die Ausrichtung zu verdeutlichen. Ich mache keinen Hehl daraus, dass ich das erste Szenario befürworte, denn es gibt viele gute und empirisch gestützte Argumente, die schulische Leistungsbeurteilung in diese Richtung weiterzuentwickeln (Winter 2018). Das zweite Szenario hat allerdings auch einige Plausibilität, vor allem wenn man in Betracht zieht, welche gewaltigen technischen Veränderungen im kulturellen Leben Platz gegriffen haben und auch dass die gegenwärtige Politik recht bedenkenlos daran arbeitet, die Schulen auf den Weg des sogenannten „digitalen“ Lernens an Bildschirmen zu bringen.

Szenario 1

Die Leistungsbeurteilung findet vermehrt kurzfristig statt. Sie wird also nah an das Lernen herangerückt und lerndienlich. Sie kann als integriertes Element des Unterrichts bezeichnet werden. Ihre Hauptziele sind es jeweils, nützliches Feedback für diesen Schüler und diese Schülerin zu erzeugen und die nächsten Schritte des Lernens ausfindig zu machen. Das Interesse an der Einstufung der Schüler tritt dabei zurück. Mittel- und langfristig werden aber Lernstandseinschätzungen vorgenommen. Hauptsächlich aber geht es um die Erkundung dessen, was die Lernenden bei ihrer Arbeit tun bzw. denken, welche Voraussetzungen (Vorwissen, Präkonzepte, Interessen) sie dabei einbringen und wie sie vorankommen. Es geht darum zu klären, was an der Arbeit angemessen und weiterführend war und was falsch. Dazu braucht man selbstverständlich die Mitarbeit der Lernenden, denn oft wissen nur sie, wie sie gerade gearbeitet und was sie gedacht haben. Die Leistungen werden in verschiedenen Formaten erbracht. Häufig werden Lernaufgaben bearbeitet und gemeinsam besprochen. Aber es gibt auch engmaschig kontrollierte Übungssituationen und größere

Projekte, in denen die Lernenden selbständig sowie interessen­geleitet arbeiten und größere Leistungen erbringen.

Der Unterricht ist so angelegt, dass das Lernen sichtbar wird. Das geschieht auf natürliche Weise vor allem indem die Lernenden sich untereinander und mit der Lehrperson darüber austauschen, wie sie gearbeitet haben und welche Produkte dabei entstanden sind (William/Leahy 2015). Es wird viel über fachliche Fragen und über die Qualität des Erarbeiteten gesprochen. Ein gutes Modell einer solchen integrierten Leistungsbeurteilung ist die Dialogische Didaktik von Urs Ruf und Peter Gallin (2005a/b). Auch das Konzept der Didaktischen Diagnostik arbeitet in dieser Weise (Prenzel 2016).

Die wichtigste diagnostische Aufgabe der Lehrperson ist es in diesem Szenario, sich Einsichten und Auskünfte darüber zu verschaffen, wie die Lernenden die Sache verstehen und mit welchen Konzepten sie arbeiten. Zur Bestimmung der Lernerfolge und des Lernstandes greift sie durchaus auch auf standardisierte Aufgaben oder auch ganze Tests zurück und prüft gesondert. Hauptsächlich wird sie beides aber aus den Arbeiten der Schüler herauslesen und direkt bestätigen oder entsprechend korrigieren. Ansonsten ist sie vor allem als Expertin für das Lernen und die zu vermittelnden Fachgegenstände gefordert.

Die „Logik“ dieser Beurteilung ist eine andere als die der Testdiagnostik. Die Lehrpersonen beurteilen die Schüler nicht nur ausschnitthaft und in standardisierten Situationen, sondern umfassend und begleitend zu den Lern- und Leistungsprozessen. Sie leiten an, beobachten und analysieren zugleich. Zudem machen sie Vorschläge zum weiteren Vorgehen. Sie können und brauchen dabei nicht „objektiv“ sein, müssen aber vielfältige Informationen nutzen (zur Person, zu den Bedingungen, zum Prozess), die sie ja besitzen. Sie sollen ihre Subjektivität bewusst und reflektiert einsetzen und müssen bereit sein, ihre Urteile kommunikativ zu validieren und im Verlauf der Arbeit zu überprüfen und ggf. zu revidieren. Ihr großer Vorteil gegenüber extern geplanten Leistungsüberprüfungen besteht darin, dass die Lehrpersonen genauer hinschauen und ihre gesamten Erfahrungen mit einzelnen Lernenden einbringen können und dass sie daraus unmittelbar Handlungen ableiten können. Das Beurteilungsvorgehen entspricht in etwa dem der Aktionsforschung und kann auch als hermeneutisches Modell bezeichnet werden (Winter 2018, S. 96ff.).

Szenario 2

Die Schüler lernen überwiegend selbständig in Lernstudios an Bildschirmen oder mit VR-Brillen und Kopfhörern. Ihr Lernverhalten wird dabei kontinuierlich von den Lernprogrammen ausgewertet, z. B. die Zeitdauer der Bearbeitung, die Pausen, die Fehler, das Aufrufen von Hilfestellungen bis hin zur Auswertung der Mimik als Indikator für die Stimmung und Motivation der Lernenden („learning analytics“). Kleine Prüfaufgaben werden immer wieder in den Lernprozess eingestreut. Entsprechend werden den Schülern vom Programm direkt angepasste Angebote und Hilfen zugewiesen: Sie erhalten Rückmeldungen über die Richtigkeit, zusätzliche Erklärungen, Aufgaben mit angemessener Schwierigkeit, Wiederholungen zur Übung und Aufmunterungen durch die Computerstimme. Das was hier nur angedeutet wird ist heute tatsächlich längst möglich, auch wenn davon in deutschen Schulen noch wenig realisiert wird (Maier 2015, S. 196ff.). Die Programme, die ich einsehen konnte, waren noch recht simpel und hauptsächlich auf das Abfragen von Wissen sowie die

Bereitstellung von richtigen Antworten sowie zusätzlichen Erläuterungen ausgerichtet. In diesem Feld ist aber viel in Bewegung und die großen Elektronik- und Internetkonzerne treiben diese Entwicklung voran (google-classroom, apple-classroom). Wobei interessanter Weise die Benotung der Schüler völlig unhinterfragt fortgeführt wird.

Diagnostisch betrachtet ist dieses Szenario anscheinend ideal, denn hier kann das Lernen der Schüler in einem Umfang kontrolliert und ausgewertet werden, wie es einem Lehrer nicht annähernd möglich ist. Es wird weitaus mehr geprüft und kontrolliert und die Rückmeldungen können sehr rasch und individualisiert erfolgen. Die Lernpfade und die gesamte Lerngeschichte eines Schülers werden maschinell aufgezeichnet sowie gespeichert und können über lange Zeiträume eingesehen und analysiert werden. Was nebenbei gesagt ein gravierendes Datenschutzproblem darstellt, wenn die Hersteller der Programme darauf Zugriff haben (Lankau 2017). Es soll hier nicht diskutiert werden, ob die Vision einer digitalen Lernfabrik wünschenswert und ob dieses Lernen besonders wirksam ist, wofür es bislang keinerlei Belege gibt. Es ging nur darum, einen Blick auf Möglichkeiten einer computergestützten Diagnostik zu werfen.

Die Leistungsüberprüfung und die Diagnostik werden bei diesem Szenario von Algorithmen gesteuert. Die Rolle der Lehrperson umfasst eher die Beaufsichtigung, technische Unterstützung und Hilfe in Krisenfällen. Mit Aufgaben der Leistungsbeurteilung und Diagnostik ist sie kaum noch befasst. Die Schüler und ihre Eltern sind ausschließlich Empfänger der Beurteilungen.

Unterschiede zwischen den Szenarios

Was in Szenario 2 auf extreme Weise gezeigt wird ist die Tendenz, die Leistungsbeurteilung an schulextern hergestellte Instrumente und Institutionen zu delegieren. Damit wird aber auch die Definitionsmacht über das, was schulische Leistung sein soll, tendenziell ausgelagert. Das passiert bereits gegenwärtig in großem Umfang, indem Verlage zunehmend Lern- und Prüfmaterial entwickeln, aber auch dadurch, dass Schulleistungstests zu einem Modell des schulischen Prüfens werden und die darin enthaltenen Kompetenzen zu Zielen des Unterrichts. Es ist zu befürchten, dass Lehrpersonen dazu neigen werden, sich die aufwändige Arbeit der Leistungsbeurteilung und Diagnose abnehmen zu lassen; was aber dazu führt, dass ihre eigenen diagnostischen Kompetenzen nicht entwickelt werden. Letztlich untergraben sie damit ihre eigene Qualifikation als Experten für Lernen und Entwicklung und können dann tatsächlich durch Programme ersetzt und auch schlechter bezahlt werden.

In Szenario 2 können die Schüler zu Lernsklaven gemacht werden, deren Entwicklung zu einem Spielball von Profitinteressen der Anbieter von Lernsoftware und der Big Data- Konzerne wird. Aber auch wenn es nicht so schlimm kommt bleiben Probleme. Denn die online-Tests können in der Regel nicht die Voraussetzungen des jeweiligen Lernstandes aufdecken und sie bieten meist nur standardisierte Lernhilfen an, also z. B. Material, das sich die Schüler dann selbst erarbeiten sollen. Bei genauem Hinsehen sind die Testungen nicht formativ im Sinne von „assessment for learning“, sondern wiederholte Lernstandserhebungen im Prozess (siehe „QUOP“). Und es ist nicht recht einzusehen, dass Tests gekauft werden müssen, wenn entsprechende diagnostische Materialien für die Hand der Lehrpersonen frei verfügbar bereits existieren (siehe ILeA). Ihre Handhabung und Nutzung erfordert aber Mühe und Lernen seitens der Lehrpersonen.

Es ist interessant, dass sich die beiden Szenarios auf den ersten Blick gar nicht so sehr unterscheiden. In beiden soll die Leistungsbeurteilung formativ und engmaschig mit dem Lernen verflochten werden – es steuern sowie fördern. Beide zielen darauf, nicht nur zu prüfen, sondern auch Lern- oder Unterrichtsentscheidungen damit zu verknüpfen, um optimal angepasste Bedingungen herzustellen. Allerdings gibt es fundamentale Unterschiede. Bezogen auf Szenario 1 lässt sich sagen:

- Die Leistungsbeurteilung erfolgt hier gemeinsam, die Schülerinnen und Schüler sind aktiv beteiligt und werden zu Diagnostikern und „Steuerleuten“ ihrer Arbeit ausgebildet. Sie reflektieren und erzeugen selbst Feedback, anhand dessen sie sich weiterentwickeln. Es wird aktiv nachgedacht, über Qualität gesprochen, erklärt und das Lernen erkundet (Clarke 2014).
- Es werden nicht nur und nicht vor allem Lernverhalten und Lernstände untersucht, sondern auch deren geistige Voraussetzungen (Konzepte) und Bedingungen.
- Die Lehrpersonen gestalten den Unterricht so, dass sie sich vermehrt für ihre Schülerinnen und Schüler interessieren können und gleichzeitig ihre eigenen Fähigkeiten entwickeln.
- Instrumente wie Lerntagebücher und Portfolios spielen eine wichtige Rolle.
- Die Beteiligten erfahren sich als (Wissens-) Bildungsgemeinschaft, die ihre eigenen Bedingungen gestalten kann.
- Die jeweils nächsten Schritte des Lernens können selbst mitbestimmt werden, sie können individuell oder gemeinschaftlich sein, praktisch oder theoretisch. Die Schlussfolgerungen sind ergebnisoffen.
- Die Erfolge werden im sozialen Kontext sichtbar gemacht und die Entwicklungen unterstützt. Alle Beteiligten können füreinander persönlich wichtig werden.
- Diese Art der Leistungsbeurteilung und Pädagogischen Diagnostik kostet nichts außer Zeit und etwas Anstrengung.

Hinter dem Szenario 1 steht die Vorstellung der Bildung zu mündigen Bürgern einer demokratischen Gesellschaft.

Freilich lassen sich auch in Szenario 1 bestimmte Tests und Testergebnisse sinnvoll einsetzen bzw. nutzen. Es gibt also Platz für Varianten zwischen den beiden Szenarios und auch zwischen ihnen und dem gegenwärtigen Stand der Leistungsbeurteilung.

Fragen an die Zukunft der Leistungsbeurteilung

Für die Schulen und die Bildungsadministrationen stellen sich auf dem Hintergrund des Dargestellten folgende Fragen.

- Welche Formen der Erbringung, der Beurteilung, der Bewertung, der Präsentation und der Dokumentation von Leistungen sollen künftig verstärkt an den Schulen eingesetzt werden und welche reduziert oder ausgeschlossen werden?
- Wie kann erreicht werden, dass lerndienliche formative Leistungsbeurteilung einen größeren Stellenwert im Unterricht erhält?
- Wie sollen die Schülerinnen und Schüler zur Beurteilung und Steuerung ihrer Arbeit und deren Ergebnisse befähigt werden?
- Wie soll die diagnostische Kompetenz der Lehrpersonen entwickelt werden?
- Wieweit sollen extern entwickelte Tests eingesetzt und genutzt werden? Welche kommen in Frage und wie häufig werden sie durchgeführt?
- Wie wird sichergestellt, dass keine Daten über Schüler abgezogen und von Dritten genutzt werden?
- Wie können hohe Kosten für externe Tests vermieden werden?

Das sind Fragen, welche die Bildungsadministratoren, die Schulen und auch die einzelnen Lehrpersonen beantworten müssen, wenn sie vermeiden wollen, dass unbedachte Verfahren Platz greifen. Aber auch die Weiterentwicklung eines schüleraktiven, demokratischen und auf vertieftes fachliches Verstehen ausgerichteten Unterrichts wird von den Entwicklungen bei der schulischen Leistungsbeurteilung abhängen.

Literatur:

- Clarke, Shirley (2008): Active Learning through Formative Assessment. London: Hodder.
- Lankau, Ralf (2017): Kein Mensch lernt digital. Weinheim: Beltz
- Maier, Uwe (2015): Leistungsdiagnostik in Schule und Unterricht. Bad Heilbrunn: Klinkhardt.
- Prenzel, Annedore (2016): Didaktische Diagnostik als Element alltäglicher Lehrarbeit – „Formatives Assessment“ im inklusiven Unterricht. In: Amrhein, B.: Diagnostik im Kontext inklusiver Bildung. Bad Heilbrunn: Klinkhardt
- Ruf, Urs; Gallin, Peter (2005a): Dialogisches Lernen in Sprache und Mathematik. Bd. 1 Austausch unter Ungleichen. Seelze: Kallmeyer.
- Ruf, Urs; Gallin, Peter (2005b): Dialogisches Lernen in Sprache und Mathematik. Bd. 2 Spuren legen – Spuren lesen. Seelze: Kallmeyer.
- Sacher, Werner; Winter, Felix (2011): Diagnose und Beurteilung von Schülerleistungen. Baltmannsweiler: Schneider.
- William, Dylan; Leahy, Siobhán (2015): Embedding Formative Assessment. West Palm Beach: Learning Sciences International.
- Winter, Felix (2018²): Lerndialog statt Noten. Neue Formen der Leistungsbeurteilung. Weinheim: Beltz.