
Beurteilen und Fördern mit Fokus auf Sprache – Fachgutachten

Grundlagen und Empfehlungen zum QUIMS-Schwerpunkt C (2019–2022)

Afra Sturm (Pädagogische Hochschule FHNW)

Werner Senn (Pädagogische Hochschule Luzern)

Mit Beiträgen von Susanne Metzger und Tim Sommer (Pädagogische Hochschule FHNW)

Im Auftrag des Programms QUIMS, Volksschulamt der Bildungsdirektion Kanton Zürich



Inhaltsverzeichnis

	Einleitung	1
1	Empfehlungen und Themenfelder	3
2	Zur Bedeutung des Beurteilens	9
3	Grundlagen der Beurteilung	12
	3.1 Zyklische Prozessmodelle zur Diagnostik und Förderung	12
	3.2 Adaptive Lehrkompetenz als diagnostische und didaktische Kompetenzen	14
	3.3 Urteilsakkuratheit und diagnostische Kompetenz von Lehrpersonen	15
4	Beurteilungsarten und ihre Ziele	18
	4.1 Bezugsnormsystem	18
	4.2 Formatives Beurteilen	18
	4.3 Summativ beurteilen	20
	4.4 Prognostisch beurteilen	21
5	Beobachtungs- und Beurteilungsinstrumente	24
	5.1 Gütekriterien von Testinstrumenten	24
	5.2 Gütekriterien von schulischen Beurteilungsinstrumenten	25
	5.3 Aufgaben zur sprachlichen Produktion und Rezeption	26
6	Sprache in allen Fächern fördern und beurteilen	29
	6.1 Mathematik	29
	6.2 Naturwissenschaft	30
	6.3 Allgemeiner Ausblick	33
7	Herausforderungen beim Beobachten und Beurteilen	34
	7.1 Erwartungen von Lehrpersonen	34
	7.2 Stereotypenbedrohung und sprachliches Lernen	36
	7.3 Sprachförderung – bildungssprachlicher Wortschatz	37
	7.4 Korrekatives Feedback	38
8	Wirksame Weiterbildung im Bereich Beobachten und Beurteilen	39
9	Die Perspektive von Lehrpersonen und Schulleitungen	42
	9.1 Die Sicht von Lehrpersonen	43
	9.2 Die Sicht von Schulleitungen	45
	9.3 Fazit	47
10	Anhang	49
11	Literatur	52

Einleitung

Beurteilen und Fördern bedingen sich gegenseitig: Für eine zielführende Förderung braucht es vorgängig wie auch im Verlauf des Lernprozesses entsprechende Beurteilungsprozesse. Eine Beurteilung muss, damit sie eine Wirkung entfalten kann, in den Förderkreislauf eingebunden werden. In diesem Sinne dient Beurteilen nicht nur dazu, Informationen über die Schüler/-innen zu gewinnen, sondern auch zur adaptiven Gestaltung des Unterrichts. Beurteilen als zentrales Element des schulischen Lehrens und Lernens ist aber auch mit verschiedenen Herausforderungen verbunden, die bspw. systematische Wahrnehmungsfehler oder zu tiefe bzw. zu hohe Leistungserwartungen betreffen.

Das vorliegende Fachgutachten behandelt sowohl das formative wie auch das summative und prognostische Beurteilen. Es fokussiert Fragen der Beurteilung im Unterricht, wobei der Blick auf Sprache im Vordergrund steht, gerade auch im Hinblick auf einen sprachbewussten Fachunterricht.

Folgende Aspekte bilden für das Fachgutachten wichtige Leitlinien:

- A Allgemeine theoretische Grundlagen zur Bedeutung des Beurteilens, zur diagnostischen Kompetenz von Lehrpersonen, zu Beurteilungsarten und ihren Zielen sowie zu Verfahren und Instrumenten, wie sie insbesondere mit Blick auf Aufgaben zur sprachlichen Produktion sowie Rezeption eingesetzt werden.
- B Mit dem Lehrplan 21 wird die formative Beurteilung gestärkt; gleichzeitig erfordert ein kompetenzorientierter Zugang ein breites Beurteilungsrepertoire (Handreichung «Kompetenzorientiert beurteilen», Koch et al., 2017, S. 4).
- C Das Programm QUIMS (= Qualität in multikulturellen Schulen) richtet sich an Schulen mit einem überdurchschnittlich hohen Anteil an Schülern und Schülerinnen

(=SuS) aus sozial unterprivilegierten und fremdsprachigen Familien. Es ist damit kein Programm für Ausländer/-innen, wie Mächler (2015, S. 3) betont, sondern erklärtes Ziel von QUIMS ist eine gute Bildung für alle Schüler und Schülerinnen, «ungeachtet ihrer Herkunft und ihrer Lernvoraussetzungen». Die damit verbundenen Herausforderungen bilden einen wichtigen Schwerpunkt des Fachgutachtens. Eine Orientierung bieten auch die für QUIMS formulierten Qualitätskriterien (Bildungsdirektion Kanton Zürich, 2009).

Ziel dieses Fachgutachtens ist es, eine Grundlage für Weiterbildungen, vor allem schulinterne, bereitzustellen. Dabei können nicht alle Befunde, Ansätze, Diagnoseinstrumente etc. dargestellt werden: Gerade mit Blick auf die Leitlinien B und C muss eine entsprechende Auswahl getroffen werden. Das heisst, das vorliegende Fachgutachten erhebt keinen Anspruch auf Vollständigkeit. Wo verfügbar wurden Meta-Analysen sowie Überblicksarbeiten, die möglichst auf aktuellen empirischen Studien basieren, einbezogen. Zusätzlich wurden einzelne Studien ausgewählt, die es erlauben, bestimmte zentrale Aspekte besser zu beleuchten.

Das Vorgehen beinhaltet insbesondere folgende vier Elemente:

- 1) Die Literatur wurde gesichtet und die Erkenntnisse überblicksartig zusammengestellt. Dabei wurden auch verschiedene Handreichungen zum Beurteilen beigezogen (v.a. Bildungsdirektion Kanton Zürich, 2013; Handreichung «Kompetenzorientiert unterrichten», Koch et al., 2017; Handreichung «Kompetenzorientiert beurteilen», Koch, et al., 2017).

Hinweis: Cohens d ist ein Mass der Effektstärke. Die Werte sind wie folgt zu interpretieren: Werte ab .20 gelten als klein, ab .50 als mittel und ab .80 als gross.

2) Am 5. und 7. September 2018 fanden mit mehreren Experten und Expertinnen sowie mit Vertretern und Vertreterinnen der Bildungsdirektion Zürich, dem Zentrum Lesen der Pädagogischen Hochschule FHNW sowie der Pädagogischen Hochschule Luzern ein Hearing zu einer Entwurfsfassung statt (noch ohne Interviews, vgl. Punkt 3).

Am 13. November 2018 fand ein zweites Hearing vor allem mit Vertretern und Vertreterinnen der Bildungsdirektion Zürich statt.

Die Rückmeldungen aus den Hearings wurden jeweils eingearbeitet.

3) Gleichzeitig wurden Leitfäden für Gruppeninterviews mit Lehrpersonen der Zyklen 1–3 sowie für Einzelinterviews mit Schulleitungen erarbeitet und mit der Vertretung des Volksschulamts vorbesprochen (vgl. Anhang 5.2). Die Interviews fanden im Oktober/November 2018 statt. Die kommunikative Validierung der durchgeführten Interviews fand am 29. November 2018 statt.

Brugg und Luzern, 18. Dezember 2018

Afra Sturm und Werner Senn

Zitiervorschlag:

Sturm, A. & Senn, W. (2018). Beurteilen und Fördern mit Fokus auf Sprache – Fachgutachten. Grundlagen und Empfehlungen zum QuIMS-Schwerpunkt C (2019–2022). Brugg: Pädagogische Hochschule FHNW.

Quelle Titelbild: Unsplash, by Tété

1 Empfehlungen und Themenfelder

Die folgende Grafik in Abbildung 1 zeigt im Überblick die Empfehlungen (Kap. 1.1), die auf Basis der gesichteten Literatur sowie auf Basis der durchgeführten Interviews formuliert wurden. Die Empfehlungen lassen sich zudem allgemeinen sowie spezifischeren Pro-

zessen zuordnen. Die Übersicht führt auch die Themenfelder (Kap. 1.2) auf, die für die Schulen angeboten werden sollen. Dabei lassen sich zwischen den allgemeinen und spezifischen Prozessen sowie Themenfeldern vielfältige Bezüge herstellen.

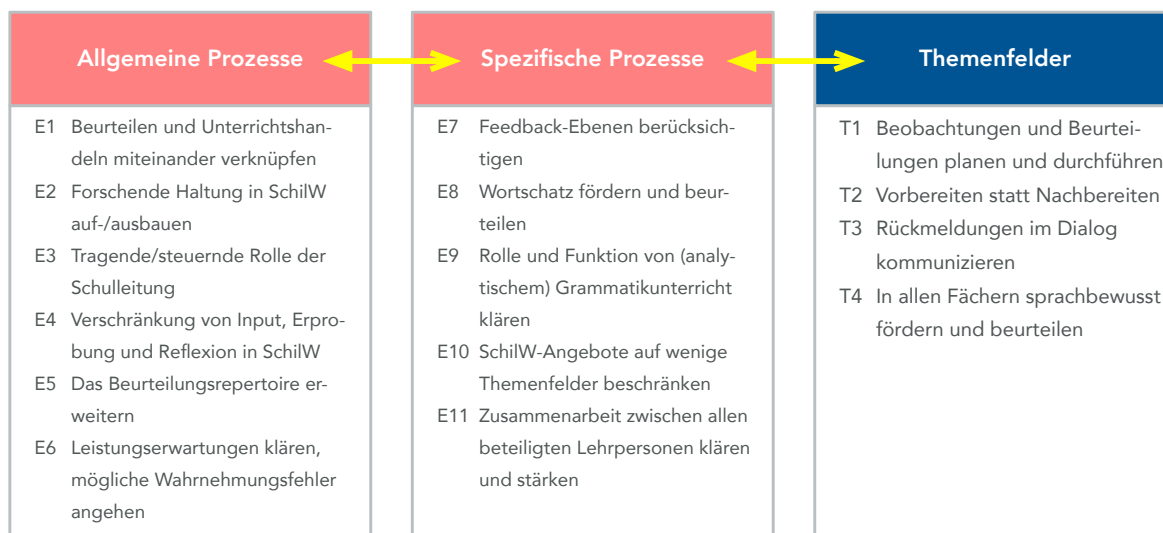


Abbildung 1: Überblick Empfehlungen und Themenfelder

1.1 Empfehlungen

Die Empfehlungen 1–11 lassen sich aus den verschiedenen Kapiteln des Fachgutachtens destillieren. Zu jeder Empfehlung wird angegeben, auf welche Kapitel sie sich hauptsächlich beziehen.

Die BRB-Ziele, die auch im Anhang (S. 51) abgebildet sind, bildeten bei der Formulierung der Empfehlungen eine zentrale Grundlage. Zum einen lassen sich die Ziele verschiedenen Empfehlungen zuordnen – das Herstellen eines gemeinsamen Verständnisses ist hauptsächlich an die Empfehlung E2 und E3 gebunden, kann aber auch in E1 oder E11 gesehen werden –; zum anderen ergeben sich die Ziele erst in der Kombination von Empfehlungen und Themenfeldern (Kap. 1.2). Das betrifft etwa das lernwirksame Formulieren von

Feedback an die Schüler/-innen, das als Zusammenspiel von E5–E7 und T1 (Themenfeld «Beobachtungen und Beurteilungen planen und durchführen») gesehen werden kann.

Nicht zuletzt ist hervorzuheben, dass Chancengerechtigkeit, wie sie im BRB-Ziel 5 angesprochen ist, für die Formulierung der Empfehlungen eine wichtige Rolle spielte. Auch dieser Aspekt findet sich in den verschiedenen Empfehlungen und Themenfeldern auf vielfältige Art und Weise abgebildet: So kann bspw. E8 (Wortschatz) in Kombination mit Themenfeld 4 «In allen Fächern sprachbewusst fördern und beurteilen» als wichtiges Element zur Herstellung von Chancengerechtigkeit gesehen werden. Das gilt aber auch für andere Empfehlungen.

Die Empfehlungen 1–6 beziehen sich zudem auf allgemeine Prozesse einer wirksamen (schulinternen) Weiterbildung. Sie sind damit auch weitgehend unabhängig von einem bestimmten Thema. Dagegen nehmen die Empfehlungen 7–11 spezifische Aspekte und Herausforderungen zum Thema «Beurteilen und Fördern mit Fokus auf Sprache» auf.

E1 Das Beurteilen soll nicht nur zur Gewinnung von Informationen über die Lernprozesse der Schüler/-innen genutzt werden, sondern die individuelle Ebene der Lernprozesse der Schüler/-innen ist mit der Unterrichtsebene zu verknüpfen. Das Beurteilen soll damit für die weitere Planung der Unterrichtsgestaltung genutzt werden können.

→ Kap. 2–4, 9

E2 In den schulinternen Weiterbildungen ist ein gemeinsames Beurteilungsverständnis im Team aufzubauen, das auf einer forschenden Haltung der Lehrpersonen beruht. Dazu sollen im Team Schwerpunkte ausgehandelt werden.

→ Kap. 5, 7 und 8

Die forschende Haltung betrifft verschiedene Aspekte wie das Interpretieren von Beobachtungen, Beurteilungen, das Hinterfragen eigener Lerntheorien, Haltungen gegenüber bestimmten Beurteilungsverfahren u.v.m.

Hinweis: Mit Blick auf Heterogenität, Chancengerechtigkeit ist zu prüfen, inwiefern in schulinternen Weiterbildungen auch Konzepte wie aus SCALA (Projektleitung: M. Neuschwander, PH FHNW) berücksichtigt werden können (vgl. auch Kap. 7.1, S. 34f.).

E3 Für die Durchführung von schulinternen Weiterbildungen ist vorzusetzen bzw. sicherzustellen, dass die Schulleitung eine tragende bzw. steuernde Rolle einnimmt.

→ Kap. 8 und 9

E4 Die schulinternen Weiterbildungen sind nach den Grundsätzen einer wirksamen Weiterbildung gestaltet und sehen insbesondere eine Verschränkung von Input, Erprobung und Reflexion vor.

→ Kap. 8 und 9

Hinweis: Anders als im Schwerpunkt «Schreiben auf allen Schulstufen» (2014–2019) ist die Verschränkung von Input, Erprobung und Reflexion flexibler zu gestalten; zudem ist je nach Schule zu prüfen, ob in einem ersten Schritt mit kleineren Teams etwas erarbeitet wird, bevor es ins ganze Team getragen wird, und ob auch andere Formen wie Coaching etc. einzusetzen wären.

E5 Das Beurteilungsrepertoire ist zu erweitern, u.a. um Verfahren, die an Lernaufgaben anknüpfen und nicht nur situativ erfolgen, sondern auch im Voraus geplant werden. Des Weiteren sollen auch Selbstbeurteilungen der Schüler/-innen verstärkt einbezogen werden.

→ Kap. 2, 5–7

E6 In allen Themenfeldern sind Leistungserwartungen auf Ebene der Schulen wie auch auf Ebene der Lehrpersonen zu klären. Dabei sind auch mögliche Wahrnehmungsfehler anzugehen.

→ Kap. 2 und 7

E7 In allen Themenfeldern sind im Rahmen einer Beurteilung – sei sie formativ oder summativ – die Feedback-Ebenen zu berücksichtigen und mit den Teams / Lehrpersonen zu erarbeiten und zu diskutieren.

→ Kap. 2, 4 und 7

Hinweis 1: Feedback auf persönlicher Ebene wirkt sich nicht positiv auf die Lernleistung der Schüler/-innen aus («Du hast dir keine Mühe gegeben»), da sich daraus keine Handlungsmöglichkeiten ergeben. Gerade mit Blick auf Chancengerechtigkeit sind die Feedback-Ebenen Aufgabe, Prozess und Selbstregulation zu stärken und möglichst auch aufeinander zu beziehen.

Hinweis 2: Der Einbezug der Feedback-Ebenen stärkt den Blick auf die Individual- und Sachnorm, da der individuelle Lernweg fokussiert wird. Ähnlich wie bei der persönlichen Ebene erweist sich der Bezug auf die Sozialnorm nicht förderlich. Wenn die Leistungen eines Schülers oder einer Schülerin mit den Leistungen von anderen verglichen werden («Das hast du nicht so gut wie X gemacht»), lassen sich daraus ebenfalls keine Handlungsmöglichkeiten ableiten.

E8 Wortschatz erweist sich über die Fächer hinweg als einer der wichtigsten Prädiktoren für den Lern- und Schulerfolg. In den schulinternen Weiterbildungen ist den Lehrpersonen/ dem Team in erster Linie zu vermitteln, wie Schul- bzw. Bildungssprache möglichst systematisch und wirksam vermittelt sowie formativ beurteilt werden kann.

→ Kap. 7, 9

E9 Die Rolle und Funktion des (analytischen) Grammatikunterrichts ist zu klären – auch für DaZ-Lernende in einem mehrsprachigen Umfeld –, und zwar in Übereinstimmung mit dem Lehrplan 21.

→ Kap. 7, 9

Hinweis: Erfahrungen aus dem Projekt ALLE haben sehr deutlich gezeigt, dass in der Sprachförderung, vor allem auch in der Schreibförderung die Rolle der Grammatik, insbesondere des analytischen Grammatikunterrichts von Lehrpersonen überschätzt wird (vgl. auch Kap. 9). Hier ist den Lehrpersonen mithilfe des Lehrplans 21 aufzuzeigen, was analytischer Grammatikunterricht leisten soll und kann, dass dieser bspw. nicht zu Lasten eines Schreibunterrichts durchgeführt werden soll.

E10 Um den Fokus auf Sprache zu gewährleisten, sind die Weiterbildungsangebote auf wenige Themenfelder zu beschränken, die «Problemzonen» aufgreifen und die zugleich grundsätzliche Fragen einer guten Beurteilungs- und Förderpraxis aufwerfen.

→ Kap. 5–7

E11 Die Zusammenarbeit zwischen Klassenlehrpersonen, schulischen HeilpädagogInnen, DaZ-Lehrpersonen, LogopädInnen, sowie HSK-Lehrpersonen ist zu klären und zu stärken.

→ Kap. 9

1.2 Themenfelder

Die Empfehlungen 1–11 bilden für die Themenfelder – wie bereits in Kapitel 1.1 ausgeführt – die zentrale Grundlage. Die Themenfelder 1 bis 3 richten sich in erster Linie an die

Lehrpersonen, die Deutsch unterrichten.
Dazu zwei Hinweise:

- Prinzipiell wären diese Themenfelder auch für andere Fächer möglich: Das würde aber voraussetzen, dass die Beurteilung fachlicher Leistungen in den jeweiligen Fächern ebenfalls im Zentrum stehen müsste. Letzteres ist jedoch für den Schwerpunkt C nicht vorgesehen.
- Schulinterne Weiterbildungen für Sekundarschulen sollten möglichst alle Lehrpersonen einbeziehen. Das ist mit Themenfeld T4 ohne Weiteres möglich. Zusätzlich kann T2 in Sekundarschulen ebenfalls mit allen Lehrpersonen bearbeitet werden, sofern die sprachlichen Aspekte im Vordergrund stehen. Dagegen wären die Themenfelder T1 und T3 fachspezifisch auszurichten, was – wie bereits angemerkt – in Schwerpunkt C nicht möglich ist.

Die Kommunikation an die Schulen, welche Angebote ihnen zur Verfügung stehen, erfolgt nicht über das vorliegende Fachgutachten. Vielmehr ist dies zum einen Gegenstand einer Fachbroschüre, die sich an die Schulen richtet, zum anderen Gegenstand des Konzepts zur schulinternen Weiterbildung.

T1 Beobachtungen und Beurteilungen planen und durchführen
(= evidenzbasiertes Unterrichten)

In diesem Themenfeld stehen u.a. folgende **Leitfragen** im Zentrum:

- Wie oft wird beurteilt? zu welchem Zweck?
 - Was wird beurteilt?
 - Wie wird beurteilt?
 - Welche Ebenen werden bei der Beobachtung, Beurteilung und Rückmeldung fokussiert? (Aufgabe – Prozess – Metakognition – Person)
- Wahl der Beobachtungs-/Beurteilungsformen und -instrumente, Interpretationen

der Ergebnisse (inkl. Wahl wirksamer, passender Förderansätze) sind weitere Stichworte in diesem Themenfeld.

Hinweis 1: Dieses Themenfeld wird in der Fachliteratur oft als «evidenzbasiertes Unterrichten» bezeichnet. Diese Bezeichnung hebt stärker hervor, dass geeignete Beobachtungsformen oder -instrumente zu wählen sind. Gerade evidenzbasiertes Unterrichten ist ein wichtiger Ansatzpunkt, um systematischen Wahrnehmungsverzerrungen entgegenwirken zu können (vgl. E6). Dieses Themenfeld sollte deshalb (auch) auf Ebene Schule und im Team angegangen werden.

Hinweis 2: Dieses Themenfeld baut idealerweise auf den Online-Lerneinheiten Lehrplan 21 der PHZH auf, v.a. auf dem Baustein «Beurteilung». Entsprechend ist es günstig, wenn die Lehrpersonen vor der Weiterbildung zu Schwerpunkt C diesen Online-Baustein zu «Beurteilung» absolviert haben, da dort wichtige Grundlagen vermittelt werden.

Hinweis 3: Möchten Schulen den bisherigen QUIMS-Schwerpunkt «Schreiben auf allen Schulstufen» um den Aspekt «Beurteilen» erweitern, bietet sich in erster Linie dieses Themenfeld an.

Für den Kompetenzbereich Schreiben können folgende Aspekte im Rahmen der Weiterbildung thematisiert werden:

- *Textprodukte beurteilen:* Eine Schule eignet sich darauf, welches Genre, welche Aufgabentypen mit welchem Lernziel beurteilt werden (inkl. wann, wie oft).
 - Hier bietet es sich an, zunächst in kleineren Gruppen mit Schüler-Texten zu starten und diese von Lehrpersonen einschätzen zu lassen. Dabei ist herauszuarbeiten, wie sie an diese Texte herangehen, gerade bei schwächeren Schülern und Schülerinnen.
 - Wichtige Bezugspunkte sind hier u.a. Urteilsfehler oder Leistungserwartungen (E2).

- *Schreibprozess beurteilen*: Wie kann der Schreibprozess sichtbar gemacht werden? Wie sind die Beobachtungen zu interpretieren? (Achtung: Im Lehrplan 21 hat – neben dem Produkt – auch der Schreibprozess ein hohes Gewicht.)
- Erste Erfahrungen mit Pilotschulen belegen, dass Lehrpersonen kaum Beobachtungs- und Beurteilungsverfahren zu Prozess und Reflexion kennen. Wichtige Bezugspunkte sind hier: Repertoire (E5), Feedback-Ebenen (E6).

T2 Vorbereiten statt Nachbereiten

In diesem Themenfeld stehen u.a. folgende Fragen im Zentrum:

- Was braucht es, damit im Unterricht alle Schüler/-innen mitkommen?
- Was ist mit jenen Schülern und Schülerinnen, die mehr Herausforderung brauchen?
- Wie kann der Lernstand bzw. das Vorwissen der Schüler/-innen in die Unterrichtsplanung einbezogen werden?
- Mit Blick auf DaZ-Lernende und mit Blick auf lernschwache Schüler/-innen besonders relevant
- Die Zusammenarbeit von Klassen-Lehrperson, SHP und DaZ-Lehrperson bildet hier einen wichtigen Grundpfeiler (E11). Mehrsprachigkeit soll dabei nicht nur als Problem angesehen werden, sondern (auch) als Ressource genutzt werden.
- Dieses Themenfeld erlaubt auch den Einbezug von Übergängen, v.a. ans Gymnasium oder auch an die Berufsschule.

T3 Rückmeldungen im Dialog kommunizieren

In diesem Themenfeld stehen sowohl Fremd- als auch Selbstbeurteilungen im Zentrum. Darüber hinaus können Fremdbeurteilungen durch Lehrpersonen wie auch durch Schüler/

-innen erfolgen. Für dieses Themenfeld bieten u.a. folgende Leitfragen eine Orientierung:

- Wie werden Rückmeldungen laufend in den Unterricht integriert?
- Wie muss eine Rückmeldung formuliert sein, damit sie für Schüler/-innen nachvollziehbar, operationalisierbar ist?
- Wie muss eine Rückmeldung formuliert sein, damit sie für andere Lehrpersonen (SHP, DaZ-Lehrpersonen etc.) sowie für die Eltern nachvollziehbar ist?
- Wie können die Schüler/-innen selbst über ihr Lernen sprechen?
- Wichtig ist, in den Schulen festzulegen, wie dies stufenübergreifend thematisiert werden kann.
- Dabei soll grundsätzlich beachtet werden, auf welcher Ebene das Feedback stattfindet (vgl. Empfehlung 7).

T4 In allen Fächern sprachbewusst fördern und beurteilen

Es stehen in diesem Themenfeld u.a. folgende Fragen im Zentrum:

- a) NT, Geografie, Mathe, Geschichte (v.a. formativ) (→ Sekundarstufe I)**
 - Wie kann die für das Fach nötige Sprache aufgebaut und gefördert werden?
 - Wie kann fachliche Leistung beurteilt werden, wenn sprachliche Fähigkeiten noch unzureichend sind?
 - Was brauchen Lehrpersonen, um in diesem Sinne sprachbewusst unterrichten zu können?
- b) Sprache in allen Fächern fördern und beurteilen (→ Kindergarten, Primarstufe)**
 - Wie kann die für das fachliche Lernen nötige Sprache aufgebaut und gefördert werden?
 - Wie können sprachliche Fähigkeiten fachübergreifend beurteilt werden?

- Was brauchen Lehrpersonen, um in diesem Sinne sprachbewusst unterrichten zu können?

Hinweis 1: «Die für das Fach nötige Sprache» kann sowohl fach- wie auch bildungssprachliche Elemente umfassen. Mit «bildungssprachlich» ist in erster Linie die für das Lernen nötige Sprache gemeint: Das schliesst bspw. Sprache ein, mithilfe derer über das eigene (fachliche) Lernen reflektiert wird.

Hinweis 2: Die sprachliche Unterstützung der im Fachunterricht sprachlich vermittelten Inhalte wird als «sprachbewusster Unterricht» bezeichnet. Die Frage, wie Mehrsprachigkeit als Ressource genutzt werden kann, ist dabei noch weitgehend ungeklärt.

Zu klären ist, inwiefern im Rahmen einer schulinternen Weiterbildung auch Fachpersonen aus nicht-sprachlichen Fächern als Leitungspersonen beigezogen werden können.

2 Zur Bedeutung des Beurteilens

Beurteilen und Diagnostizieren gehört zu den Kernkompetenzen einer Lehrperson. Dieser Kernbereich umfasst die Tätigkeiten von Lehrpersonen, die dazu dienen, Informationen über Lernvoraussetzungen, Lernprozesse und Lernergebnisse von Schülerinnen und Schülern zu beschaffen und zu interpretieren. Diese Informationen und Interpretationen bilden die **Grundlage für pädagogische und didaktische Entscheide**, die einerseits einzelne Lernende, aber andererseits auch die gesamte Unterrichtsgestaltung betreffen (vgl. Schrader, 2013, S. 154).

So zeigt insbesondere formatives Beurteilen (inkl. Feedback) mit $d = .90$ über verschiedene Fächer hinweg einen grossen positiven Effekt auf die Lernleistungen (Hattie, 2009). Für die einzelnen Fächer und auch Kompetenzbereiche fällt dies i. d. R. sehr ähnlich aus, so etwa für den Bereich Schreiben (mit $d = .87$, vgl. Graham, Hebert & Harris, 2015).

Die schulische Beurteilung einzelner Schüler/-innen folgt deshalb den didaktischen Prinzipien des Unterrichts, da sie die Informationen für vier unterschiedliche Typen von Entscheidungen liefern muss, die eine Lehrperson fällen muss (Hosp, 2012, S. 89):

- 1) **Informationen gewinnen:** Wo steht die Schülerin, der Schüler zu einem festgelegten Zeitpunkt? Wie gross ist ihr oder sein Förderbedarf?
Lehrpersonen prüfen die individuellen Lernvoraussetzungen der Schüler/-innen und beurteilen, inwiefern die Schüler/-innen mit hoher Wahrscheinlichkeit ein bestimmtes Lernziel erreichen werden. Sie klären ab, welchen Förderbedarf einzelne Schüler/-innen haben. Entscheidungen dieses Typs haben damit sowohl formativen wie auch prognostischen Charakter.
- 2) **Lernfortschritte beobachten:** Wie gross ist ihr/sein Lernfortschritt über eine bestimmte Zeitperiode?

Die Zeitperiode, die Entscheidungen dieses Typs zugrunde liegt, ist abhängig vom jeweiligen Zweck oder Lernziel. Entscheidungen können zum einen auf der Basis von Verfahren oder Instrumenten erfolgen, die nah an den Lernaufgaben oder am Klassenziel sind, zum anderen mit standardisierten Instrumenten, die einen eher «externen» Blick erlauben. Letztere werden i. d. R. ein- bis zweimal pro Schuljahr eingesetzt, während Erstere einem formativen Beurteilungskreislauf folgen und je nach Instrument auch häufiger eingesetzt werden können, so etwa, wenn es sich um so genannte Curriculum-basierte Messverfahren (= CBM) handelt (vgl. Deno, 1985).

- 3) **Förderplanungen erstellen:** Welche Fördermassnahme braucht eine Schülerin/ ein Schüler aufgrund der bisherigen Informationen und Beobachtungen?
Entscheidungen dieses Typs betreffen damit in erster Linie die Frage, welche Art der Vermittlung zielführend ist (bspw. implizite vs. explizite Vermittlung von Lesestrategien). Das schliesst nicht aus, dass die Frage, welcher Bereich gefördert werden soll, nochmals in den Blick geraten kann.
- 4) **Lernergebnisse überprüfen:** Welche Wirkung zeigt die (kompetenzorientierte) Förderung bei den Schülern, Schülerinnen?
Hierbei handelt es sich um summative Beurteilungen und in der Regel auch um dichotome Entscheidungen: Haben die SuS eine bestimmte Lerneinheit verstanden, ein bestimmtes Lernziel erreicht?

Damit Entscheidungen getroffen werden können, müssen Methoden und Verfahren – dazu zählen auch Gespräche – bzw. Instrumente eingesetzt werden, die möglichst reliabel und valide, dennoch aber auch ökonomisch einsetzbar sind.

Darüber hinaus müssen Verfahren oder Instrumente – das können sowohl alltagsnahe wie auch stärker standardisierte Verfahren und Instrumente sein – je nach Entscheidungstyp weitere Merkmale aufweisen. So ist es für formative Entscheidungen (z.B. Förderbedarf feststellen, Förderplanung festlegen) wichtig, dass die Beurteilungsinstrumente die Grundlage für ein Feedback liefern, das hilfreich für die weitere Instruktion ist, dass die Beobachtung und Beurteilung leicht durchzuführen und mit den Lehr-/Lern-Arrangements gut abgestimmt ist (Hosp, 2012). Das setzt seitens Lehrperson einiges voraus (Timperley, 2011): Nicht nur müssen sie geeignete Verfahren bzw. Instrumente kennen und auswählen, sondern die Informationen daraus auch interpretieren. Persönliche Theorien beeinflussen dabei nicht nur die Interpretation, sondern bereits die Auswahl der Verfahren oder Instrumente, vor allem aber auch, welche Informationen berücksichtigt werden. Das gilt es gerade auch bei der Konzipierung geeigneter Massnahmen für schulinterne Weiterbildungen (=SchilW) zu berücksichtigen.

Des Weiteren können diese unterschiedlichen Entscheidungen sowohl aus einer langfristigen oder eher kurzfristigen Perspektive heraus erfolgen und haben entsprechend einen *vorausplanenden Aspekt* (Burns, 1984):

- **Langfristige Entscheidungen** beziehen sich auf eine grössere Zeitperiode und umfassen dabei auch grössere Einheiten. Eine wichtige Grundlage bilden dabei Bildungsstandards und Lehrpläne sowie Lehrmittel (dabei v.a. Jahresplanungen).
- **Kurzfristige Entscheidungen** beziehen sich in erster Linie auf Lektionen oder Lernsequenzen, die mehrere Lektionen oder auch nur Segmente einer Lektion umfassen können. Entsprechend spielt die Wahl von Aufgaben, die instruktionale

Umsetzung etc. dabei eine entscheidende Rolle. Entscheidungen dieser Art werden hauptsächlich im wöchentlichen oder täglichen Rhythmus gefällt.¹

Wie Burns (1984) hervorhebt, müssen Lehrpersonen im Verlauf ihres Unterrichts zudem zahlreiche Entscheidungen fällen, die aus einer Situation heraus erfolgen und damit nicht vorausschauend eingeplant werden können:

- **Situative bzw. interaktive Entscheidungen** beziehen sich auf einzelne Episoden und erfolgen in erster Linie in der Interaktion mit den Schülern und Schülerinnen. Solche Interaktionen können eng getaktet erfolgen – sie werden deshalb auch als «minute-to-minute interactions with students» bezeichnet (Burns, 1984, S. 98) – und setzen voraus, dass die Lehrperson die Schüler/-innen beim Lösen bzw. Bewältigen einer Aufgabe beobachten bzw. überwachen und bei Bedarf die Schüler/-innen stärker unterstützen oder auch direkt in der Situation eine Rückmeldung geben kann.

Instruktionale Routinen wie das Bilden von Lerngruppen ordnet Burns (1984) den langfristigen Entscheidungen zu, während Klassenmanagement u.Ä. als Teil der situativen Entscheidungen betrachtet werden.

Ein zentrales Element von kurzfristigen vorausplanenden Entscheidungen ist, dass die Beobachtung und Beurteilung an die entsprechenden Aufgaben gebunden ist, wie Burns (1984, S. 98) verdeutlicht. Das bedeutet nicht zuletzt, dass die Frage, wie beobachtet und beurteilt werden soll, bereits bei der Wahl oder Entwicklung einer Lerneinheit und ihren einzelnen Aufgaben zu beantworten ist.

Bei diesem formativen Beurteilen spielt das Feedback eine entscheidende Rolle. Feedback kann jedoch nicht nur auf der Ebene der

¹ Die QUIMS-Handreichung zu Qualitätsmerkmalen räumt kurzfristigen und vorausplanenden Entscheidungen ein hohes Gewicht ein (Bildungsdirektion Kanton Zürich, 2003, S. 14).

Aufgabe erfolgen, sondern auch auf der Prozessebene und der Ebene der Selbstregulation. Hattie & Gan (2011) formulieren zu den einzelnen Ebenen u.a. folgende Leitfragen:

- a) **Ebene Aufgabe:** Erfüllt die Antwort oder Lösung die zuvor festgelegten Kriterien? Was hat der Schüler, die Schülerin gut bzw. weniger gut gelöst? Was wäre eine angemessene oder korrekte Lösung?
- b) **Ebene Prozess:** Welche Strategie hat der Schüler, die Schülerin angewendet? Was ging dabei schief? Wie hat er/sie die Aufgabe verstanden?
- c) **Ebene Selbstregulation:** Wie kann der Schüler, die Schülerin das eigene Vorgehen, die bisherige Arbeit überwachen? Kann er/sie die eigene Arbeit, das eigene Vorgehen evaluieren?

Wirksames Feedback – sei es im Rahmen von kurzfristigen oder situativen Entscheidungen – setzt zumindest auf einer dieser Ebenen an. Die Wirkung kann verstärkt werden, wenn das Feedback mehrere dieser Ebenen einbezieht. Wie Hattie & Gan (2011) aber betonen, zeigt Feedback, das auf der persönlichen Ebene ansetzt, keinen positiven Effekt auf die Lernleistung der Schüler/-innen. Gemeint sind damit nicht individuelle Rückmeldungen, sondern solche Rückmeldungen, die sich auf das Verhalten, auf «innere Zustände» des Schülers, der Schülerin beziehen.² Feedback dieser Art führt i. d. R. dazu, dass die Verantwortung für das Lernen mehrheitlich an die Lernenden delegiert wird, indem sie aufgefordert werden, sich mehr Mühe zu geben, sorgfältiger zu arbeiten u.Ä. Gleichzeitig bedeutet dies auch, dass auf diese Weise keine Implikationen für die Unterrichtsgestaltung abgeleitet werden können. Gerade jedoch niedrige Lernleistungen sollten als Zusam-

menspiel von Unterricht und Lernen betrachtet werden, in diesem Sinne auch zu einem systematischen Hinterfragen führen und als Gelegenheit des professionellen Lernens verstanden werden (Timperley, 2011). Das Zusammenspiel von Unterricht und Lernen hängt nicht zuletzt mit Leistungserwartungen zusammen und wird von Timperley (2011, S. 68f.) auch zur Einschätzung des Weiterbildungsbedarfs von Schulen eingesetzt:

Expectations of students

Basic: Low student achievement and engagement was presented as a problem resting with students, not with teaching.

[...]

Integrated: Low student achievement and engagement was presented as a teaching/learning/relationship issue that requires a systematic approach to professional learning.

Beurteilungen und Rückmeldungen wirken sich nicht nur auf Lernleistungen i.e.S. aus, sondern auch auf motivationale Einstellungen der Schüler und Schülerinnen (Bong & Skaalvik, 2003). So beeinflussen summative Beurteilungen in erster Linie das Selbstkonzept in einem bestimmten Bereich, oftmals bezogen auf ein Schulfach (Bsp.: «Ich bin gut in Deutsch», «Ich bin schlecht im Schreiben von Texten»).

Mit der Wahl von Aufgaben, der instruktionalen Umsetzung und formativen Rückmeldungen zum weiteren Lernweg können dagegen die Überzeugung und Zuversicht von Schülern und Schülerinnen, eine bevorstehende Aufgabe zu lösen, beeinflusst werden. Dies stärkt entsprechend die Selbstwirksamkeit der Lernenden. Letzteres stellt besondere didaktische Herausforderungen an die Lehrpersonen, da Lernaufgaben für Schüler/-innen mit unterschiedlichen Voraussetzungen bewältigbar, gleichzeitig aber auch herausfordernd sein sollten.

² Beispiel aus einem studentischen Leistungsnachweis an der PH FHNW: «Das Erste, was mir auffiel, war seine Ignoranz gegenüber Linien.»

3 Grundlagen der Beurteilung

In diesem Kapitel werden die allgemeinen Grundlagen kurz dargelegt, die für die Arbeit am neuen Schwerpunkt «Beurteilen und Fördern mit Fokus auf Sprache» von besonderer Relevanz sind, insbesondere mit Blick auf die schulinternen Weiterbildungen, die mit möglichst vielen der gegenwärtig 130 QUIMS-Schulen durchzuführen sind.

Um die verschiedenen Typen von pädagogischen Entscheidungen nach Hosp (2012) oder Burns (1984) fällen zu können, benötigt eine Lehrperson also für die schulische Beurteilung vielfältige und möglichst verlässliche Informationen, die sie durch genaue Beobachtung bzw. Abklärungen beschaffen muss.

3.1 Zyklische Prozessmodelle zur Diagnostik und Förderung

Diese Informationsbeschaffung und -verarbeitung im Rahmen der Diagnostik und Förderung versteht Klug (2017) mit Abstützung auf das zyklische Modell der Selbstregulation von Zimmerman (2002) als einen Kreislauf in drei Phasen:

- 1) Die präaktionale Phase umfasst die Vorbereitungshandlungen der Diagnostik (klare Ziele setzen, Bezugsnorm definieren, diagnostisches Wissen aktivieren, Beobachtungs- und Beurteilungsprozesse planen).
- 2) Die aktionale Phase beschreibt das eigentliche diagnostische Handeln (Informationen aus verschiedenen und verlässlichen Quellen sammeln, auswählen, gewichten und bewerten).
- 3) Die postaktionale Phase beinhaltet Nachbereitungshandlungen (individuelle Förderpläne zusammenstellen, Feedback geben).

Klug (2017) setzt damit die Diagnosekompetenz und den diagnostischen Prozess der schulischen Beurteilung in den Kontext des selbstregulierten Lernens und zeigt auf, wie drei Phasen in einem Kreislauf aufeinander

bezogen werden können. Das Modell wurde in einer Studie empirisch validiert, in der die befragten Teilnehmerinnen und Teilnehmer (N = 293 Lehrpersonen, Referendarinnen und Referendare, Lehramtsstudierende) mittels Fragebogen verschiedene Lernprozesse in einem Szenario beurteilen sollten (vgl. Klug, 2017, S. 57). Insbesondere die dreidimensionale Struktur konnte durch die empirischen Daten abgestützt werden. Der diagnostische Prozess, zu dessen Validierung auch nach möglichen Ursachen der im Szenario dargestellten Lernprozesse gefragt wurde, korreliert mit der Genauigkeit der Diagnose. Als wichtige Punkte zur Voraussagbarkeit einer hohen Diagnosekompetenz stellten sich die Motivation zum Diagnostizieren und das Wissen über Diagnostik heraus, während reflektierte Erfahrung entgegen der Erwartung eine geringe Voraussagbarkeit aufwies.

Das zyklische Prozessmodell zur Diagnostik und Förderung von Klug (2017) legt als Rahmenmodell eine Klammer um die wichtigen diagnostischen Prozesse der Beobachtung, des Beurteilens und der Förderung. Es betont auch mit der Abstützung auf die Konzepte des selbstregulierten Lernens die Wichtigkeit von genauen Zielen für den diagnostischen Prozess, die in der vorbereitenden präaktionalen Phase gesetzt werden. Der zyklische Diagnoseprozess kann aus Sicht der Lehrpersonen als Fremdbeurteilungs-, aus Sicht der Lernenden als Selbstbeurteilungskreislauf dargestellt werden (vgl. Abbildung 2, S. 13). Senn (2018) unterscheidet dabei im Kreislaufmodell der Beurteilung die vier Teilprozesse Ziele setzen, Beobachten, Evaluieren und Feedback geben, indem die aktionale Phase des diagnostischen Handelns weiter in Beobachtungs- und Evaluationsprozesse differenziert wird, um die Beschaffung der Information von deren Evaluation zu trennen (vgl. Senn, 2018, S. 190 ff.; Lötscher, Tanner Merlo, & Joller-Graf, 2017, S. 22; Senn, Lötscher, &

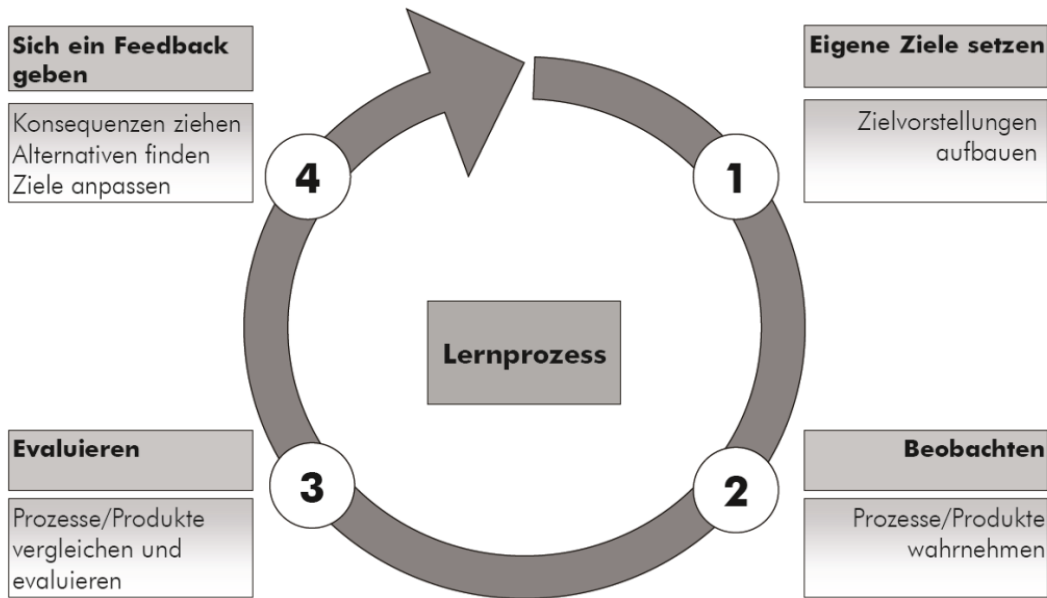


Abbildung 2: Kreislaufmodell der Selbstbeurteilung (vgl. Senn, 2018, S. 192)

Malti, 2005, S. 14). Die Ziele regulieren dabei die Beobachtungen, indem sie die Aufmerksamkeit auf die fokussierten Beobachtungspunkte richten, sei es auf Lernprozesse oder Lernprodukte, und so die Informationsbeschaffung steuern.

Ein wesentlicher Aspekt dieses Selbst- bzw. Fremdbeurteilungsmodells betrifft die dialogischen Feedbackgespräche zwischen der Lehrperson und den Lernenden. In diesen Gesprächen werden Fremd- und Selbstwahrnehmungen bzw. -einschätzungen ausgetauscht und überprüft. Die Lernenden bauen dadurch validere und realistischere Ziele, Kriterien oder Gütestandards für ihre Selbsteinschätzungen auf, die Lehrperson lernt im Gespräch die Sicht und die Überlegungen der Lernenden kennen und kann sie in ihrer Einschätzung und der weiteren Förderplanung berücksichtigen.

Mit Bezug auf die pädagogischen Entscheidungstypen von Hosp (2012) kann diese Beobachtungssituation eine *Lernstandserfassung* sein, wenn sie zu Beginn einer grösseren Unterrichtseinheit eingesetzt wird. Sie kann aber auch den *Lernfortschritt* ermitteln, wenn

sie nach einer bestimmten Zeitperiode eingesetzt wird, um formativ zu ermitteln, inwieweit ein Lernziel im Rahmen eines Lernprozesses bereits erreicht ist. Die gesetzten Ziele steuern deshalb auch den zweiten Beurteilungsprozess, das eigentliche Evaluieren, in dem die wahrgenommenen Beobachtungen mit den gesetzten Zielvorstellungen verglichen werden. Diese Diagnose führt dann zu einem Feedback, in dem Konsequenzen für den weiteren Lernverlauf formuliert werden. Diese Konsequenzen können Teil einer *Förderplanung* sein, die auf die gesetzten Ziele ausgerichtet ist und welche möglichst konkret definiert, welche Massnahme als nächster Schritt im Lernprozess ergriffen werden soll. Eine Konsequenz aus diesen Beurteilungsprozessen kann natürlich auch sein, dass die Ziele selbst revidiert werden, indem bspw. anspruchsvollere oder weniger anspruchsvolle Ziele gesetzt werden.

Das zyklische Verständnis unterstreicht den engen Zusammenhang von Diagnose- und Förderprozessen beim Lernen. Es entspricht einem formativen Beurteilungsverständnis (vgl. Handreichung «Kompetenzorientiert be-

urteilen», Koch et al., 2017, S. 6). Das formative Beurteilen und die Selbstbeurteilung stellen gemäss Hattie (2009) sehr wirkungsvolle unterrichtsgestaltende Massnahmen dar (vgl. Kap. 4.2).

Im konkreten Unterrichtsgeschehen können diese Fremd- oder Selbstbeurteilungsmassnahmen nicht nur kurzfristiger oder situativer Art im Sinne von Burns (1984) sein, sondern auch einen unterschiedlichen Systematisierungsgrad aufweisen. Groeben (2009) unterscheidet dabei beispielsweise zwischen unsystematisch erhobenen Daten aus zufälligen Beobachtungen in unterschiedlichen Unterrichtssituationen, teilsystematischen Daten, die beispielsweise in einer geplanten Vorlesesituation erhoben werden, und systematisch erhobenen Daten mit standardisierten Testinstrumenten, z.B. mit dem Instrument LDL (Lernfortschrittsdiagnostik Lesen) von Walter (2009). Entsprechend dem zyklischen Verständnis führen die Beurteilungsprozesse zu den entsprechenden Fördermassnahmen und diese wiederum zu einem weiteren Zyklus von Beurteilungs- und Förderprozessen, und zwar unabhängig vom Systematisierungsgrad der durchgeführten Beobachtung. Zeitigt also bspw. eine Fördermassnahme nicht den gewünschten Erfolg, kann dies dazu führen, dass die bisher routiniert eingesetzte Fördermassnahme überdacht und bei Bedarf verändert oder durch eine andere ersetzt werden muss.

Teachers with adaptive expertise, therefore, have the capability to identify when known routines do not work and to seek new information about different approaches when needed. (Timperley, 2011, S. 12)

Timperley (2011) geht dabei davon aus, dass Lehrpersonen nicht immer auf bestehendes Wissen zurückgreifen können, sondern dass sie sich bei Bedarf über alternative Förderansätze informieren. Dies verlangt eine kontinuierliche Weiterentwicklung der diagnostischen und didaktischen Kompetenzen der Lehrpersonen. Bedenkt man, dass es Renkl

(2008) zufolge eine Vielzahl verschiedener instruktionaler Modelle oder Ansätze gibt – so projektorientiertes Lernen und Lehren, Lernen durch Selbsterklären, Beobachtung von Peers oder direkte bzw. explizite Vermittlung, um nur wenige zu nennen –, kommt einer «forschenden» Haltung seitens Lehrperson, es genauer wissen zu wollen, ein nicht zu unterschätzendes Gewicht zu. Vor allem schulinterne Weiterbildungen zu formativem Beurteilen sind Timperley (2011) zufolge so anzulegen, dass sie diese «forschende» Haltung fördern, da sich sonst Widerstände gegen die Weiterbildung, gegen die betreffende Lernstandserfassung etc. breit machen können. Dies ist eine zentrale Konsequenz für die Planung von Weiterbildungen im Bereich Beurteilen.

3.2 Adaptive Lehrkompetenz als diagnostische und didaktische Kompetenzen

Diagnosekompetenz im schulischen Kontext bezieht sich nicht nur auf die individuelle Ebene der Beurteilung des einzelnen Schülers, der einzelnen Schülerin, auf die sich die Entscheidungsprozesse von Hosp (2012) und die zyklischen Modelle von Klug (2017) und Senn (2018) vorwiegend beziehen, sondern können auch den gesamten Unterricht betreffen. Deshalb kann eine Lehrperson eine Lernstandserfassung oder eine Lernfortschrittsüberprüfung nicht nur als Informationsquelle über ihre Schüler/-innen, sondern auch über ihren Unterricht betrachten. Diagnostische Kompetenzen sind also auch wichtig für eine adaptive Unterrichtsvorbereitung bzw. Klassenführung.

Brühwiler (2014, 2017) hat diese adaptive Lehrkompetenz von Lehrpersonen untersucht. Er setzt adaptive Lehrkompetenz mit kognitiven Prozessen gleich, die sich als Zusammenspiel von Sach-, diagnostischer, didaktischer und Klassenführungskompetenz gestaltet. Dabei betont er, dass adaptive Lehrkompetenz sich nicht nur auf adaptiven

Unterricht bezieht, sondern auch die Unterrichtsplanung, also die präaktionale Phase, umfasst. In seinem Modell beinhaltet die adaptive Lehrkompetenz die Planungskompetenz ebenso wie die Handlungskompetenz der Klassenführung, und beide Bereiche setzen sich aus diagnostischen und didaktischen Kompetenzen zusammen (vgl. Brühwiler, 2017, S. 125). Die diagnostische Planungskompetenz bezieht sich dabei auf das Antizipieren möglicher Unterrichtsverläufe mit dem Einplanen von Beobachtungs- und Beurteilungsanlässen, auf den Aufbau von diagnostischem Wissen sowie auf das Klären von Lernvoraussetzungen der Schüler/-innen und Zielsetzungsprozesse. Die diagnostische Handlungskompetenz würde sich vorwiegend auf das Prüfen des Lernstands (Lernstandserfassung) bzw. der Zielerreichung (Lernfortschritt) beziehen.

In seiner Untersuchung zu diagnostischen und didaktischen Kompetenzen von Lehrpersonen im naturwissenschaftlichen Fachbereich – durchgeführt im 4./5. und im 7./8. Schuljahr – kann Brühwiler (2014, 2017) zeigen, dass diese diagnostischen bzw. didaktischen Planungs- und Handlungskompetenzen einen Prädiktor für den Lernzuwachs der SuS darstellen. Dabei stellt sich die diagnostische Seite in der Unterrichtsplanung als besonders wichtig heraus, insbesondere die vorgängige Abklärung der Lernvoraussetzungen (v.a. des Vorwissens), während bei der Unterrichtsdurchführung die didaktische Handlungskompetenz, nämlich der Einsatz geeigneter Fördermassnahmen, im Vordergrund steht. Dies bedeutet, dass genaue Abklärungen, die in die Unterrichtsplanung einfließen, also nur dann zu einem Lernzuwachs führen, wenn sie mit passenden, im Unterricht umgesetzten Fördermassnahmen gekoppelt sind. Wirksame adaptive Handlungskompetenz setzt sich demzufolge aus einer genauen diagnostischen Planung und einer angemessenen didaktischen Umsetzung als Fördermassnahme zusammen. Dies entspricht einer formativen Beurteilung (vgl. Kap. 4.2).

3.3 Urteilsakkuratheit und diagnostische Kompetenz von Lehrpersonen

Südkamp & Praetorius (2017) präsentieren einen Überblick über verschiedene Ansätze zur Modellierung diagnostischer Kompetenz von Lehrpersonen. Dabei halten Karst & Förster (2017) fest, dass der überwiegende Anteil der wissenschaftlichen Studien diagnostische Kompetenz mit Urteilsakkuratheit gleichsetze. Dennoch zeigen sich vielfältige Unterschiede zwischen den Ansätzen, sodass von einer einheitlichen Vorstellung keine Rede sein kann. Nicht zuletzt wird oft kritisch eingewendet, dass eine Reduzierung auf Urteilsakkuratheit den Kompetenzbegriff vernachlässige. Es ist hier nicht der Ort, überblicksartig die verschiedenen Ansätze und Vorstellungen darzustellen und detailliert zu diskutieren. Deshalb werden im Folgenden nur einige zentrale Befunde zur Urteilsakkuratheit von Lehrpersonen kurz dargestellt.

Allgemein formuliert kann Urteilsakkuratheit als die zutreffende Vorhersage künftiger Schülerleistungen bestimmt werden, ohne dass dabei die Vorhersage die Schülerleistungen beeinflusst (Jussim, Robustelli & Cain, 2009, S. 362). Einen wichtigen Ansatzpunkt bildet die Arbeit von Schrader (1989), der zufolge zwei Arten des Urteils sowie drei Komponenten von diagnostischer Kompetenz zu unterscheiden sind:

Urteilsart: a) aufgabenbezogen (Wie viele Lernende werden diese Aufgabe lösen?), b) personenbezogen (Wie viele Aufgaben werden von einem bestimmten Schüler/einer bestimmten Schülerin richtig gelöst?)

Komponenten: a) Niveau (Akkuratheit in der Einschätzung der mittleren Aufgabenschwierigkeit bzw. der mittleren Schülerleistung), b) Differenzierung (Akkuratheit in der Einschätzung der Heterogenität von Aufgabenschwierigkeiten bzw. von Schü-

lerleistungen), c) Rangordnung (Akkuratheit in der Einschätzung der Rangreihe der Aufgabenschwierigkeiten bzw. der Schülerleistungen)

In der Forschung zur Urteilsakkuratheit spielen die Niveau- und die Differenzierungskomponente eher eine untergeordnete Rolle. In vielen Studien wurde, wie schon bei Schrader (1989), vor allem die Rangordnungskomponente (In welcher Reihenfolge befinden sich die Aufgaben in Bezug auf ihre Schwierigkeit? In welcher Reihenfolge befinden sich die Schüler und Schülerinnen in Bezug auf ihre Leistungen?) als Indikator der Urteilsakkuratheit beigezogen, insbesondere die Ausrichtung auf die Schülerleistung (vgl. Helmke, Hosenfeld & Schrader, 2004). Diese doppelte Reduktion (1. Rangordnungskomponente für Urteilsakkuratheit, 2. Urteilsakkuratheit für die gesamte diagnostische Kompetenz) wird von Karst (2017, S. 24f.) als problematisch kritisiert. Behrmann & Glogger-Frey (2017, S. 135f.) weisen zudem darauf hin, dass die Urteile der Lehrpersonen präziser sind, wenn diesen mehr Informationen über das einzuschätzende Merkmal zur Verfügung standen, insbesondere, wenn sie ein Ankerkriterium mit demselben Spezifitätsgrad nutzen konnten. Diese Informationen sind jedoch sehr fach- und situationsspezifisch, weshalb Behrmann & Glogger-Frey (2017, S. 135f.) festhalten, dass die über die Urteilsakkuratheit gemessene diagnostische Kompetenz nicht von einem auf einen anderen Bereich übertragen werden kann. Sie vermuten sogar, dass die drei Komponenten gar nicht geeignet sind, um die diagnostische Kompetenz umfassend zu erheben, was sich vor allem in denjenigen Studien zeige, die die Auswirkung der Urteilsakkuratheit auf die Schülerleistungen analysieren.

So scheinen positive Effekte der Urteilsakkuratheit auf die Leistung der Schüler und Schülerinnen erst dann aufzutreten, wenn sie durch ein förderndes Verhalten im Unterricht unterstützt werden, worauf die Studie von

Behrmann und Souvignier (2013) hinweist. Die beiden Autoren haben den Einfluss der Urteilsgenauigkeit und der Unterrichtsgestaltung auf die Lesekompetenz untersucht, wobei die Unterrichtsgestaltung v.a. die Feedbacks umfassten, die die Lehrpersonen den Schülern und Schülerinnen beim Lesen gaben. Als wesentlichen Befund halten die Autoren fest, dass sich die Feedbackhäufigkeit positiv auf die Lesekompetenzentwicklung auswirkt, wenn sie mit einer hohen Urteilsgenauigkeit einhergeht. Eine hohe Urteilsgenauigkeit mit wenig Feedbacks hingegen zeigt einen deutlich geringeren positiven Effekt. Bei Lehrpersonen mit einer geringeren Urteilsgenauigkeit wirkt sich unterschiedliche Feedbackhäufigkeit nicht auf die Lesekompetenz aus. Damit weist diese Studie im Kontext des Leseunterrichts in dieselbe Richtung wie die oben erwähnte Untersuchung von Brühwiler (2014, 2017) im Rahmen des naturwissenschaftlichen Unterrichts.

Auch die Studie von Karing, Pfof & Artelt (2011) stellte einen Zusammenhang zwischen dem aufgabenspezifischen Treffer und der Entwicklung der Lesekompetenz fest, wenn zur Urteilsakkuratheit – hier über die aufgabenspezifischen Treffer ermittelt (Welche Aufgaben können die Schüler/-innen lösen?) – noch didaktische Fördermassnahmen wie Vorentlastungen zu Beginn des Lesens oder Zusammenfassungen nach Abschluss eingesetzt werden. Ein wichtiger Befund der Studie ist, dass eine höhere Individualisierung als unterrichtsgestaltende Massnahme die Wirkung der diagnostischen Fähigkeiten verstärkt. Einschränkend zeigt der Einsatz von Strukturierungshilfen nur bei den leistungsschwächeren Schüler/-innen einen positiven Effekt. Offensichtlich sind sie anders als leistungstärkere Schüler/-innen auf eine grössere Strukturierung angewiesen. Dieser Effekt zeigte sich allerdings nur in Bezug auf die Leseförderung. Im Fach Mathematik konnte die Studie keinen Zusammenhang zwischen

Diagnosegenauigkeit, mathematischer Kompetenzentwicklung und der Unterrichtsgestaltung feststellen.

Südkamp, Kaiser und Möller (2012) können in ihrer Meta-Analyse zur Urteilsakkuratheit von Lehrpersonen eine recht gute Urteilsgenauigkeit feststellen. Auf der Basis von 73 Studien ermitteln sie eine mittlere Korrelation (Median) von .53, die aber auch auf ein Verbesserungspotenzial der Studien hinweist. Das zeigt sich u.a. daran, dass teilweise erhebliche Unterschiede zwischen den einzelnen Studien bestehen. In der weiteren Analyse streichen Südkamp et al. (2012) zwei weitere Befunde heraus:

- Die Urteilsakkuratheit fällt bei ‹informiertem› Beurteilen höher aus als bei ‹nicht-informiertem› Verfahren. Zu ‹informiertem› Beurteilen zählen sie Verfahren, in denen die Lehrpersonen bspw. angeben sollen, wie viele Items jeder Schüler und jede Schülerin korrekt lösen wird. Sollen die Lehrpersonen die Leistung der Schüler/-innen in einem bestimmten Bereich grob auf einer Likert-Skala einschätzen, zählt dieses Verfahren Südkamp et al. (2012) zufolge zu ‹nicht-informiertem› Beurteilen.
- Hinsichtlich der beiden Bereiche Sprache (mehrheitlich handelt es sich dabei um Studien zu Lesekompetenzen) und Mathematik zeigen sich keine signifikanten Unterschiede in der Urteilsgenauigkeit.

Das Beurteilen von Textprodukten mit Kriterienrastern, wie es in jedem Lehrmittel angeboten wird, zählt – folgt man Südkamp et al. (2012) – entsprechend zu ‹nicht-informiertem› Beurteilen. So stellen Graham, Harris & Hebert (2011) fest, dass eine zuverlässige Beurteilung von Textprodukten in deutlich mehr als der Hälfte der von ihnen einbezogenen Studien nicht gegeben war, unabhängig davon, ob analytische oder holistische Beurteilungsraster zum Einsatz kamen. Dies legt die Vermutung nahe, dass dieser Bereich entsprechend dem Schluss von Behrmann &

Glogger-Frey (2017) sich doch stark von anderen Bereichen unterscheiden könnte.

Zusammenfassend lässt sich feststellen, dass die Verbesserung von diagnostischen Kompetenzen, insbesondere der Urteilsakkuratheit, die in den Studien häufig als Indikator für die gesamte diagnostische Kompetenz verwendet wird, sich nicht automatisch auf bessere Lernleistungen der Schüler/-innen niederschlägt. Diagnostische Kompetenzen zeigen Auswirkungen auf den Lernerfolg, wenn sie mit angemessenen didaktischen Umsetzungen kombiniert werden. Fließt also das Ergebnis einer Lernstandserfassung in die Unterrichtsplanung bzw. eine Förderplanung ein, beispielsweise indem das Vorwissen der Schüler/-innen abgeklärt wird, braucht es zusätzlich angemessene, darauf abgestützte didaktische Massnahmen, die im Unterricht bzw. in einer Förderung umgesetzt werden, z.B. förderliche Lernstrukturen wie Vorentlastung beim Lesen, gezieltes Feedback oder fokussierende Zusammenfassungen nach Abschluss, um die Leistung der Lernenden zu unterstützen.

Dies entspricht einem zyklischen und dialogischen Verständnis von Lernen und Beurteilen, wie es von den Modellen zum selbstregulierten Lernen von Zimmerman (2002) bzw. in Bezug auf Diagnostizieren von Klug (2017) angenommen wird und welches die Basis des Beurteilungs- und Förderkreislaufes von Senn (2018) darstellt. In diesem Kreislauf wird aufgezeigt, wie Ziele setzen, beobachten, evaluieren und Feedback geben sich gegenseitig bedingen und ein formatives Beurteilen ermöglichen. Diese Befunde müssen für eine Planung von Weiterbildungen der Lehrpersonen im Bereich der diagnostischen Kompetenz berücksichtigt werden. Darüber hinaus werden im Feedbackgespräch Fremd- und Selbsteinschätzung ausgetauscht und in einem dialogischen Prozess überprüft, was zu einer kommunikativen Validierung von Zielen, Kriterien und Gütestandards führt (vgl. Kap. 5.2).

4 Beurteilungsarten und ihre Ziele

4.1 Bezugssystem

Beurteilen heisst immer, eine gemachte Beobachtung im Rahmen eines Beurteilungssystems mit der entsprechenden Norm in Bezug zu setzen und zu vergleichen, inwiefern die Beobachtung von der gesetzten Norm abweicht. Grundsätzlich werden drei unterschiedliche Bezugssysteme unterschieden: Die *Sachnorm*, die *Individualnorm* und die *Sozialnorm* (Rheinberg, 2006, 2008):

- Mit dem Bezugssystem der **Sachnorm** wird die Beobachtung mit gesetzten Kriterien verglichen. Diese Kriterien werden aus den Kompetenzbeschreibungen des Lehrplans 21 abgeleitet (vgl. Handreichung «Kompetenzorientiert beurteilen»: Koch et al., 2017, S. 6). Beurteilungen mit diesem Bezug geben Auskunft, inwiefern eine Schülerin oder ein Schüler die aus dem Lehrplan abgeleiteten Lernziele erreicht hat.
- Die **Individualnorm** legt den Fokus auf den individuellen Lernverlauf und dokumentiert die individuellen Lernwege und -fortschritte. Sie steht bei der individuellen Förderung im Vordergrund. Mit diesem Bezugssystem werden die individuellen Leistungen von Lernenden mit einem früheren Leistungsstand verglichen.
- Im System der **Sozialnorm** werden die einzelnen Leistungen der Lernenden einer Gruppe miteinander verglichen. Dieser Vergleich kann schweizweit mit standardisierten Instrumenten geschehen, wie dies bspw. bei der Überprüfung der Grundkompetenzen des Lehrplans 21 geschieht. Diese standardisierten Leistungsmessungen gehören nicht auf die Unterrichtsebene, da sie primär zur Überprüfung des gesamten Bildungssystems eingesetzt werden, wie Koch et al. (2017, S. 6) in der Handreichung «Kompetenzorientiert beurteilen» betonen.

Bei der Beurteilung im Klassenvergleich sollte die Sozialnorm mit Vorsicht angelegt werden, da die Klassenzusammensetzung zwischen einzelnen Klassen stark variieren kann (vgl. dazu Kap. 4.3). Zudem ist das Leistungsbild innerhalb einer heterogenen Klasse relativ stabil, weshalb individuelle Lernfortschritte, die alle in der Klasse machen, nicht sichtbar werden (vgl. Rheinberg, 2006, S. 57).

Beziehen sich die Lehrpersonen in einem Fach auf die Individualnorm, hat dies im Vergleich zur Sozialnorm positivere Auswirkungen auf die Motivation der Schüler und Schülerinnen in diesem Fach, insbesondere auf das Selbstkonzept und die Selbstwirksamkeitserwartungen. Dies zeigen Studien vor allem in den Fächern Mathematik und Englisch (vgl. Rheinberg, 2006). Eine für Deutschland repräsentative Studie von Köller (2004) im Fach Mathematik zeigt auf, dass diese motivationalen Effekte tatsächlich eintreten, und zwar als Klasseneffekt.

4.2 Formatives Beurteilen

Die Individualnorm ist der Bezugsrahmen für formatives Beurteilen und zeigt – wie in Kap. 2 ausgeführt – über verschiedene Fächer hinweg einen grossen positiven Effekt auf die Lernleistungen (Hattie, 2009). Zudem kann formatives Beurteilen und Rückmelden – Glogger-Frey & Herppich (2017) sprechen von formativer Diagnostik – insbesondere zur Individualisierung sowie für eine Anpassung des Unterrichts genutzt werden. Der hohe positive Effekt kann damit erklärt werden, dass formative Diagnostik der Lernförderung dient, während summatives Beurteilen (siehe Kap. 4.3), das Hattie (2009) zufolge geringe Effekte auf die Lernleistungen zeigt, auf eine abschliessende Bewertung eines bestimmten Bereichs abzielt und damit von den Lernenden auch anders wahrgenommen wird.

Metaanalysen und besonders Meta-Metaanalysen fassen notwendigerweise ganz unterschiedliche Ansätze zusammen. Umso wichtiger ist es, formatives Beurteilen konzeptionell zu klären und aufzuzeigen, wie verschiedene Komponenten zusammenspielen. Ein solches Rahmenkonzept zum formativen Beurteilen präsentieren Black & Wiliam (2009). Dabei unterscheiden sie fünf Aktivitäten:

- 1) **Klärung der Lernziele und der Erfüllungskriterien**, die für eine gute Leistung stehen. Diese Ziele müssen auf allen drei Ebenen klar sein, der Ebene der Lehrperson, der Ebene der Peers (z.B. werden sie unter den Schülern und Schülerinnen diskutiert) und auf der Ebene der einzelnen Lernenden (sie werden von jedem einzelnen Schüler/jeder einzelnen Schülerin verstanden). Dies bedingt, dass die Lehrperson immer wieder das Gespräch mit den Lernenden sucht, um zu überprüfen, ob die vorgegebenen mit den von den Lernenden selbst gesteckten Zielen übereinstimmen (vgl. auch Kap. 3.1).
- 2) Es müssen **wirksame Lerngelegenheiten entwickelt und zusammengestellt** werden, die das beabsichtigte Ziel fokussiert und das Lernen bzw. die Lernergebnisse sichtbar machen.
- 3) Die Lernenden erhalten ein **Feedback, das sie in ihrem Lernprozess weiterbringt** und konkret aufzeigt, wie sie das Ziel erreichen können. Bei Fremdbeurteilungen durch die Lehrperson geschieht dies in einem Feedbackgespräch, in dem Ziele, Kriterien und mögliche weitere Lernschritte zwischen den Lernenden und der Lehrperson ausgetauscht und bei Bedarf auch ausgehandelt werden.
- 4) Der Unterricht aktiviert die Lernenden, sich gegenseitig beim Lernen zu unterstützen und sich ein **Peerfeedback** zu geben.

- 5) Der Unterricht aktiviert die Lernenden zur **Selbstbeurteilung und zum selbstregulierten Lernen**. Das Feedback der Lehrperson bzw. der Peers ist darauf ausgerichtet, eigene Selbstbeurteilungsprozesse anzuregen und diese zu reflektieren.

Diese Aktivitäten lassen sich einerseits auf die Ebenen Lehrperson, Peers und einzelne Lernende, andererseits auf die Leitfragen des formativen Beurteilens beziehen, wie sie auch Hattie & Timperley (2007) als Strukturierung eines wirkungsvollen Feedbacks verwenden (vgl. Tabelle 1, S. 20).

Bennett (2011) verweist darauf, wie wichtig eine solch präzisere Fassung ist, um die positiven Effekte maximieren zu können. Zudem betont er, dass Lehrpersonen für das formative Beurteilen ein detailliertes Domänenwissen benötigen, also eine genaue Vorstellung davon, wie Lesen, Schreiben etc. erworben werden, welche Förderansätze wirksam sind. Kingston & Nash (2011) bestätigen dies und zeigen in ihrer Analyse zum formativen Beurteilen auf, dass vor allem die Domänen (neben Klassenstufe und der entsprechenden Beurteilungsmodalität) als Moderatoren des formativen Beurteilens wirksam sind.

Im Rahmen eines solchen Konzepts der formativen Beurteilung ist die Diagnostik auf die individuelle Förderung der Schülerinnen und Schüler ausgerichtet und entspricht dem zyklischen und dialogischen Verständnis des Selbst- bzw. Fremdbeurteilungskreislaufs von Senn (2018).

Koch et al. (2017, S. 6) postulieren in der Handreichung «Kompetenzorientiert beurteilen», dass das formative Beurteilen im Zentrum eines lernförderlichen Unterrichts stehen soll, weil es mit Blick auf die Ziele aus dem Lehrplan 21 und der individuellen Lernvoraussetzungen der einzelnen Schülerinnen und Schüler deren Lernen unterstützt. Dies entspricht auch den Befunden von Brühwiler

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1.) Clarifying learning intentions and criteria for success	2.) Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	3.) Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4.) Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	5.) Activating students as the owner of their own learning	

Tabelle 1: Aspekte der formativen Beurteilung (Black & Wiliam, 2009, S. 8)

(2017), die ebenfalls zeigen, dass die Berücksichtigung des Vorwissens der Schüler/-innen im Unterricht (didaktische Handlungskompetenz) positive Effekte auf die Leistung der Schüler/-innen haben.

4.3 Summativ beurteilen

Um zu entscheiden, ob eine Schülerin oder ein Schüler nach einer längeren Förderphase beispielsweise die Grundansprüche oder die Orientierungspunkte einer bestimmten Kompetenz im Lehrplan erreicht hat, müssen die Lernergebnisse abschliessend ermittelt werden. Dazu braucht es eine *summative Beurteilung*, die am Ende einer Förderung Bilanz zieht und den Lernstand der Schülerin oder des Schülers zu diesem Zeitpunkt ermittelt. Summative Beurteilung ist im schulischen Kontext auch für die Notensetzung im Zeugnis wichtig.

Die Handreichung «Kompetenzorientiert beurteilen» (Koch et al., 2017, S. 7) zeigt auf, wie die Sachnorm, ausgehend von den Kompetenzbeschreibungen des Lehrplans 21 und die davon abgeleiteten Lernziele als Grundlage für die Notensetzung beigezogen werden können. Werden Noten auf diese Art gesetzt, zeigen sie auf, auf welchem Niveau die Kompetenzen eines Fachs erreicht worden sind. Wie die Handreichung zur Beurteilung der Bildungsdirektion Kanton Zürich (2018, S. 19)

zudem hervorhebt, werden Noten nur im Zusammenhang mit dem Zeugnis verlangt und stellen in diesem Sinne die Gesamtleistung eines Schülers einer Schülerin bezogen auf die Vorgaben im Lehrplan 21 dar:

Sie geben Auskunft darüber, in welchem Grad eine Schülerin, ein Schüler in einem bestimmten Fachbereich die angestrebten Lernziele während der Zeugnisperiode erreicht hat. (Ebd., S. 18)

Sie ist nicht das Ergebnis einer Durchschnittsrechnung von gesammelten schriftlichen Prüfungen, sondern beruht auf einem professionellen Ermessensentscheid der Lehrperson. (Ebd., S. 19)

Oft wird in der Schule für summative Beurteilungen in Zusammenhang mit Noten jedoch der Gruppenvergleich in der Klasse herangezogen, bei dem die Sozialnorm im Vordergrund steht. Obwohl die Sozialnorm generell geringere positive Effekte auf die Lernleistung der Schüler/-innen hat, hat sie eine grosse Bedeutung für den Schulalltag, insbesondere für die Notensetzung, da Noten als «... ökonomisch, überschaubar und verständlich» (Tent, 2006, S. 879) gelten.

Zeinz & Köller (2006) untersuchten beim Beurteilen den Einfluss der Referenzgruppe auf das Selbstkonzept von 2.-Klässler/-innen, und zwar im Fach Mathematik. Sie stellen fest, dass die Einführung der Zeugnisnote den Fischteich-Effekt verstärkt. Das bedeutet, dass der Vergleich mit den Leistungen ande-

rer Schüler/-innen das mathematische Selbstkonzept beeinflusst. Konkret: Eine Schülerleistung ist abhängig von der durchschnittlichen Leistung einer Klasse. Entsprechend wird die Leistung eines guten Schülers in einer durchschnittlich eher schwächeren Klasse besser eingeschätzt als in einer überdurchschnittlichen Klasse. Des Weiteren zeigen sie auch auf, dass mit der Zeugnisnote der soziale Vergleich zunimmt und sich damit der Fischteich-Effekt verstärkt (Zeinz & Köller, 2006, S. 186).

Zusammengefasst: Bei einer summativen Beurteilung hat die Lehrperson weniger die Frage vor Augen, welche passende Fördermassnahme als Nächstes den Lernenden angeboten werden soll, sondern sie möchte rückblickend zu dem abschliessenden Entscheid gelangen, inwiefern eine im Lehrplan verlangte Kompetenzstufe erreicht worden ist. Während der Lehrplan die Sachnorm vorgibt, ist im Unterricht im Zusammenhang mit der Notensetzung oftmals die Sozialnorm leitend, was einen Fischteich-Effekt zur Folge hat.

Solche abschliessenden Beurteilungen – vor allem in Form von Noten – können weitreichende Konsequenzen für die SuS haben, gerade auch, wenn sie wie bei Übertritten mit einer weiteren Beurteilungsart kombiniert wird: der prognostischen Beurteilung.

4.4 Prognostisch beurteilen

Gewisse Entscheide in der Schullaufbahn von Schülerinnen und Schülern haben eine besonders grosse Tragweite, da sie nicht nur rückwirkend bilanzieren, wie dies in summativen Beurteilungssituationen geschieht, sondern zudem noch eine Prognose darstellen und damit den zukünftigen Weg der Lernenden bestimmen (Bildungsdirektion Kanton Zürich, 2018). Sie finden zu bestimmten vordefinierten Zeitpunkten statt, wie beispielsweise zum Schuljahresende oder beim Wechsel von der Primar- zur Sekundarstufe. Solche

Promotionsentscheide verlangen eine *prognostische Beurteilung*. Die meisten der im schulischen Kontext gefällten Prognosen basieren auf Noten. Obwohl Noten in Bezug auf ihre Objektivität in Verruf geraten sind, nicht zuletzt bei der Beurteilung von Texten, die nach wie vor in Bezug auf Objektivität grosse Probleme bereitet, wie Birkel (2003) in seiner Replikationsstudie aufzeigt, stellen Schulnoten Prädiktoren weiterer Bildung dar, wie Schuler (2006, S. 540) in seiner Analyse vor allem von deutschen und amerikanischen Studien aufzeigt. Die prognostische Qualität ist vor allem für die weitere Bildung recht gut, während sie sich in geringerem Masse für die Vorhersage des Berufserfolgs eignet. So beträgt die mittlere prognostische Validität der Abiturnote in Bezug auf Studienleistungen $r = .46$ (vgl. Schuler, 2006, S. 537), die Vorhersage des Berufserfolgs (aufgrund von amerikanischen Studien, da die deutsche Datenglage bescheiden ist) $r = .30$, mit abnehmender Tendenz bei neueren Studien (vgl. Schuler, 2006, S. 540).

Roos, Wandeler & Mosimann (2013, S. 16f.) untersuchen in ihrem Schlussbericht zur externen Evaluation des Luzerner Übertrittsverfahrens von der Primarschule auf die Sekundarstufe I die prognostische Validität von Noten in einzelnen Kantonen der Schweiz. Sie weisen darauf hin, dass Noten zwar statistisch auf der Ebene des Gesamtsystems gut abschneiden (im Kanton Bern erklären Noten in Deutsch und Mathematik 52 % der Varianz der unterschiedlichen Zuweisung auf die Sekundarstufe I; im Kanton Zürich sind es sogar 66 %). Moser & Rhyn (2000) fanden dabei für den Kanton Zürich am Ende der Primarstufe eine Korrelation zwischen Schulnoten und Testleistung von $r = .74$ (Deutsch: $r = .67$; Mathematik: $r = .69$), was auf einen engen Zusammenhang hinweist – und dies obwohl in den Tests nur einzelne Aspekte der fachlichen Kompetenzbereiche getestet wurden (bspw. keine produktiven Leistungen im Fach

Deutsch). Dieser Zusammenhang stimmt zudem mit den internationalen Befunden überein.

Richtet man den Blick jedoch auf die prognostische Aussagekraft von Noten für ein einzelnes Kind, kann sich diese Interpretation entscheidend ändern. Noten und Testleistung können im Einzelfall stark variieren, wie Moser, Angelone, Hollenweger & Buff (2011, S. 63ff.) für den Kanton Zürich differenziert nachweisen. So zeigen Moser et al. (2011, S. 73) bezogen auf eine bestimmte Testleistung, dass die Notenbandbreite doch beachtlich ist (z.B. streuen die Noten für Mathematik bei 900 Punkten Testleistung von 3.5 bis 6; im Fach Deutsch bei 800 Punkten zwischen 4 und 5.5), was dann auch unterschiedliche Schulzuweisungsentscheide zur Folge haben kann. Insofern relativiert sich die prognostische Validität auf der Ebene der einzelnen Schülerinnen und Schüler deutlich. Die Autoren führen dies darauf zurück, dass es, obwohl die Schulnote sich vorwiegend auf die fachliche Leistung abstützt, zu Verzerrungen in Bezug auf soziale und sprachliche Faktoren wie auch auf Gruppeneffekte kommt, da die Referenzgruppe bei der Notensetzung grundsätzlich die Klasse ist (Moser et al., 2011, S. 63). So gibt es stärkere Prädiktoren als Noten: Die soziale Herkunft selbst – aber nicht Deutsch als Zweitsprache – zeigt die stärkste Vorhersagekraft (Moser, Oostlander & Tomasik, 2017, S. 71).

Wegen ihrer grossen Tragweite müssen die Instrumente für die prognostische Beurteilung besonders sorgfältig ausgewählt werden und sehr valide sein (vgl. Kap. 5.1). Wie mit Blick auf die fachliche Leistung in Mathematik oder Naturwissenschaft in Kapitel 6 dargelegt wird, übt die Sprachfähigkeit der Schülerinnen und Schüler – das betrifft sowohl mono- als auch multilinguale Schüler/-innen – einen starken Einfluss auf deren fachliche Leistungen aus. Das heisst, in den meisten fachspezifischen Leistungsüberprüfungen werden immer auch die Sprachfähigkeiten der Schüler/

-innen überprüft. Bei prognostischen Entscheidungen muss deshalb in der Gesamtbeurteilung der Einfluss der sprachlichen Aspekte entsprechend berücksichtigt werden (vgl. Bildungsdirektion Kanton Zürich, 2013, S. 5).

Verschiedene Studien belegen den Einfluss von Eltern auf Motivation, Leistung und Verhalten der Schüler/-innen (Neuenschwander & Nägele, 2017, S. 14). Zudem wurde der Einfluss der Eltern auf Übertrittsentscheidungen bzw. auf entsprechende Empfehlungen durch die Lehrpersonen breit nachgewiesen (Roos, Wandeler & Mosimann, 2013). Es reicht entsprechend nicht aus, dass Lehrpersonen geeignete prognostische Instrumente einsetzen, sondern dass sie sich auch dieses Einflusses durch die Eltern bewusst sind. Dies wäre in Weiterbildungen zu berücksichtigen.

Zusammenfassend in Bezug auf die verschiedenen Beurteilungsarten lässt sich feststellen, dass die formative Beurteilung, insbesondere, wenn sie darauf ausgerichtet ist, die Schüler/-innen zur Selbstbeurteilung anzuleiten, eine sehr wirkungsvolle unterrichtsgestaltende Massnahme darstellt. Sie steht deshalb im Zentrum eines unterrichtlichen Handelns, das entsprechend der Individualnorm die individuelle Förderung des Lernens fokussiert. In der formativen Diagnostik steht das Feedback im Zentrum, und im Sinne einer Anleitung zur Selbstbeurteilung auch das Peer-Feedback. In summativen Beurteilungssituationen wird Bilanz darüber gezogen, inwiefern bestimmte, aus dem Lehrplan von Kompetenzstufenformulierungen abgeleitete Lernziele erreicht worden sind. Zu bestimmten Zeitpunkten in der Schullaufbahn erhalten diese summativen Beurteilungen zusätzlich eine prognostische Funktion. Dann gilt es abzuschätzen, inwiefern für einen Schüler oder eine Schülerin die Voraussetzungen für das erfolgreiche Absolvieren einer weiteren Phase gegeben sind.

Formative und summative Beurteilung ergänzen sich im unterrichtlichen Handeln, sie sollten jedoch wegen ihrer unterschiedlichen

Funktion nicht in derselben Beurteilungssituation eingesetzt werden. Weinert (1999, S. 33) unterscheidet in diesem Zusammenhang zwischen Leistungs- und Lernsituationen, die aufgrund der unterschiedlichen psychologischen Gesetzmässigkeiten voneinander getrennt werden sollten. Summative (und prognostische) Beurteilungssituationen mit ihrer grossen Bedeutung für das weitere Lernen stellen dabei nach Weinert (1999) Leistungssituationen dar, in denen Schüler/-innen sinnvollerweise Fehler möglichst vermeiden,

um das Risiko eines Misserfolgs zu minimieren, während formative Beurteilungssituationen mit ihrer Ausrichtung auf die Lernförderung Lernsituationen darstellen, in denen Fehler einen Erkenntnisgewinn haben, als wertvolle Information für das weitere Lernen dienen. In Bezug auf Weiterbildungen von Lehrpersonen sollten deshalb die Unterscheidung der Beurteilungsarten eine wichtige Rolle spielen, inklusive der zu diesen Beurteilungssituationen passenden Beobachtungs- und Beurteilungsinstrumente.

5 Beobachtungs- und Beurteilungsinstrumente

Die Verfahren oder Instrumente, die in Beobachtungs- und Beurteilungssituationen eingesetzt werden, müssen bestimmte Kriterien erfüllen. Dabei ist es zentral, zwischen Verfahren und Instrumenten zu unterscheiden, die im unterrichtlichen Rahmen eingesetzt werden, und solchen, die für Testzwecke in sogenannten *Large-Scale-Assessments* eingesetzt werden (wie bspw. bei den Schweizer Bildungsmonitoring-Projekten der EDK HarmoS Bildungsstandards oder Überprüfung der Grundkompetenzen [ÜGK]).

5.1 Gütekriterien von Testinstrumenten

Testinstrumente müssen ganz spezifischen Anforderungen genügen, da sie einen hohen Standardisierungsgrad aufweisen. Zudem werden sie basierend auf Groeben (2009) in systematischen Beobachtungssituationen eingesetzt. Als Gütekriterien müssen Testinstrumente folgende drei Merkmale aufweisen:

- *Objektivität* bezeichnet den Grad, in dem ein Test unabhängig von der untersuchten Person ist, und zwar hinsichtlich der Durchführung, der Auswertung und der Interpretation.
- *Reliabilität* bezieht sich auf die Genauigkeit der Beurteilung und des eingesetzten Beurteilungsinstruments.
- *Validität* bezeichnet das Ausmass dessen, was der Test zu messen vorgibt, sei es in Bezug auf den Inhalt oder den Zusammenhang der Testleistung mit einem oder auch mehreren Kriterien.

Objektivität und Reliabilität sind aus messtheoretischer Sicht grundlegende Kriterien, Validität ist aus fachdidaktischer Sicht zentral, da es die inhaltliche Seite von Tests bestimmt. Zusätzlich zu diesen Kriterien werden oft auch weitere Kriterien wie Normierung,

Vergleichbarkeit, Ökonomie – z. B. hinsichtlich der Durchführungszeit – oder Nützlichkeit herangezogen (Bühner, 2004). Gelten diese Kriterien als Richtlinie für die Erstellung von Testinstrumenten im Rahmen eines Forschungskontexts, verlangt das schulische Umfeld einen pragmatischen Umgang mit diesen Kriterien.

An die schulische Beurteilung im normalen Schulalltag dürfen nicht unverhältnismässige Ansprüche gestellt werden, da die primäre Aufgabe von Lehrpersonen «die Initiierung, Unterstützung und Lenkung von Lernprozessen der Lernenden» sei (Groeben, 2009, S. 111), was mit alltagsnäheren Beurteilungsformen verbunden ist (vgl. Kap. 2). Während in einer Testsituation die Durchführung eines Tests stark normiert ist, weshalb sie auch grossen Einschränkungen unterliegt, bedeutet eine Vielfalt in der Durchführung von Unterricht (wie Lerntests, Prüfungen, Lernaufgaben, Gruppenarbeiten, Projektarbeit, Beteiligung am Unterricht oder auch Hausaufgaben) eine grosse Unterrichtsqualität, wie Helmke (2009) mit dem Hinweis auf Methodenvielfalt feststellt.

Die vielfältige Unterrichtspraxis generiert zudem sehr unterschiedliche Beobachtungssituationen für die schulische Beurteilung (Jussim et al., 2009, S. 362), in denen die genannten Kriterien nicht nur in unterschiedlichem Masse zum Zuge kommen, sondern auch je nach Unterrichtsform unterschiedlich interpretiert werden müssen. Das Kriterium der Objektivität beispielsweise dürfte bei der Beurteilung sprachlicher Leistungen generell nur schwer umsetzbar sein. Wie bspw. Birkel (2009) mit Blick auf Diktate, die oft als objektive Form der Überprüfung von Rechtschreibleistung angesehen werden, zeigen kann, erfüllen sie diesen Anspruch nicht, da unter anderem die Zahl der identifizierten Fehler von Lehrperson zu Lehrperson stark variieren kann. Ähnlich können Ehl et al.

(2018) zeigen, dass die Durchführungs- und Auswertungsobjektivität für die «Profilanalyse nach Griesshaber» zu gering ausfällt.

Dies ist jedoch nicht so zu verstehen, dass die Gütekriterien Objektivität, Reliabilität oder Validität überhaupt keine Rolle im schulischen Kontext mehr spielen sollten. Schrader (2013, S. 155) unterscheidet diesbezüglich die beiden Ebenen Mikro- und Makroadaptationen von diagnostischen Handlungen. Auf der Mikroebene der Lehr-Lern-Prozesse basieren die diagnostischen Informationen zwangsläufig auf Unterrichtsbeobachtungen und subjektiven Einschätzungen. Auf der Makroebene hingegen besteht oftmals die Möglichkeit, gezieltere Abklärungen zu treffen (z.B. formelle Diagnostik), da auf dieser Ebene längerfristige, auf grössere Unterrichtseinheiten bezogene Entscheide im Zentrum stehen.

Aber auch auf der Mikroebene leisten präzise erarbeitete Beurteilungsinstrumente, die auf die entsprechende sprachliche Handlung abgestimmt sind, einen Beitrag in diese Richtung. Praktikable Kriterienraster für mündliche Leistungen etwa müssen zwar einfach handhabbar sein, also für die schulische Praxis aus wenigen Kriterien bestehen, bei der Beschreibung von Kriterien und Indikatoren hingegen sollte darauf geachtet werden, dass diese eine ausgewogene Mischung aus eindeutig beobachtbaren Oberflächenmerkmalen (z.B. Blickkontakt zur Zuhörerschaft) und weniger eindeutig beobachtbaren Tiefenmerkmalen (z.B. Verständlichkeit oder klare inhaltliche Struktur) bestehen. Vor allem diese Tiefenmerkmale bewirken zwar, dass der Interpretationsspielraum bei diesen Kriterien grösser ist, was die Messgenauigkeit des Instruments und die Objektivität der Auswertung beeinträchtigt. Diese Kriterien betreffen jedoch die wesentlichen Aspekte der

Sprachhandlung, weshalb es sich um validere Beobachtungspunkte handelt.

5.2 Gütekriterien von schulischen Beurteilungsinstrumenten

In eine schulische Beurteilung fliessen Daten, die meist auf Basis teilsystematischer Beobachtung erhoben wurden, manchmal handelt es sich sogar um Daten aus unsystematischen Beobachtungen. Ein wesentliches Ziel schulischer Beurteilung muss deshalb sein, dass der Beurteilungsvorgang so gestaltet wird, dass er intersubjektiv nachvollziehbar ist. Bohl (2001) nennt dies **kommunikative Validierung**. Er stellt fest, dass insbesondere die Erweiterung der Lehr- und Lernformen zu einem erhöhten Einsatz von offenen Lernaufgaben führt, die auch «neue Beurteilungsformen» nach sich ziehen, die insbesondere mit einer strengen Auslegung des Gütekriteriums Objektivität in Konflikt geraten (vgl. Bohl, 2001, S. 38ff.).³ Er schlägt deshalb vor, dass solche Beurteilungssituationen einer **kontrollierten Subjektivität** mit Kern einer kommunikativen Validierung als Gütekriterium unterliegen sollen.

Kommunikative Validierung zeichnet sich primär durch den kommunikativen Charakter aus. Durch diese kommunikative Ausrichtung wird die Beurteilungssituation einerseits zum Dialog zwischen Lehrperson und Lernenden. Die Lehrperson versucht in diesem Gespräch eine möglichst hohe Transparenz über Beurteilungsgrundlagen und Entscheid zu erreichen. Gleichzeitig soll dieses Gespräch zu einem gemeinsamen Verständnis der Kriterien führen und die Selbsteinschätzung der Lernenden einbeziehen. Andererseits findet der Validierungsprozess auch im Team der Lehrpersonen einer Schule statt, wo Beurteilungsaufgabe, -verfahren und -kriterien (inkl. deren Gewichtung) besprochen werden, um zu

³ Handelt es sich dabei um offene Formen wie Lernjournale oder mündliche Präsentationen, können dabei aber auch grössere Herausforderungen für Schülerinnen und Schüler aus bildungsfernen Familien wie auch für solche mit Deutsch als Zweitsprache einhergehen. Dies wäre bereits bei der Einführung solcher Verfahren zu berücksichtigen.

gemeinsamen Zielvorstellungen zu gelangen. Bei der sprachlichen Produktion beispielsweise ist ein wesentlicher Schritt dieses Validierungsprozesses der Austausch über die Beurteilungen der Schülerleistungen und damit das Gespräch über die Einschätzung schriftlicher Texte oder videografiertes mündlicher Leistungen der Schülerinnen und Schüler. Im Validierungsprozess ist auch zentral, dass sich die Lehrpersonen darauf einigen, welches Bezugsnormensystem sie bei der Beurteilung anlegen: die *Sachnorm*, die Kriterien aus den Kompetenzbeschreibungen des Lehrplans ableitet, die *Individualnorm*, die den Fokus auf den individuellen Lernverlauf legt, oder die *Sozialnorm*, nach der die einzelnen Leistungen der Lernenden einer Gruppe miteinander verglichen werden (vgl. Kap. 4.3).

In einem qualitativ hochstehenden Unterrichtsgeschehen mit vielfältigen Lehr- und Lernformen braucht es also auch offene Beurteilungsformen, bei denen die Kriterien Objektivität, Reliabilität und Validität nur in Ansätzen umsetzbar sind. Kommunikative Validierung stellt demzufolge den Versuch dar, den Beobachtungs- und Beurteilungsprozess einerseits für die Lernenden, andererseits für Aussenstehende (Eltern, andere Lehrpersonen) intersubjektiv nachvollziehbar zu gestalten.

5.3 Aufgaben zur sprachlichen Produktion und Rezeption

Kommunikative Validierung ist auch bei der Beurteilung sprachlicher Leistungen wichtig, die gerade wegen ihrer Komplexität häufig den Ansprüchen der Testkriterien nicht genügen können. Diese bezieht sich auch auf den Austausch im Team über die zur Beurteilung eingesetzten Aufgaben, da diese Aufgaben oft den Kern der eingesetzten Beurteilungsinstrumente darstellen.

Dabei ist zwischen Aufgaben zur Sprachproduktion und zur Sprachrezeption zu unter-

scheiden (vgl. Senn, 2009), da diese unterschiedliche Anforderungen stellen. Auch der Lehrplan 21, Fach Deutsch, unterscheidet zwischen den rezeptiven und den produktiven Sprachhandlungen. Sprachproduktive Aufgaben verlangen grundsätzlich ein sprachliches Produkt, dessen Qualitäten dann beurteilt werden, während sprachrezeptive Aufgaben von einem solchen Produkt ausgehen und das Verständnis der Schülerinnen und Schüler von diesem Produkt überprüfen.

Bei **Aufgaben zur sprachlichen Rezeption** ist der Schwierigkeitsgrad und die Komplexität der Aufgabe weitgehend durch die Aufgabenstellung vorbestimmt, womit auch die Lösungserwartungen grösstenteils geklärt sind. Sie setzt sich aus der Verständlichkeit des Primärtexts sowie der Art und Komplexität der gestellten Fragen an den Text zusammen. Eine solche Aufgabe stellt also gewissermassen eine doppelte Rezeptionsleistung dar, diejenige des Primärtexts sowie das Verstehen der Fragen. Die Verständlichkeit des Primärtexts hängt von verschiedenen Faktoren ab: Bekanntheit und Komplexität des Themas, Strukturiertheit, grafische Gestaltung, Informationsdichte, Sprache, Implizitheit bzw. Explizitheit bspw. in der Leserführung, Illustrationen, Bilder, Situierung und unterstützende Strukturierungshilfen etc. Die Verständlichkeit der Fragen wird durch deren Komplexität, deren sprachliche Ausformulierung, der Art des Frageformats und bei Multiple-Choice-Fragen der Komplexität und der Sprache der Antwortvorgaben bestimmt. Gerade für Schüler/-innen mit anderssprachlichem Hintergrund kann nicht nur der Primärtext, sondern auch die Fragestellung, der Auftrag selbst eine grosse sprachliche Herausforderung darstellen. Der Lehrplan 21 gibt wenig Hinweise zur Art des Primärtexts (er unterscheidet jedoch zwischen literarischen und Sachtexten und nennt einzelne Genres und Textsorten). Zur Art der Fragestellung macht er bei der Kompetenz «Verstehen von Sachtexten» (D.2.B) Vorgaben zur Abfrage von expliziten bzw. impliziten Informationen

aus dem Text und verlangt Fragen, die das lokale oder globale Textverständnis betreffen bzw. einen Bezug zur eigenen Lebenswelt herstellen wollen. Diese Art der Strukturierung der Verständnisfragen findet sich bspw. im Lehrmittel *Die Sprachstarken*. Sie basiert grundsätzlich auf dem Lesekompetenzmodell, das zur Überprüfung der Lesekompetenz bei PISA verwendet wird (vgl. Naumann, Artelt, Schneider & Stanat, 2010). Der Hörprozess im monologischen Bereich ist auf dieselbe Art strukturiert. Oft haben Aufgaben zur Sprachrezeption auch offene Formate, die wiederum zur Lösung eine Sprachproduktion verlangen, z. B. eine Zusammenfassung des Inhalts oder eine Gewichtung der dargestellten Punkte. In solchen Mischformen können die Bedingungen der Sprachproduktion die Überprüfung der rezeptiven Fähigkeiten mit beeinflussen.

Bei **Aufgaben zur Sprachproduktion** ist die Lösungserwartung meist nur punktuell in der Aufgabenstellung geklärt. Ein wesentlicher Teil des Schwierigkeitsgrads wird dadurch bestimmt, wie strukturiert die Aufgabe ist und welche Informationen zum Schreiben/Sprechen und zum herzustellenden (mündlichen oder schriftlichen) Text gegeben sind. Hochstrukturierte Aufgaben beinhalten eine kontextuelle Situierung (Adressat/-innen, Handlungskontext, Handlungsziel, Angaben zum Textmuster, Mustertexte, Textprozeduren), einen klaren Auftrag mit Anleitungen und Lösungserwartungen in Form von Textkriterien (vgl. Senn, 2009, S. 83). Zudem enthalten sie Hilfen, die die Schüler/-innen im Textproduktionsprozess unterstützen oder die sie zur Erarbeitung von notwendigem Hintergrundwissen anleiten (Sturm & Weder, 2016, S. 118). Die Aufgabe strukturiert damit einerseits den Sprachproduktionsprozess, indem dieser in überschaubare Schritte gegliedert wird, die als Strategien gebündelt werden können. Sie

gibt zusätzlich Hinweise auf das herzustellende Sprachprodukt. Hochstrukturierte Aufgaben verlangen jedoch auch eine nicht unbeträchtliche Leseleistung, was vor allem im Kontext von DaZ-Schülern und Schülerinnen sowie von Schülern und Schülerinnen mit eher geringem (bildungssprachlichem) Wortschatz zu bedenken ist.

Diese Prozess- und Produktorientierung zeigt sich auch im Lehrplan 21, beim Schreiben, indem Schreibprodukt (D.4.B) und Schreibprozesse (D.4.C, D, E, F) als eigene Kompetenzen ausformuliert wurden, und bei der mündlichen Sprachproduktion, indem nicht nur die Ergebnisse, sondern auch der Erarbeitungsprozess dargestellt wird (z. B. beim «Präsentieren» [D.3.B.1.g]: Inhalte recherchieren, auswählen, strukturieren und das Wissen präsentieren). Als Techniken oder als Anleitungen finden sich diese auch in den Lehrmitteln (z. B. im Werkbuch des Lehrmittels *Sprachwelt Deutsch* oder auf den Karteikarten des Lehrmittels *Die Sprachstarken*). Die Wirksamkeit von expliziter Strategievermittlung ist im Bereich Schreiben empirisch überprüft – v. a. das sogenannte SRSD-Modell (SRSD = selfregulated strategy development) und zeigt eine grosse Wirkung auf die Schreibleistungen, wie etwa Graham & Harris (2012) mit $d=1.20$ oder auch Koster et al. (2015) mit $d=.96$ aufzeigen.⁴

Es sei betont, dass insbesondere Reflexion als eigene Kompetenz – beim Schreiben (D.4.F) etwa beinhaltet dies das Nachdenken über ihre Schreibprodukte wie auch über ihren Schreibprozess – für Schüler/-innen mit Deutsch als Zweitsprache und vor allem auch für Schüler/-innen aus bildungsfernen Familien eine grosse Herausforderung darstellt, da dies das Verfügen über bestimmten bildungssprachlichen Wortschatz voraussetzt. Ansätze wie die erwähnte explizite Vermittlung von Lese- und Schreibstrategien vermitteln im Rahmen des Modellierens – es handelt

⁴ Ähnliches gilt für das Lesen als rezeptiver Sprachhandlung (vgl. Philipp, 2017).

sich dabei um eine didaktisierte Form des lauten Denkens – den für die Reflexion notwendigen Wortschatz. Nicht zuletzt werden im Rahmen dieses Ansatzes die Schüler/-innen immer wieder aufgefordert, die zentralen Schritte zu memorieren und dabei zu versprachlichen. Dies gilt es auch im Rahmen einer formativen Beobachtung und Beurteilung zu berücksichtigen.

Formative Beurteilung verlangt eine genaue Prozessbeobachtung, damit das fördernde Feedback an der richtigen Stelle ansetzen kann (z. B. können die Grundfertigkeiten der einzelnen Sprachhandlungen automatisiert genug eingesetzt werden? Welche Strategien und Techniken werden angemessen eingesetzt? Wo treten Probleme auf?). Dies ist bei längeren Sprachproduktionsprozessen leichter beobachtbar als bei den Rezeptionsprozessen. Bei beiden gilt jedoch, dass durch einen Austausch oder ein kooperatives Vorgehen, das eine Verbalisierung der Teilprozesse verlangt, solche kognitiven Prozesse beobachtbar werden und damit auch reflektierbar. Diese Sprachreflexion der einzelnen Sprachhandlungen – aber auch die Reflexion in den andern Fächern – wird vom Lehrplan 21 verlangt und durch die entsprechenden Kompetenzbeschreibungen sichtbar gemacht. In den neueren Sprachlehrmitteln wie auch in den Lehrmitteln anderer Fächer werden reflexive Fähigkeiten ebenfalls oft verlangt. Für Schüler/-innen mit bildungsfernem Hintergrund kann diese Ausrichtung auf Reflexion mit einer spezifischen Herausforderung einhergehen, da ihnen oft die sprachlichen Mittel zur Verbalisierung fehlen. So

sollten gerade bei Peer-Feedback-Situationen oder Selbstbeurteilungen entsprechende Formulierungen angeboten werden. Hier muss eine Weiterbildung einen entsprechend starken Akzent setzen.

Die Analyse dieser Aufgabenformate ist auch deshalb wichtig, weil sie in den meisten Fächern auf die eine oder andere Art ebenfalls zur Überprüfung der Fachinhalte eingesetzt werden. Wird im Fach Geschichte eine Präsentation verlangt, in einem naturwissenschaftlichen Fach das Verständnis eines Schul- oder Sachbuchtexzes überprüft, wird eine Zusammenfassung (mündlich oder schriftlich), ein Protokoll verlangt oder wird ein Feedback-Gespräch eingesetzt, werden immer auch sprachliche Anforderungen – sei es im rezeptiven oder im produktiven Teil – verlangt. In der Regel werden diese sprachlichen Anforderungen jedoch nicht als Teil der Aufgabe aufgefasst, da diese ja die entsprechende Fachausrichtung hat (vgl. Kap. 6).

Zusammenfassend lässt sich feststellen, dass sprachrezeptive und sprachproduktive Aufgaben (bzw. Mischformen davon) in vielen schulischen Kontexten zur Überprüfung von schulischen Leistungen eingesetzt werden. Insbesondere offene Lehr- und Lernformen befördern offene Beurteilungsformen (wie Beurteilungsgespräche, Feedbackgespräche, Portfolio, Lernjournale), die eine höhere Versprachlichung verlangen. Es ist deshalb wichtig, die (sprachlichen) Anforderungen dieser Formate bei der Beurteilung zu reflektieren und als Aufgabenschwierigkeit zu berücksichtigen.

6 Sprache in allen Fächern fördern und beurteilen

«Sprache als kognitives Werkzeug» bildet Thürmann (2012, S. 5) zufolge eine wichtige Voraussetzung für «fachbezogene Kognitionen». Hinzu kommt, dass fachliche Leistungen primär durch «verbale Filter festgestellt und bewertet werden». Als ein besonderes Problem wird der Umstand bewertet, dass damit verbundene sprachliche Muster kaum explizit gemacht oder vermittelt werden:

Somit führen nicht eigentlich die heterogenen, durch die außerschulische Sozialisation vor allem in der Familie erworbenen sprachlichen Voraussetzungen der Schüler/innen zu Bildungsungleichheit, sondern erst und vor allem die Tatsache, dass die Schule ihre sprachlichen Anforderungen und Bewertungsmaßstäbe nicht offenlegt bzw. zum expliziten Vermittlungs- und Lerngegenstand macht. Erst dadurch kann das Verfügen über bildungssprachliche Kompetenzen als Eintrittskarte zu Lerngelegenheiten und Bildungsabschlüssen innerhalb der Schule als *gate-keep-ing*-Institution [...] fungieren. (Morek & Heller, 2012, S. 11)

Dies betrifft in erster Linie Schüler und Schülerinnen aus eher bildungsfernen Familien, aber auch Schüler und Schülerinnen mit Deutsch als Zweitsprache.

Generell zeigt sich über die verschiedenen Fächer hinweg, dass etwa bezogen auf Wortschatz zwischen den sprachlichen und nichtsprachlichen Fächern grosse Unterschiede bestehen. So belegt eine kanadische Studie auf der Basis von umfangreichen Unterrichtsbeobachtungen, dass in den Klassen 1–3 durchschnittlich 6% der Unterrichtszeit für Wortschatzarbeit verwendet werden, während Fächer wie Mathematik, Naturwissenschaft oder Geschichte lediglich 1,4% der Unterrichtszeit dafür aufwenden (McKeown, Deane, Scott, Krovetz & Lawless, 2017, S. 5). Wie im Folgenden deutlich werden wird, bedeutet die sprachbewusste Gestaltung des Fachunterrichts mehr als nur die Vermittlung von fachsprachlichem Wortschatz.

Des Weiteren zeigt die Meta-Analyse von Bangert-Drowns, Hurley & Wilkinson (2004),

dass sich schreibendes Lernen mit $d = .17$ auf die fachlichen Leistungen positiv auswirken kann. Die positive Wirkung lässt sich verstärken, wenn die damit verbundenen Aufgaben kurz sind ($d = .52$), das heisst max. 10 Minuten beanspruchen.

6.1 Mathematik

Sprachliche Kompetenzen bilden einen zentralen Einflussfaktor im Mathematikunterricht und sind als entscheidende Bedingung für einen erfolgreichen Kompetenzerwerb in diesem Fach zu sehen (Meyer & Tiedemann, 2017, S. 48). Vor diesem Hintergrund konnten für das Fach Mathematik in diversen Studien Leistungsunterschiede zwischen Schülerinnen und Schüler mit und ohne Migrationshintergrund offengelegt werden, die auf sprachliche Hindernisse zurückzuführen sind (Bochnik & Ufer, 2017). Diese Leistungsdifferenzen sind für den vorschulischen Bereich, für die Primar- und die Sekundarstufe belegt (Bochnik & Ufer, 2017, S. 81). Lernende mit Migrationshintergrund zeigen nicht nur insgesamt schwächere Leistungen, sondern sie stagnieren teilweise auch auf tieferem Niveau (Meyer & Tiedemann, 2017, S. 48).

Mit Blick auf Test- und Prüfungsaufgaben kann gezeigt werden, dass sie oft sprachliche Hürden enthalten, die das Lösen der Aufgabe erschweren oder sogar verhindern, die aber dennoch nicht als Teil der Aufgabenschwierigkeit definiert werden, so etwa in TIMSS 2011 (Walzebug, 2015). Tatsächlich belegen verschiedene Studien, dass Textaufgaben als Gegenstand des Sachrechnens schlechter gelöst werden als Aufgaben in rein numerischer Form (Duarte, Gogolin & Kaiser, 2011). Zu bedenken ist, dass in solchen Fällen also nicht nur mangelnde mathematische Kompetenzen am Ursprung stehen, sondern vor allem sprachliche. Ein besonderes Augenmerk gilt deshalb Textaufgaben.

Walzebug (2015) kann in einem qualitativen Zugang aufzeigen, dass sprachlich anspruchsvolle Textaufgaben ein Repertoire an sprachlichen Kenntnissen und Vorgehensweisen erfordern, das nicht privilegierte, monolingual deutsche SuS benachteiligt. Sie spricht deshalb auch von einer sprachlich bedingten sozialen Ungleichheit.

Zusammenfassend listet Walzebug (2015, S. 236) folgende sprachlichen Herausforderungen auf:

- Die Aufgabe ist als Textaufgabe formuliert.
- Sie enthält Schlüsselwörter.
- Die Textaufgabe weist eine hohe sprachliche Komplexität auf.
- Sie sind in einem offenen Antwortformat dargestellt.

Zusätzlich zeigen sich ihr zufolge grössere Herausforderungen für die kognitiven Bereiche Reproduzieren und Problemlösen sowie in den mathematischen Inhaltsbereichen Arithmetik und Geometrie/Messen.

Allerdings lässt sich bisher nicht überzeugend belegen, dass ein Vereinfachen der Textaufgaben auf lexikalischer oder syntaktischer Ebene förderlich für den Lösungsweg ist, unabhängig davon, ob es sich um Test- oder Lernaufgaben handelt (Stephany, 2017). Selbst wenn auf Zweitsprachlernende fokussiert wird, zeigen sich keine Verbesserungen. Weitgehend fehlen Studien, die nicht auf lexikalische oder syntaktische Aspekte fokussieren, sondern die Textkohärenz in den Blick nehmen. Im deutschsprachigen Raum liegen zwei Studien vor, die vielversprechende Wege aufzeigen:

- Eine Erhöhung der Textkohärenz in mathematischen Textaufgaben hat im Mittel für alle Kinder unabhängig von leserseitigen Faktoren wie der Lesekompetenz oder der Erstsprache einen positiven Einfluss auf Lösungswege und die Lösung (Stephany, 2017, S. 57).

- Walzebug (2015, S. 218) arbeitet heraus, dass privilegierte, monolingual deutsche Schüler/-innen Schwierigkeiten beim Lösen einer Aufgabe kompensieren, indem sie gezielt Strategien des Lese- und Textverstehens anwenden.

Zusammenfassend ist vor dem gegenwärtigen Forschungsstand festzuhalten, dass ein Handlungsbedarf bezogen auf sprachliche Anforderungen breit akzeptiert ist und auch vermehrt angegangen wird (vgl. für einen Überblick: Schilcher, Röhl & Krauss, 2017). Des Weiteren kann festgehalten werden, dass vor allem die Textebene zu fokussieren ist, und dass davon alle Schülerinnen und Schüler profitieren können. Mit Blick auf die Bildungsetappen bzw. Zyklen 1–3 nach Lehrplan 21 kann im Anschluss an Walzebug (2015) empfohlen werden, bereits in der frühkindlichen Sprachbildung – dazu zählt sie Kindertagesstätten sowie Kindergärten – eine systematische frühkindliche Sprachbildung zu etablieren, die die Kinder dabei unterstützt, «formelle» Register auf familialen Registern aufzubauen.

6.2 Naturwissenschaft

Ähnlich wie für das Fach Mathematik zeigt sich im Rahmen der PISA-Studie 2006 für Deutschland, dass «die Testleistungen der Jugendlichen mit Migrationshintergrund durchschnittlich 73 Punkte unterhalb der Testleistungen der Vergleichsgruppe der Jugendlichen ohne Migrationshintergrund lagen», was ungefähr einem Leistungsrückstand von drei Schuljahren entspreche (Riebling, 2013, S. 33). Dieser Rückstand könne nicht allein durch den sozioökonomischen Status der Jugendlichen erklärt werden, wie Riebling an derselben Stelle anmerkt.

Nicht zuletzt aus solchen Gründen ist die Bedeutung von Sprache im naturwissenschaftlichen Unterricht – im Rahmen von «Natur, Mensch, Gesellschaft (NMG)» im 1./2. Zyklus und in «Natur und Technik (NT)» im 3. Zyklus

– unbestritten (u. a. Hand, Norton-Meier, Gunel & Akkus, 2016; Härtig, Bernholt, Prechtl & Retelsdorf, 2015; Heitzmann, 2010; Kunz et al., 2016; Rincke, 2010). Von besonderem Interesse ist dabei oft die Fachsprache, meist in Abgrenzung zur Alltagssprache (u. a. Härtig, Pehlke, Fischer & Schmeck, 2012; Leisen, 2011; Ralle, 2015; Schmiemann, 2011). Ebenso nimmt die Beurteilung – formativ wie summativ – der verschiedenen Facetten im naturwissenschaftlichen Unterricht eine immer grössere Rolle ein, insbesondere im Zusammenhang mit kompetenzorientiertem Unterricht (z. B. Adamina, 2010; D-EDK, 2016).

Analog zum Fach Mathematik ist jedoch davon auszugehen, dass die sprachlichen Anforderungen in der Regel nicht als Teil der Aufgabenschwierigkeit definiert werden. Wie Riebling (2013, S. 47) ausführt, unterscheiden sich Prüfungsaufgaben sowie Schulbuchtexte hinsichtlich des «sprachlichen Registers» erheblich von fragend-entwickelnden Unterrichtsgesprächen, wie sie häufig im naturwissenschaftlichen Unterricht zu beobachten sind. Sie illustriert dies mit drei unterschiedlichen Diskurstypen sowie Textsorten, die alle das gleiche Thema – die Glimmspanprobe – behandeln (vgl. dazu Abbildung 3 rechts).

Gleichzeitig kann Riebling (2013) in einer Befragung von 229 Hamburger Lehrpersonen der Sekundarstufe I zeigen, dass die Lehrpersonen zu einem überwiegenden Teil Sprachbildung als eine wichtige Aufgabe auch ihres Faches betrachten, dass sie aber gleichzeitig überzeugt sind, nicht ausreichend für solche Aufgaben qualifiziert zu sein. Das geht nicht zuletzt damit einher, dass sie den Unterricht kaum aktiv bildungssprachlich gestalten, sondern mehrheitlich auf das Mittel der sprachlichen Entlastung zurückgreifen. Dazu sind u. a. folgende Massnahmen zu subsumieren:

- Anstelle von Fachbegriffen werden Näherungsbegriffe oder Umschreibungen verwendet. Die Lehrpersonen selbst greifen allgemein auf Alltagssprache zurück.

Text 1: (Zwei Schülerinnen experimentieren im Chemieunterricht

... Halt das mal. ... Jetzt glüht es ... Tu es da rein. ... Ah, die Spitze leuchtet! ... Vorbei.

Text 2: (Eine Schülerin stellt das Gruppenergebnis vor.)

Wir haben einen Holzstäbchen angezündet und dann ausgepustet. Den Stab haben wir dann so in unser Reagenzglas gehalten. Die Spitze hat wieder angefangen zu brennen.

Text 3: (Ein Text wird aus dem Lehrbuch übertragen.)

Die Glimmspanprobe

«Sauerstoff ermöglicht und unterhält die Verbrennung. Diese Eigenschaft kann man nutzen, um Sauerstoff durch die so genannte Glimmspanprobe nachzuweisen. Hält man einen an der Luft nur glimmenden Holzspan in Sauerstoff, so flammt er auf. Mit Gasgemischen, deren Sauerstoffgehalt wesentlich größer ist als in der Luft, beobachtet man ein Aufglühen des Spans.» (Asselborn et al., 2006, S. 58)

Abbildung 3: Beispiele für unterschiedliche sprachliche Anforderungen; Quelle: Riebling (2013, S. 48)

- Sie setzen möglichst spracharme Materialien ein.
- Sie verzichten ganz auf Texte oder auch auf ein Schulbuch.

Auch in forschungsbasierten Erhebungen zu naturwissenschaftlichen Kompetenzen wird zunehmend versucht, die allenfalls nicht vorhandenen sprachlichen Kompetenzen möglichst zu kompensieren. Beispielsweise wurde im angeleiteten Test im Rahmen der Validierungsstudien von HarmoS Naturwissenschaften für das 4. Schuljahr mit vielen Bildern gearbeitet: Die Schüler/-innen konnten dabei Schreibunterstützung anfordern, ohne dass sich dies negativ auf die Bewertung der naturwissenschaftlichen Kompetenzen ausgewirkt hätte (Konsortium HarmoS Naturwissenschaften+, 2008).

Schulbuchtexte stellen i. d. R. hohe Ansprüche an die Schüler/-innen, die je nach Bereich – Biologie, Geschichte etc. – unterschiedlich begründet sein können. Für Biologie-Texte können Dittmar, et al. (2017) u.a. zeigen, dass das aktive Herstellen von Text-Bild- sowie Bild-Bild-Bezügen, das Verstehen von Darstellungskodes in den Bildern oder das Verstehen der verschiedenen Textblöcke – Merkbox, Frageblock etc. – eine wesentliche Rolle im Verstehensprozess spielen. Die Lehrmitteltex-te enthalten jedoch kaum explizite Steuerungselemente, die ein solches Lesen des Textes unterstützen würden. Wird das Verstehen der Texte mit textseitigen Massnahmen unterstützt, die im Wesentlichen auf der Textebene ansetzen, zeigen sich im Leseprozess deutlich weniger problematische Ereignisse. Die Ergebnisse zeigen auch, dass die Verwendung von Bildern nicht per se den Verstehensprozess erleichtert. Des Weiteren halten Dittmar et al. (2017) fest, dass durch textseitige Massnahmen ein Vermitteln von fachspezifischen Lese-strategien nicht hin-fällig wird.

Mit Blick auf den Schwerpunkt «Beurteilen und Fördern mit Fokus auf Sprache» ist zu-dem nicht unerheblich, dass durch den einsei-tigen Fokus auf das Nennen von Begriffen «die sprachlichen Lernbedürfnisse der Schü-lerinnen und Schüler, die vor allem im Bereich der komplexen Sprachstrukturen der domä-nenspezifischen Bildungssprache liegen, von den Lehrkräften häufig unerkannt» bleiben (Riebling, 2013, S. 42). Mit Blick auf das for-mative Beurteilen besteht entsprechend auch ein Bedarf an geeigneten Beobach-tungsmöglichkeiten. Ausserdem deuten ex-emplarische Ergebnisse von Härtig et al. (2012) darauf hin, dass das Begriffsverstän-dnis als Aspekt der Fachsprache von Fachwis-sen und Unterrichtssprache unterschieden werden kann.

Konkrete Anleitungen oder Forschungsergeb-nisse zum Beurteilen und Fördern in NMG und NT mit Fokus auf Sprache stellen damit

weitgehend ein Desiderat dar (vgl. Härtig et al., 2015; Rautenstrauch & Busker, 2015). So stellt Leisen (2008) zwar verschiedene Me-thodenwerkzeuge zur Förderung der Sprach-und Lesekompetenz für Schülerinnen und Schüler vor und propagiert deren Einsatz im Unterricht, weist aber auch darauf hin, dass Sprachübungen im Physikunterricht immer dem Physiklernen und der Kommunikation in der Physik und über Physik dienen sollten (und nicht in erster Linie dem Sprachlernen). Dennoch scheinen Naturwissenschaftslehr-personen bei der Beurteilung fachlicher Lei-stungen auch die sprachlichen Leistungen als Fachleistungen mit zu bewerten (Feser, Höt-tecke, Ehmke, 2016; Tajmel, 2010).

Insgesamt bleibt festzustellen, dass noch eini-ges an Entwicklungsarbeit im Bereich Beurtei-len und Fördern im naturwissenschaftlichen Unterricht mit Fokus auf Sprache zu leisten ist. Es sind aber in dieser Hinsicht verschie-dene Bemühungen festzustellen:

- Feser & Höttecke (2017) haben ein Krite-rienraster zur Beurteilung von Erklärun-gen von Schülerinnen und Schülerinnen entwickelt, das auf vier Qualitätsmerkma-len basiert: Rahmenbau der Erklärung, Rolle von Evidenzbezügen, Tiefe und Kon-sistenz.
- Aktuell erarbeitet die Koordinations-gruppe SWiSE (Swiss Science Education; www.swise.ch) Kriterienraster für die Beur-teilung verschiedener Facetten des natur-wissenschaftlichen Unterrichts. Unter an-derem entsteht ein Kriterienraster zum Dokumentieren/Präsentieren von natur-wissenschaftlichen Sachverhalten, in dem auch das Kriterium Alltagssprache/Fach-sprache berücksichtigt wird.
- Im neuen Lehrmittel für den «Natur und Technik»-Unterricht der Sekundarstufe I (Lehrmittelverlag Zürich, erscheint 2019) wird bspw. für das Beurteilen von Bau und Dokumentation eines Modells neben den Kriterien «Modell», «Analyse des Mo-dells», «Dokumentation und Reflexion des

Vorgehens» und «Gestaltung des Modells» auch die «Sprache der Dokumentation» explizit thematisiert.

6.3 Allgemeiner Ausblick

Ansätze, die über «blosse» Sprachübungen hinausgehen, die insbesondere fachspezifische sprachliche Handlungen fokussieren – etwa mit Blick auf das Lesen von fachspezifischen Texten oder mit Blick auf das schreibende Lernen sowie das Verfassen fachspezifischer Texte –, können weitgehend als Desiderat gelten. Zwar können bspw. Hand, Wallace & Yang (2004) oder Akkus, Gunel & Hand (2007) mit dem Förderansatz «Science Writing Heuristic» einen sehr lernwirksamen Ansatz vorlegen, der in erster Linie dem fachlichen Lernen dient, dabei aber auch fachspezifische Arbeits- und Schreibformen integriert: Ein solcher Förderansatz setzt jedoch ein bestimmtes Verständnis des naturwissenschaftlichen Unterrichts sowie bestimmte Rahmenbedingungen wie Lektionsumfang u. a. voraus, was so im naturwissenschaftlichen Unterricht der Schweiz nicht gegeben ist. Basierend auf einer Sichtung von Unterrichtsvideos aus dem Projekt QUIP (Fischer, Labudde, Neumann & Viiri, 2014) vertreten Beerenwinkel & Sturm (2018) die These, dass die Implementation umfangreicher Förderansätze wie «Science Writing Heuristic» nicht an bestehende Unterrichtsformen anknüpfen kann. Vielmehr empfehlen sie einen Ansatz, der in kleinen Schritten und mittels kleiner Aufgaben epistemisches Schreiben in den Fachunterricht integriert, dies durchaus im Sinne von Bangert-Drowns, Hurley und Wilkinson (2004). Das heisst, kürzere Aufgaben

zeigen die positiveren Effekte als zeitlich umfangreiche Aufgaben (besonders umfangreiche Schreibaufgaben können sogar einen negativen Effekt auf das fachliche Wissen ausüben).

Erschwerend kommt hinzu, dass der Lehrplan 21 lediglich implizite Hinweise auf Lese- oder Schreibaktivitäten in Mathematik oder Naturwissenschaft enthält. Ein Beispiel kann das illustrieren:

Die SuS können beobachten, beschreiben und dokumentieren, wie ein bestimmter Reiz eine entsprechende Reaktion auslöst. (Bildungsdirektion Kanton Zürich, 2017, S. 78)

Der Lehrplan 21 enthält auch keine Hinweise dazu, wie die mit «Beschreiben» oder «Dokumentieren» verbundenen sprachlichen Anforderungen zu vermitteln sind. Da mit dem Lehrplan 21 eine erhöhte Versprachlichung einhergeht, braucht es zur Orientierung – gerade mit Blick auf schulinterne Weiterbildungen zu diesem Bereich – Materialien sowie Hilfestellungen:

- Für untere Klassenstufen enthält die Online-Lerneinheit zum Lehrplan 21 einen Lehrfilm, der aufzeigt, wie mit einer 3. Primarklasse ein Forschungsbericht zu einem Waldtag verfasst wird. Dabei wird gezeigt, wie in Zusammenarbeit mit der DaZ-Lehrperson insbesondere DaZ-Lernende unterstützt werden:⁵ Eine grosse Rolle spielen dabei Textprozeduren.
- Handreichungen wie jene von Lindauer et al. (2013) enthalten für Sekundarstufe I Orientierungshilfen für die Arbeit in Regelklassen, die gerade auch für schulinterne Weiterbildungen einsetzbar sind.

⁵ Der Film wird jedoch als Beispiel zum Schreibunterricht im Deutschunterricht abgehandelt.

7 Herausforderungen beim Beobachten und Beurteilen

Ziel dieses Kapitels ist es, spezifische Herausforderungen mit Blick auf Chancengerechtigkeit oder Heterogenität herauszuarbeiten, die im Rahmen der schulinternen Weiterbildungen zum Schwerpunkt «Beurteilen und Fördern mit Fokus auf Sprache» eine wichtige Rolle spielen könnten.

7.1 Erwartungen und Schulerfolg

Eine Schulkultur lässt sich als Gefüge von stillschweigenden Annahmen und Überzeugungen auffassen (Tschannen-Moran, Salloum & Goddard, 2015): Die Handlungen und das Verhalten der Lehrpersonen sind dabei beeinflusst vom Umfeld. Wie Oester, Fiechter & Kappus (2008) in einer ethnografischen Studie zu drei Berner Quartierschulen mit unterschiedlich hohem Migrationsanteil zeigen, neigen Lehrpersonen bei hohem Migrationsanteil dazu, das soziale Lernen zulasten des akademischen Lernens ins Zentrum zu stellen, während sie bei geringerem Anteil das akademische Lernen stärker im Blick haben (vgl. auch Kap. 7.3).

Auch wenn Lehrpersonen als Einzelpersonen für den Schulerfolg eine zentrale Rolle spielen, sollten kollektive Faktoren nicht ausgeblendet werden. Tschannen-Moran, Salloum & Goddard (2015) heben dabei vor allem die kollektiven Selbstwirksamkeitserwartungen von Lehrpersonen hervor, ihre Schüler und Schülerinnen erfolgreich fördern zu können. Eine Schule bzw. ein Kollegium mit hohen Selbstwirksamkeitserwartungen kann dabei auf verschiedene Arten und Weisen das Lernen positiv beeinflussen (Tschannen-Moran et al., 2015, S. 303):

- Die Auswirkungen von niedrigem SES können gemildert werden.
- Eine kollektive Selbstwirksamkeitserwartung beeinflusst, wie einzelne Lehrpersonen unterrichten, wie sie ihre Klassen managen, die Schüler/-innen motivieren und

wie sie auf Hürden oder Schwierigkeiten reagieren.

- Die Lehrpersonen von Schulen mit hoher Selbstwirksamkeitserwartung verwenden weniger Zeit und Engagement in Bezug auf das Verhalten der Schüler/-innen: Vielmehr akzeptieren sie unzureichende Schulleistungen etwa von Lernenden mit niedrigem SES nicht als unumgängliche Tatsache.
- Schulen mit hoher kollektiver Selbstwirksamkeitserwartung zeichnen sich durch eine starke Zusammenarbeit der Lehrpersonen aus, nicht zuletzt mit dem Ziel, den Unterricht zu verbessern. Ein Nebeneffekt ist, dass dies die Widerstandskraft der Lehrpersonen bei Schwierigkeiten erhöhe.

Wenn die Schule als Ganzes ein hohes Vertrauen in ihre Fähigkeit und gleichzeitig auch in ihre Schüler/-innen zeigt, dann geht dies auch mit besseren Schulleistungen der Schüler/-innen einher. Schulen sind damit Tschannen-Moran et al. (2015) zufolge gefordert, Bedingungen zu schaffen, unter denen sich ein positives Vertrauen entwickeln kann. Sie weisen darauf hin, dass gerade der Schulleitung dabei eine zentrale Rolle zukommt:

- Lehrpersonen nehmen die Haltung der Schulleitung wahr und reagieren auch entsprechend darauf.
- Die Lehrpersonen erhalten ausreichend Gelegenheiten für die Zusammenarbeit im Kollegium wie auch mit externen Fachpersonen sowie für das gemeinsame Vor- und Nachbereiten.

Auch auf Ebene der einzelnen Lehrperson spielen (Leistungs-)Erwartungen eine zentrale Rolle.

Urteilsakkuratheit, wie sie in Kap. 3.3 ausgeführt wurde, erfolgt zu 75% ohne Beeinflussung der Schüler/-innen, <lediglich> zu 25%

kann sie als selbsterfüllende Prophezeiung bezeichnet werden, indem falsche Erwartungen ihre eigene Wirklichkeit kreieren (Jussim et al., 2009). Übers Ganze gesehen handelt es sich eher um geringe Effekte, die sich Jussim et al. (2009, S. 354) zufolge zwischen 0 und .40 bewegen, in den meisten Fällen zwischen .10 und .20. Die Befunde können etwas differenziert werden, wenn man spezifische Gruppen, die stärker anfällig für niedrige Erwartungen sind, betrachtet: Häufig trifft es ganze Gruppen wie Schüler/-innen mit Migrationshintergrund oder niedrigem sozioökonomischen Status (Möller, Machts & Retelsdorf, 2016, S. 17). Das zeigt auch der Überblick in Jussim et al. (2009):

- a) Für schwächere Schüler/-innen beträgt die Effektstärke ca. $d = .30$ –.40.
- b) Für Schüler/-innen mit geringem SES beträgt die Effektstärke ca. $d = .20$ –.30.
- c) Für schwächere Schüler/-innen mit geringem SES kann sogar ein Effekt von bis zu $d = .60$ nachgewiesen werden.
- d) Für afro-amerikanische Schüler/-innen zeigte sich ein durchschnittlicher Effekt von $d = .40$ bis .60.

Auffällig ist, dass sich niedrige Erwartungen bei Schülern und Schülerinnen mit geringeren Leistungen in Kombination mit geringem SES gewissermassen kumuliert auswirken. In diesem Zusammenhang erwähnenswert ist, dass es Lehrpersonen generell schwerer fällt, die Leistungen von schwächeren Schülern und Schülerinnen akkurat zu beurteilen (Südkamp et al., 2012).

Niedrige Erwartungen der Lehrpersonen führen dazu, dass sie starke und schwache Schüler/-innen anders behandeln (Jussim et al., 2009, S. 364ff.), indem schwache Schüler/-innen öfters Feedback erhalten, das sich nicht auf ihre Leistung, sondern vor allem auf ihr Verhalten bezieht. Wie Sturm, Lindauer & Sommer (i. Vorb.) im Rahmen einer Fallstudie

zum Schreibunterricht von Lehrpersonen in ihrem ersten Berufsjahr zeigen, gehen solche Erwartungen, die als Rückmeldung auf der persönlichen Ebene formuliert werden (vgl. auch Kap. 2) auch mit der Auffassung einher, dass die Verantwortung fürs Lernen allein bei den Schülern und Schülerinnen liegt.

Hinzu kommt, dass Lehrpersonen auch dazu neigen, starken Schülern und Schülerinnen mehr Lerngelegenheiten zu verschaffen, die zudem hierarchiehöhere Prozesse fokussieren, während bei schwachen Schülern und Schülerinnen gerne repetitive Aufgaben oder auch Aufgaben, die das Auswendiglernen ins Zentrum stellen, eingesetzt werden.

Des Weiteren können negative Erwartungen dazu führen, dass bestimmte Fördermethoden grundsätzlich nicht angewendet werden: So berichtet in der Fallstudie von Sturm et al. (2018) eine Lehrperson, dass sie kooperatives Schreiben nicht einsetze, da Schüler/-innen, die bequem seien, sich zu sehr auf ihre Partner verlassen würden. Ähnlich erwähnt eine Lehrperson in einem Gruppeninterview zum neuen Schwerpunkt C (vgl. Kap. 9), dass Peer-Feedback in QUIMS-Schulen nicht durchführbar wäre, da die Schüler/-innen nicht über die entsprechenden sprachlichen Voraussetzungen verfügen würden.

Bölsterli et al. (2018) können im Schweizer Projekt SCALA systematisch verzerrte Leistungserwartungen von Lehrpersonen in den Fächern Mathematik und Deutsch ebenfalls nachweisen (in Deutsch wurden Testaufgaben zu Leseverstehen sowie Rechtschreibung und Grammatik eingesetzt), und zwar bezogen auf Migrationshintergrund, sozioökonomischen Status und Geschlecht der Schüler/-innen.⁶ Bei jenen Lehrpersonen, die an einer Weiterbildung teilnahmen, in der in Kleingruppensitzungen sensibilisiert und informiert, dabei auch die eigene Haltung reflektiert wurde, die zudem um Elemente wie

⁶ An dieser Studie nahmen auch QUIMS-Klassen teil.

Projektcoaching (Hospitation und Reflexion) sowie Peercoaching (inkl. Entwicklung konkreter Massnahmen) angereichert wurde, konnte der Effekt der erfassten Stereotypen beseitigt werden (nicht aber bei der Kontrollgruppe).

Falsche Erwartungen können nicht nur zu tief, sondern auch zu hoch sein. Wie Jussim et al. (2009, S. 376) anmerken, können «irrtümlich» hohe Erwartungen einen deutlich grösseren Effekt selbsterfüllender Prophezeiung auslösen – was in diesem Falle durchaus als ein erwünschter positiver Effekt gesehen werden kann – als zu niedrige Erwartungen.

7.2 Stereotypenbedrohung und sprachliches Lernen

Schüler/-innen mit Migrationshintergrund weisen im Vergleich zu Schülern und Schülerinnen ohne Migrationshintergrund einen geringeren schulischen Kompetenzerwerb auf und stehen mit Blick auf den Erwerb sprachlicher Kompetenzen auch vor grossen Herausforderungen: Um Lernangebote überhaupt nutzen zu können, ist insbesondere ein umfassender Wortschatz zentral (Sander, Ohle, McElvany, Zander & Hannover, 2018). Letzteres zeigt sich nicht zuletzt daran, dass sowohl allgemeiner wie auch schulischer Wortschatz mit den Leistungen in verschiedenen Schulfächern zusammenhängt (Philipp, 2012). Allerdings wird auch oft darauf hingewiesen, dass die teilweise erheblichen Unterschiede im Wortschatz eher auf den sozioökonomischen Hintergrund zurückführbar seien (Beck, McKeown & Kucan, 2013).

In diesem Zusammenhang untersuchten Sander et al. (2018), inwiefern die sogenannte Stereotypenbedrohung eine mögliche Ursache für geringeren Wortschatz bzw. für geringeren Wortschatzzuwachs gelten kann.⁷ Unter «Stereotypenbedrohung» verstehen

sie eine Situation, der zufolge ein Schüler oder eine Schülerin um ein negatives Stereotyp über eine Gruppe, zu der er oder sie gehört, weiss und dass dieses Wissen «die Bedrohung auslöst, das Stereotyp selbst zu bestätigen» (Sander et al., 2018, S. 179). Ein solches Stereotyp könnte bspw. darin bestehen, dass Schüler/-innen aus dem arabischen Raum in einem bestimmten Bereich für weniger kompetent gehalten werden. Kennen die Schüler/-innen dieses Stereotyp – unabhängig davon, ob sie selbst daran glauben oder nicht –, kann es zu einer Bestätigung dieses Stereotyps kommen.

In ihrer Studie gingen sie der Frage nach, inwiefern sich eine explizite sowie implizite Stereotypenbedrohung im Vergleich zu einer Kontrollgruppe ohne Stereotypenbedrohung in einer Lernsituation auswirkt: Die explizite Bedrohung wurde hergestellt, indem den Schülern und Schülerinnen im Fach Deutsch gesagt wurde, dass Schüler/-innen mit einer nicht-deutschen Familiensprache schlechter neue deutsche Wörter lernen. Implizit wurde sie hergestellt, indem die Lernenden vor der Lerneinheit Fragen zu ihrer Familiensprache, zu ihrem Migrationshintergrund zu beantworten hatten. Die Ergebnisse können einen negativen Effekt nachweisen, der für die explizite Stereotypenbedrohung erwartungsgemäss höher ausfällt. Wie ein solcher Effekt verhindert werden kann, wurde von Sander et al. (2018) nicht untersucht. Sie verweisen lediglich darauf, dass bereits das Wissen um einen solch negativen Effekt wichtig sein könne.

Wright und Neuman (2014), die 55 Kindergärten während jeweils 12 Stunden, verteilt über vier Tage, beobachteten, können überdies zeigen, dass in Kindergärten, die einen hohen Anteil an Schülern und Schülerinnen mit niedrigem sozioökonomischen Status (SES) aufweisen, signifikant weniger Gelegenheiten

⁷ So kann etwa für die Mathematikleistung in Testsituationen ein negativer Effekt durch die Stereotypenbedrohung in verschiedenen Studien nachgewiesen werden (Sander et al., 2018, S. 182).

zur Wortschatzvermittlung geschaffen werden als in Kindergärten, die einen hohen Anteil an Schülern und Schülerinnen mit hohem SES haben. Zudem werden in Ersteren deutlich weniger anspruchsvolle Ausdrücke vermittelt. Und nicht zuletzt erklärten die Kindergarten-Lehrpersonen den Schülern und Schülerinnen mit vergleichsweise hohem SES deutlich häufiger Wörter, die sie nicht verstanden.

7.3 Sprachförderung – bildungssprachlicher Wortschatz

Mehrfach wird betont, dass sich die sprachlichen Fähigkeiten der Kinder beim Schuleintritt teilweise erheblich unterscheiden, und zwar abhängig in erster Linie vom sozioökonomischen Status (u. a. Beck et al., 2013; Isler, Künzli & Wiesner, 2014). Für das Hörverstehen lassen sich dagegen weder in Bezug auf den sozioökonomischen Status noch für die familiäre Sprache – monolingual deutsch vs. nicht monolingual deutsch – überproportionale Leistungsunterschiede nachweisen: Vielmehr bereiten bildungssprachlich anspruchsvolle Hörtexte allen Schülern und Schülerinnen grössere Schwierigkeiten, und zwar in erster Linie bezogen auf bildungssprachlichen Wortschatz, weniger mit Blick auf grammatisch komplexe Strukturen (Heppt et al., 2014).

Unterschiede in den sprachlichen Fähigkeiten scheinen sich fortzusetzen – allerdings nur, wie Beck et al. (2013, S. 2) betonen, wenn nichts dagegen unternommen wird. So vergleichen McKeown & Beck (2014) drei Förderansätze – a) einmaliges Vorlesen und Erläutern eines Ausdrucks während des Vorlesens, b) mehrmaliges Vorlesen und c) wiederholtes Erläutern, Vorlesen und Vermitteln von ausgewählten Fokuswörtern, die zentral für das Verstehen des Textes sind –, und zwar bei Kindergarten-Kindern. Dabei zeigt sich zugunsten des robusten Wortschatz-Trainings, das auf ein vertieftes Verstehen abzielt und

dazu mit interaktiven Übungsformaten arbeitet, einen grösseren positiven Effekt, und zwar bezogen auf das Anwenden des vermittelten Wortschatzes und die Verwendung im Kontext. Ähnlich zeigen Ennemoser, Kuhl & Pepouna (2013) für Vorschul-Kinder mit Migrationshintergrund und sprachlichen Schwierigkeiten, dass ein Training zum dialogischen Lesen, das sprachförderliche Interaktionsmerkmale aufweist, gegenüber einem regulären Ansatz mit $d = .47$ den grösseren Leistungszuwachs generiert. Bedenkt man zudem, dass Wortschatz – vor allem bildungssprachlicher Wortschatz – ab etwa 3./4. Klasse hauptsächlich über das Lesen erworben wird (McKeown et al., 2017), kommt einer Verknüpfung von Leseunterricht mit Wortschatzvermittlung eine nicht zu unterschätzende Rolle zu.

Allgemein kann festgestellt werden, dass Wortschatzvermittlung in erster Linie situativ mit kurzen Erklärungen zur Wortbedeutung erfolgt, dass diese Erklärung jeweils durch die Lehrperson in Form einer kurzen Definition gegeben wird (McKeown et al., 2017). Verschiedene Untersuchungen zeigen jedoch, dass Wortschatz im Rahmen vielfältiger Kontexte und in Verbindung mit anderen sprachlichen Ausdrücken erworben wird. Das bereits erwähnte robuste Wortschatz-Training erzielt gegenüber einer situativen Wortschatzvermittlung deutlich bessere Resultate.

Isler et al. (2014) führen keine Interventionsstudie durch, sondern richten ihr Augenmerk auf die Qualität von sprachlichen Bildungsprozessen im Kindergarten-Unterricht, da diese eine zentrale Rolle spielt. Qualitativ hochstehende Bildungsprozesse vermitteln nicht nur curriculare, sondern auch nicht-curriculare Bildungsinhalte:

Informell erworben werden zum Beispiel die diskursiven Fähigkeiten, zwischen Referenzräumen zu wechseln, verschiedene Sprachhandlungen zu realisieren oder Gesprächsrollen situationsgemäss auszugestalten. Diese Befunde unterstreichen die Funktion des Kindergartens, möglichst allen Kindern frühzeitig Zugang zu jener Alltagsbildung zu

verschaffen, die in der Schule stillschweigend vorausgesetzt wird, aber nicht in allen Familien gleichermaßen präsent ist [...]. (Isler u. a., 2014, S. 473)

Um Kindergarten-Lehrpersonen dabei besser unterstützen zu können, schlagen Isler et al. (2014) Videofeeds in der Aus- und Weiterbildung vor.

7.4 Korrekatives Feedback

Korrektives Feedback findet besonders im Kontext des Zweitspracherwerbs statt, wobei dessen Wirksamkeit im mündlichen sowie schriftlichen Bereich umstritten ist. So belegen etliche Studien einen positiven Effekt des korrektiven Feedbacks auf sprachliche Kompetenzen, während in anderen Studien keine Nachweise dafür festzumachen sind (Bitche-ner, Young & Cameron, 2005). Der Grund für diese ambivalente Forschungslage ist methodischer Natur und liegt hauptsächlich in der Konzeptualisierung des korrektiven Feedbacks (Russell & Spada, 2006). Korrektives Feedback kann implizit oder explizit erfolgen, es ist unter anderem abhängig von der Menge, der Häufigkeit und davon, ob es mündlich oder schriftlich erfolgt. Des Weiteren spielt es eine Rolle, wer das korrektive Feedback erteilt und wie die Lernenden damit umgehen (ebd.). Je nach dem, welche Art des korrektiven Feedbacks, welchen Aspekt die unterschiedlichen Studien fokussieren,

sind entsprechend auch unterschiedliche Ergebnisse zu erwarten.

Aufgrund der unübersichtlichen Forschungslage gingen Russell & Spada (2006) in einer Metaanalyse der Frage nach, wie effektiv korrektives Feedback übers Ganze gesehen für den Zweitspracherwerb ist. Mit einer Effektstärke von $d=1.16$ verweisen die Resultate auf einen hohen Einfluss des korrektiven Feedbacks im mündlichen und schriftlichen Bereich, und zwar weniger auf hierarchiehöhere Aspekte wie bspw. Textqualität, sondern sprachformale Korrektheit. Der hohe positive Effekt ist darüber hinaus mit Vorsicht zu interpretieren, da korrektives Feedback nur effektiv ist, wenn es fokussiert erfolgt. In diesem Zusammenhang betont Truscott (1999), dass aus korrektivem Feedback, das auf grammatikalische Fehler und damit auf die Wort- und Satzebene fokussiert ist, in diesem Bereich zwar Verbesserungen resultieren können, jedoch nicht auf inhaltlicher Ebene, die im Wesentlichen die eigentliche Textebene umfasst (bezogen auf die Gesamtidee, die Kohärenz u. Ä.). Des Weiteren ist zu beachten, dass andere Feedbackformen je nach Fokus geeigneter sind. So konnten Shintani & Ellis (2013) zeigen, dass das Aneignen von explizitem Grammatikwissen mit metalinguistischen Erklärungen begünstigt wird, während korrektives Feedback keinen positiven Effekt auf den Erwerb hat.

8 Wirksame Weiterbildung im Bereich Beobachten und Beurteilen (mit Fokus auf Sprache)

Im bisherigen Schwerpunkt räumten die Auftraggeberin (Volksschulamt, QUIMS) sowie die Auftragnehmerin (vgl. Neugebauer, 2015) – basierend auf Lipowsky und Rzejak (2012) – der **Verschränkung von Input, Erprobung und Reflexion** einen hohen Stellenwert in der schulinternen Weiterbildung ein. Dieser Ansatz hat sich insgesamt bewährt (Neugebauer, 2017; Roos, 2017): Hauptsächlich die Erprobung eines Förderansatzes und die Reflexion der dabei gemachten Erfahrungen (inkl. der Analyse von Schülerbeispielen) stellt ein Schlüsselmoment der Schul- und Unterrichtsentwicklung dar. Ohne konkrete Erprobungen besteht die Gefahr, dass Diskussionen isoliert über Methoden geführt werden und dabei auch kein Bezug zum Lernen der Schüler/-innen oder zu einer lerntheoretischen Auffassung hergestellt wird. Gerade die Auseinandersetzung, die kollegiale Debatte über das Lernen der Schüler/-innen, in Verbindung mit einer Diskussion über die Wirkung des eigenen Unterrichts ist entscheidend, wie Hattie (2012, S. 162f.) festhält. Das setzt allerdings voraus, dass sich eine Schule als lernende Einheit versteht, dass sich nicht zuletzt auch Lehrpersonen selbst als Lernende betrachten.

Ebenso bewährte sich, dass die schulinterne Weiterbildung einen **fachdidaktischen Fokus** legte (Roos, 2017; Sturm, 2017). Ein inhaltlicher Fokus stellt – so Lipowsky und Rzejak (2012, S. 5) – gerade «in Verbindung mit einer Förderung des „*assessment knowledge*“ eine wichtige Voraussetzung für die Veränderung beruflichen Lehrerhandelns» dar, wofür sich ihnen zufolge einige empirische Belege finden. Umso wichtiger ist es, dass im Schwerpunkt «Beurteilen und Fördern mit Fokus auf Sprache» für die Arbeit mit Schulen ein klarer fachdidaktischer Fokus gelegt wird (vgl. dazu die vorgeschlagenen Themenfelder in Kap. 1.2).

Der Schwerpunkt C «Beurteilen und Fördern mit Fokus auf Sprache» stellt jedoch im Vergleich zum Schwerpunkt A «Schreiben auf allen Schulstufen» zusätzliche Ansprüche an die Arbeit mit Schulen. Stellvertretend zeigt sich dies in der Studie von Hebbecke & Souvignier (2018), die das formative Beurteilen im Kontext der Leseförderung untersuchten. Sie unterscheiden in ihrer Studie für das formative Assessment die drei Komponenten Diagnostik, Feedback und Förderung, die im Unterricht zu implementieren sind. Dabei gehen sie der Frage nach, in welchem Mass die Implementation dieser drei Komponenten gelingt und welche Wirksamkeit die Nutzung von Feedback- und Fördermaterialien im Vergleich zu einer reinen Lernverlaufdiagnostik auf das Lesen hat. Sie stellen fest, dass die Implementation der Feedback- und Fördermaterialien nicht zufriedenstellend gelungen ist. Die Lehrpersonen zeigen zwar eine hohe Akzeptanz für die Vorgehensweisen, jedoch schätzen sie die Umsetzbarkeit und die Nutzung bedeutend tiefer ein. Ebenfalls konnten Hebbecke & Souvignier (2018) für die beiden unterstützenden Massnahmen – Materialien zum Feedback sowie zusätzliche Fördermaterialien – keinen zusätzlichen positiven Effekt nachweisen.

Die Studie weist damit auf einen zentralen Punkt hin: Feedback und Fördermassnahmen wirken sich im Rahmen einer formativen Beurteilung dann positiv auf das Lernen aus, wenn die Lehrperson sie genügend in ihren Unterricht integriert. Hebbecke & Souvignier (2018, S. 27) betonen, dass mit Blick auf «eine praxistaugliche und wirksame Integration der drei Komponenten Diagnostik, Feedback und Förderung sowie deren nachhaltige Implementation» noch viel Entwicklungsarbeit zu leisten sei, die mit Blick auf die Schulen auch langfristiger zu denken sei:

After five years of work, our euphoria devolved into a reality that formative assessment, like so many other education reforms, has a long way to go before it can be wielded masterfully by a majority of teachers to positive ends. (Shavelson, 2008, zitiert nach Hebbecke & Souvignier, 2018, S. 27)

Timperley (2011) stellt in ihrem Buch einen Ansatz vor, den sie nicht nur basierend auf einer Synthese relevanter empirischer Literatur entwickelt hat, sondern auch basierend auf über 10 Jahren Erfahrung mit schulinternen Weiterbildungen im Rahmen eines gross angelegten Programms zur Schul- und Unterrichtsentwicklung. Sie zeigt dabei u.a. auf, dass Lehrpersonen Informationen aus Beobachtungen und Beurteilungen selbst interpretieren sollten – möglichst im Team und nicht auf sich allein gestellt –, damit sie ihre Theorien, Überzeugungen prüfen und gegebenenfalls auch verändern können. Wird ihnen dieser Teil wie in der Studie von Hebbecke & Souvignier (2018) abgenommen, tangiert dies entsprechend ihre Theorien oder Überzeugungen nicht.

Timperley & Parr (2010) sowie Timperley et al. (2007) haben u.a. basierend auf Arbeiten von Paul Black und Dylan William einen «Kreislauf» entwickelt, der für die Arbeit mit Schulen einen **forschenden Zugang** sowie ein gemeinsames professionelles Lernen favorisiert (Abbildung 4, S. 41, illustriert diesen Kreislauf). Der Kreislauf ist so angelegt, dass die Leitfragen auch die verschiedenen Ebenen – Schule, Lehrperson, Schüler/-innen – in den Blick nehmen. Wie sich die Arbeit mit Schulen auf dieser Basis gestaltet, hat Timperley (2011) ausführlich dargelegt: Mithilfe verschiedener Fallbeispiele zeigt sie auf, dass Lehrpersonen selbst in der Lage sein müssen, Beobachtungen und Ergebnisse aus Beurteilungsinstrumenten zu interpretieren – das trifft auch auf summatives oder prognostisches Beurteilen zu –, im Team zu diskutieren und dann geeignete Fördermassnahmen

zu treffen. Das beinhaltet im Sinne von Hattie (2012) ein Verständnis von Schule als lernender Einheit.

Ein Beispiel aus Timperley (2011, S. 35ff.) kann dies illustrieren (vgl. auch Sturm, 2018): So zeigte sich eine neuseeländische «Brennpunkt»-Schule mit den Leseleistungen ihrer Schüler/-innen nicht zufrieden und beschloss Massnahmen. Diese zeigten im Verlauf des Schuljahres Wirkung. Nach den grossen Sommerferien jedoch war davon nichts mehr zu sehen. In der Diskussion meinten einige Lehrpersonen, dass die Schüler/-innen während der Sommerferien wohl nicht gelesen hätten, dass man – wolle man die Leseleistung konstant halten – die Eltern einbeziehen müsste. Andere Lehrpersonen waren skeptisch. Dies führte dazu, dass sie genauer hinschauten und bspw. realisierten, dass sie mehrere Wochen vor den Sommerferien keine eigentliche Leseförderung mehr betrieben und vor allem die Schüler/-innen selbst nicht dazu anleiteten, über die langen Sommerferien Verantwortung für ihr Lesen übernehmen zu können. Als Ergebnis dieses Hinterfragens beschloss die Schule, gemeinsam daran zu arbeiten, den Sommereffekt zu vermeiden, einerseits durch gezielte Elternarbeit, andererseits aber vor allem durch eine schulische Leseförderung, die die Schüler/-innen selbst auf das Lesen während der Sommerferien vorbereitet.

Inwiefern ein solcher Kreislauf, ein solch forschender Zugang in das neue Konzept der schulinternen Weiterbildung integriert werden kann, wird noch zu prüfen sein. Der Kreislauf ist dabei so zu lesen, dass er nicht nur die kooperative Zusammenarbeit von externen Fachpersonen und der Schule beschreibt, sondern auch die Zusammenarbeit von Schulleitung und Team.

**Teacher inquiry and knowledge-building cycle
to promote valued student outcomes.**

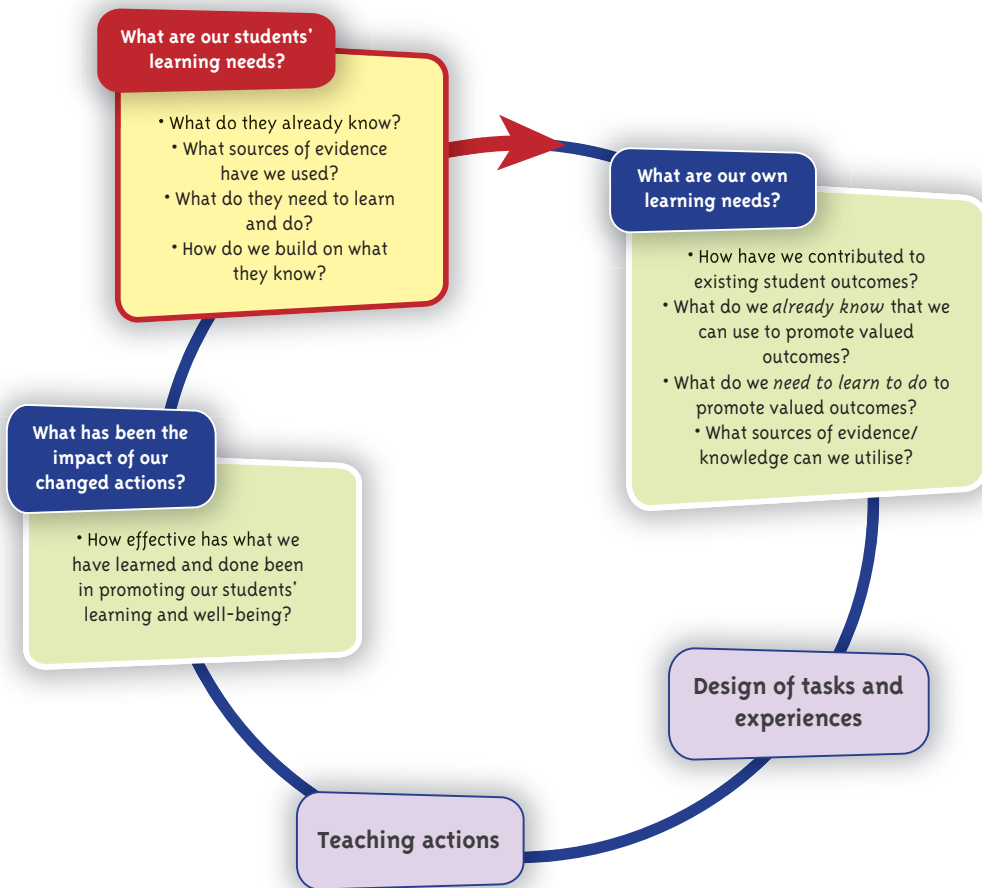


Abbildung 4: Quelle: Timperley et al. (2007)

9 Die Perspektive von Lehrpersonen und Schulleitungen

Während die Kapitel 1–8 den theoretischen Hintergrund aufgearbeitet haben, bietet dieses Kapitel einen Einblick in das Praxisfeld: So wurden Interviews mit Schulleitungen sowie mit Lehrpersonen aus allen Schulstufen durchgeführt, um zumindest einen Einblick in verschiedene QUIMS-Schulen und ihren Zugang zum Schwerpunkt «Beobachten und Beurteilen mit Fokus auf Sprache» zu erhalten:

- a) Mit den Lehrpersonen wurden Gruppeninterviews durchgeführt, da in der Reaktion aufeinanderhaltungen und Überzeugungen schärfer zu Tage treten können als in Einzelinterviews (Troia & Maddox, 2004).
- b) Mit den Schulleitern und Schulleiterinnen wurden Einzelinterviews durchgeführt, damit die Besonderheit der jeweiligen Schule besser gefasst werden kann und die Befragten ihre Sicht möglichst frei äussern können.

Für a) und b) wurden separate Interview-Leitfäden erstellt, die im Anhang aufgeführt sind. Die einzelnen Fragen richten sich stark an den Zielen des Schwerpunkts aus, wie sie vom Bildungsrat am 30. Oktober 2017 verabschiedet wurden (die Ziele sind ebenfalls im Anhang enthalten). Darüber hinaus nehmen sie – vor allem auch in den Detailfragen – wichtige Aspekte auf, wie sie in den vorangehenden Kapiteln herausgearbeitet wurden.

Tabelle 2 gibt einen Überblick über die Schulen, die seitens der befragten Schulleiter/-innen involviert waren. Die Schulen wurden so ausgewählt, dass alle Schulstufen sowie Stadt und Kanton Zürich berücksichtigt sind. Zusätzlich wurde darauf geachtet, dass sie unterschiedlich lange am Programm QUIMS teilnehmen und möglichst auch unterschiedliche Erfahrungen zum neuen Schwerpunkt aufweisen.

Schule A	<ul style="list-style-type: none"> – Agglomeration Zürich – 8 Kindergarten-Klassen und 13 Primarschul-Klassen – Seit zwei Jahren eine eigenständige QUIMS-Schule, davor aber im Verbund mit anderer Schule schon dabei (insgesamt nun 10 Jahre) – Absolvierte SchilW zum Schwerpunkt «Schreiben auf allen Schulstufen», arbeitet nun am Thema Wortschatz
Schule B	<ul style="list-style-type: none"> – Kanton Zürich – 7 Kindergarten-Klassen und 17 Primarschul-Klassen – Neu als QUIMS-Schule eingestuft – Schloss vor Kurzem eine dreijährige Weiterbildung zum Thema Beurteilen ab
Schule C	<ul style="list-style-type: none"> – Stadt Zürich – 18 Sekundarschul-Klassen (Niveaus A, B und C), zusätzlich eine Aufnahmeklasse – Seit 8 Jahren eine QUIMS-Schule – Absolvierte eine Kurz-Weiterbildung zum Schwerpunkt «Schreiben auf allen Schulstufen»
Schule D	<ul style="list-style-type: none"> – Kanton Zürich – 15 Sekundarschul-Klassen – Seit 2 Jahren eine QUIMS-Schule – Keine Weiterbildung zum Schwerpunkt «Schreiben auf allen Schulstufen» über PHZH besucht, einzelne Lehrpersonen nahmen aber an der Veranstaltungsreihe teil

Tabelle 2: Übersicht über die Schulen, die die interviewten Schulleitungen leiten

Die Lehrpersonen wurden so angefragt, dass ein möglichst breites Bild entstehen kann. Folgende Kriterien wurden herangezogen:

- Teilnahme an den bisherigen QUIMS-Veranstaltungen
- QUIMS-Beauftragte
- Schulstufe
- Klassenlehrperson, DaZ-Lehrperson
- Erfahrung mit dem Programm QUIMS

Eine der interviewten Lehrpersonen unterrichtet an einer Schule, die bereits als Pilot-schule am neuen Schwerpunkt beteiligt ist. Eine weitere Lehrperson erprobt erste Materialien für den neuen Schwerpunkt.

Die Interviews wurden zwischen dem 19. Oktober 2018 und dem 21. November 2019 durchgeführt. Die Interviews wurden aufgezeichnet und entlang der Leitfragen zusammengefasst. Die Zusammenfassungen wurden in vereinfachter Weise inhaltsanalytisch ausgewertet.

9.1 Die Sicht von Lehrpersonen

In den Interviews zeigt sich ein breites Verständnis von Beurteilungsformen: In erster Linie wird formatives Beurteilen mit guter Beurteilungspraxis gleichgesetzt, summatives Beurteilen dagegen als Notwendigkeit gesehen (prognostisches Beurteilen wird nicht explizit genannt):

- Mit Blick auf formatives Beurteilen wird betont, dass dies die Schüler/-innen weiterbringe, dabei die Bezugsnormen klar seien und eine Orientierung an den Grundansprüchen des Lehrplans 21 hilfreich sei (im Sinne von «weniger ist mehr»).
- Mit Blick auf den Übertritt in die Sekundarstufe und die Einteilung in Sek A und B/C werde der Sprachstand zu stark in den Vordergrund gestellt. Das gelte so auch für die Aufnahmeprüfung ins Gymnasium oder in die Berufsschulen.

- Summatives Beurteilen wird dabei v. a. von Lehrpersonen des Zyklus 2 als kritisch gesehen, da es ein zu starkes Gewicht im Schulalltag einnehme und so den Unterricht auch negativ beeinflussen könne.

Dagegen betonen die befragten Lehrpersonen des Zyklus 3 summatives Beurteilen – in diesem Zusammenhang betonen sie eine starke Ausrichtung an Stellwerk und Multicheck –, v. a. mit Blick auf weiterführende Schulen. Beklagt wird dabei, dass die Aufnahmeprüfungen zu wenig auf den Lehrplan abgestimmt seien.

Eine Lehrperson fügte an, dass sie formatives Beurteilen als zunehmend wichtig erachte, jedoch nicht wisse, wie dies umzusetzen sei.

Mit Blick auf kommende Weiterbildungen wünschen sich einige vor allem Unterstützung zum formativen Beurteilen, da zum summativen Beurteilen vieles vorhanden sei, gerade wenn man eng mit einem Lehrmittel arbeite. Allerdings fügen sie an, dass die Frage nach Effizienz beim summativen Beurteilen für sie sehr wichtig wäre und sie sich dazu mehr Hilfestellungen wünschen. Gleichzeitig merken sie an, dass der Begriff «Beurteilung» extrem unterschiedlich besetzt sei und eine Klärung im Team notwendig wäre.

Auffällig ist, dass die Lehrpersonen in den Interviews mehrheitlich situative Beobachtungen und Beurteilungen hervorheben (in der Regel als «kleine Beurteilungsformen» bezeichnet). Im Sinne kurzfristiger Entscheidungen, wie sie in Kap. 2 ausgeführt werden, nennen sie Peer-Feedback als festes Element oder auch Lernjournale. Lediglich eine Lehrperson erwähnt, dass sie auch mehrere gezielte Beobachtungen durchführe, die sie dann zusammenfasse. Gegenüber standardisierten Verfahren zeigen sie sich mehrheitlich sehr skeptisch, da sie diese so einschätzen, dass bei den Schülern und Schülerinnen «Gleiches» vorausgesetzt werde.

Hinsichtlich des Lehrplans 21 sehen die Lehrpersonen keine spezifischen Herausforderungen für QUIMS-Schulen, da dieser für alle gleich sei. Teilweise wird angemerkt, dass der Bereich Hören schwierig zu beurteilen sei, gerade mit Blick auf das Zeugnis. Des Weiteren wird erwähnt, dass die Zyklen 1–3 für sie teilweise zu grosse Zeiträume umfassen, weshalb sie sich pro Klasse spezifische Kompetenzraster wünschen. So fügt eine Lehrperson an, dass sie selbst deshalb den Lehrplan 21 für jede Klasse «übersetzt» hätten). Erwähnt wird im Zusammenhang mit dem Lehrplan 21 auch die Online-Weiterbildung der Pädagogischen Hochschule Zürich, die als hilfreich, aber auch ausbaufähig eingeschätzt wird.

An mehreren Stellen kommen die Lehrpersonen auf Rechtschreibung und Grammatik zu sprechen. Dabei zeigen sich zwei sehr konträre Positionen:

- a) Beides habe oftmals ein zu grosses Gewicht im Schulalltag, was nicht dem aktuellen didaktischen Stand entspreche. Wer allerdings eng mit einem Lehrmittel arbeite, habe eher ein vernünftiges Mass.
- b) Insbesondere Grammatik sei für DaZ-Schüler/-innen zentral. Da müsse bspw. gerade in ihren Texten jeder Satz stimmen. Hier könne das Gewicht der Grammatik nicht unterschätzt werden.

Hinsichtlich Lehrmittel betonen einzelne, dass die Arbeit mit einem offiziellen Lehrmittel zentral und hilfreich sei, dass das Zusammenstellen verschiedenster Materialien – aus dem Internet, aus mehreren Lehrmitteln, aus eigener Produktion – eher problematisch sei. Betont wird, dass in künftigen Weiterbildungen genauer aufgezeigt werden sollte, was gute Lehrmittel oder Hilfsmaterialien wären und was es an Beurteilungsmöglichkeiten gebe.

Schwierigkeiten werden mit Blick auf den neuen Schwerpunkt folgende genannt:

- Oft sei es schwierig, einzuschätzen, ob ein Sprachproblem oder ein anderweitiges Problem vorliege.
- Die Gewichtung der einzelnen Kompetenzbereiche (Hören, Lesen, Sprechen, Schreiben, Sprache(n) im Fokus, Literatur im Fokus) wirft Fragen auf.
- Fächer wie NMG und Deutsch seien zu stark voneinander abgekoppelt. Die Lehrmittel im Bereich NMG seien sprachdidaktisch nicht gut aufbereitet.
- Die Vernetzung von DaZ und Deutschdidaktik wird als unzureichend erlebt.

In den Interviews wird der Bezug von Beurteilung und eigenem Unterricht kaum hergestellt. Eine Lehrperson meint denn auch, dass sie streng genommen nicht testen müsse, da sie eigentlich ganz genau wisse, was ihre Schüler/-innen könnten.

Die Frage, wie vorhandene Instrumente oder Materialien sprachbewusst adaptiert werden, wird mehrheitlich mit der Zusammenarbeit von Klassen-Lehrperson und schulischen Heilpädagoginnen und Heilpädagogen bzw. DaZ-Lehrpersonen beantwortet. Ebenso wird erwähnt, dass besonders Aufträge grosse Herausforderungen darstellen und deshalb besonders gut aufbereitet sein müssten. Dennoch brauche es gerade auch bei Aufträgen Worterklärungen. Eine Lehrperson erwähnt, dass schwierige Aufträge oder Anweisungen vor allem dann auffallen, wenn man anderen Lehrpersonen zuschauen; selber sei das schwierig zu merken.

Die befragten Lehrpersonen der Sekundarstufe I betonen, dass sie durch ihre Ausbildung nicht ausreichend dafür vorbereitet worden wären. Die sprachbewusste Adaption von Aufgaben oder Aufträgen sei zudem viel zu aufwendig. Sie merken auch an, dass sie verstehen, wenn bspw. eine Mathematik-Lehrperson der Auffassung sei, dass sie mit ihren Schülern und Schülerinnen nicht auch noch schreiben müsse, was aber vielleicht auch auf fehlendes Wissen oder Verstehen zurückzuführen sei.

Mit Blick auf schwächere Schüler/-innen wird erwähnt, dass der Nachteilsausgleich das Beste wäre. Andere Formen der Instruktion oder zusätzliche Unterstützungsmassnahmen werden nicht erwähnt.

Auffällig ist, dass die Lehrpersonen auf die Frage, wie mögliche Verzerrungen bei der Beurteilung vermieden werden könnten, mehrheitlich mit Beispielen reagieren, in denen die Schüler/-innen nicht die Leistung zeigen, die sie zeigen könnten. Gründe werden auf der persönlichen Ebene angesiedelt (psychische Probleme, Bequemlichkeit u. Ä.) oder auf zu geringe Unterstützung durch die Lehrperson zurückgeführt. Eine Lehrperson erwähnt, dass Stellwerktest, Multicheck oder die gymnasialen Aufnahmeprüfungen nicht auf multilinguale Schüler/-innen ausgerichtet seien.

Austausch wie auch Absprachen im Team, in der Schule gibt es gemäss den Auskünften der Lehrpersonen kaum, wird auch als «heikles Thema» erlebt. Zudem finde der Austausch nicht organisiert, sondern mehr zufällig statt. Insgesamt wünschen sich die befragten Lehrpersonen eine weitaus stärkere Rolle der jeweiligen Schulleitung. Das wird nicht zuletzt mit Blick auf Kollegen und Kolleginnen erwähnt, die anders arbeiten, die etwas falsch verstehen oder anwenden würden (z.B. Scaffolding). Betont wird, dass sie nicht wüssten, wie sie sich in solchen Fällen verhalten sollten. Da sei es nicht ihre Aufgabe, Kollegen oder Kolleginnen mitzuteilen, wie sie es (nicht) machen sollen. Als ein positives Beispiel wird erwähnt, dass sie gemeinsam an Kompetenzrastern gearbeitet hätten, was zu einem besseren gemeinsamen Verständnis geführt habe.

Eine Form des Austauschs ist die Notenkonferenz. Der Austausch kann aber auch darin bestehen, dass sie in pädagogischen Teams beschliessen, dieselbe Prüfung durchzuführen. So könnten sie besser Erfahrungen und Ergebnisse austauschen. Auch mit Projekten in Jahrgangsstufen wurden gute Erfahrungen gemacht:

Wir wollten eine Form von Feedback-Gesprächen finden, wie wir das regelmässig machen können. Wir haben es so organisiert: Jede Schülerin, jeder Schüler hat alle 6 Wochen die Möglichkeit, 10–15 Min. – immer nach dem gleichen Schema – mit der Lehrperson zu sprechen. Haben wir jetzt ein Jahr lang gemacht, war super! Wir haben viel erfahren über die Kinder.

Mit Blick auf die kommenden Weiterbildungen wird den Lehrpersonen der Zyklen 1 und 2 erwähnt, dass unbedingt das ganze Kollegium daran beteiligt sein sollte. Gerade wenn nicht alle von Anfang an dabei seien, sei es schwierig, alle auf den gleichen Stand zu bringen. Zudem wird betont, dass eine verpflichtende Weiterbildung wünschenswert sei, da dies Verbindlichkeiten schaffe, die auch eingefordert werden können. Des Weiteren sei wie im Schwerpunkt Schreiben eine gute Begleitung bei Weiterbildungen wichtig sowie Musteraufgaben sehr hilfreich.

Die Lehrpersonen des Zyklus 3 halten auch in einem ersten Zugang die Arbeit mit Pilotgruppen für eine gangbare Alternative, und erst in einem zweiten Schritt den Einbezug des ganzen Teams.

Weitere Nennungen sind:

- Klärung von «individuell», «Binnendifferenzierung»
- Das Team bestärken, auch durch Befunde aus der Forschung (was ist gut, was weniger).
- Den Lehrpersonen Mut machen, dass man bei einzelnen Schülerinnen und Schülern nur bis zu einem bestimmten Punkt komme.
- Ausreichend Zeit für die Umsetzung einplanen, Coaching und Unterstützung durch externe Fachpersonen während der Umsetzung.

9.2 Die Sicht von Schulleitungen

Die Frage, was eine gute Beurteilung ausmache, wird mehrheitlich mit Verweis auf ganzheitliches Beurteilen beantwortet. Zum ganzheitlichen Beurteilen wird einerseits die

familiäre Herkunft gerechnet, was ein Schüler oder eine Schülerin von zuhause mitbringe, andererseits der Vergleich mit früheren Leistungen der Schüler/-innen. Hervorgehoben wird, dass es dabei nicht um das arithmetische Mittel aus verschiedenen Noten gehen könne oder dass es mit der vorhandenen Schülerschaft besonders schwierig sei. Besonders einfach sei gutes Beurteilen, wenn man in Kompetenzrastern Kreuzchen setzen könne.

Eine Schulleitung betont, das vorgängige Festlegen von Lernzielen sei das A und O, dass erst dies zu einer Transparenz bei der Beurteilung führe (dazu gehören vor allem auch Beobachtungen und weniger Prüfungen). Dabei sei es schwierig, den Prozess einzubeziehen. Wenn ein Schüler gut gearbeitet habe, aber das Lernziel nicht erreiche, könne man das im Zeugnis nicht abbilden.

Der Austausch findet vor allem innerhalb der Stufen, der pädagogischen Teams oder in den Fachschaften statt. Dabei steht jeweils die Notengebung im Vordergrund, Absprachen zum Zeugnis, die Einteilung von Schülerinnen und Schülern in Leistungsgruppen. Hospitationen scheinen eher selten und im freiwilligen Rahmen eingesetzt zu werden.

Damit ein Austausch und Abmachungen möglich seien, müsse die Schulleitung auch die strukturellen Voraussetzungen schaffen, wie eine Schulleiterin betont. Geeignet seien dazu Sitzungsgefäße, in denen die Lehrpersonen gemeinsam Unterricht vor- und nachbereiten. So könne zum einen Verbindlichkeit hergestellt, zum anderen seitens Schulleitung auch eingefordert werden.

Es fällt auf, dass die Rolle der Schulleitung sehr unterschiedlich gesehen wird: Während wie im zuvor erwähnten Beispiel die Schulleitung eine starke Rolle übernimmt, die Verantwortung auch bei sich sieht und stark steuert, sehen zwei Schulleiter/-innen die Hauptverantwortung in erster Linie bei den QUIMS-Beauftragten. Eine Person meint zudem, dass

eine Schulleitung grundsätzlich nur etwas anstossen könne, sodass die Lehrpersonen selbst einsehen, weshalb etwas so sei.

Das Erarbeiten eines gemeinsamen Verständnisses zum Beurteilen fand in einer Schule im Rahmen einer zweijährigen Weiterbildung statt. Zwei Schulen haben sich dieses Themas noch gar nicht angenommen, eine Schule möchte das Beurteilen vor allem in Bezug zu SOL, was in den letzten zwei Jahren ein wichtiger Schwerpunkt war, vertiefen.

Von besonderem Interesse ist der mehrfache Hinweis, dass es beim Austausch oftmals um schwache bzw. auffällige Schüler/-innen gehe, dass dabei den schulischen Heilpädagoginnen und Heilpädagogen (= SHP) eine besonders zentrale Rolle zukomme. Teilweise werden Lernstandserfassungen zudem von SHPs initiiert, die dann auch in allen Klassen durchgeführt und von den SHPs ausgewertet werden. So wird wiederholt erwähnt, dass mehr Informationen zur Zusammenarbeit von DaZ-Lehrpersonen, IF-Lehrpersonen, schulischen HeilpädagogInnen und Klassenlehrpersonen wichtig wären.

Bei schwachen Schülern und Schülerinnen werden angepasste Lernziele in Erwägung gezogen, auch DaZ-Intensivkurse. Eine Person meint jedoch, man könne es auch einfach stehen lassen, da sowieso alles ein zweites Mal durchgenommen werde. Dagegen betont eine Befragte, dass die Verantwortung bei der Schule liege, dass es ihre Aufgabe sei, Settings so zu gestalten, dass alle Schüler/-innen die Lernziele erreichen könnten. Dazu müsse man aber genauer klären, wo die Schwierigkeiten liegen. «Sonderschulungen» seien nur als letzte Möglichkeit in Betracht zu ziehen. Darüber hinaus dürften die begabteren Schüler/-innen nicht vergessen gehen: Auch diese brauchen spezifische Angebote.

Prinzipiell wird der Lehrplan 21 nicht als spezifische Herausforderung für QUIMS-Schulen gesehen. Die Herausforderung liege vielmehr darin, Schüler/-innen mit Migrationshintergrund nicht per se schlechter einzuschätzen.

QUIMS wird als integratives Programm verstanden, das nicht nur Schüler/-innen mit Migrationshintergrund besonders fokussiere, sondern allgemein Schüler/-innen mit höherem Förderbedarf. Kritisch erwähnt wird mit Blick auf den Lehrplan 21 lediglich, dass unklar sei, welcher Spielraum bei der Beurteilung für DaZ-Schüler/-innen bestehe.

Besondere Herausforderungen stellen sich aus Sicht der Befragten mit Blick auf den Übertritt von der Mittel- in die Sekundarstufe und mit Blick auf die Einteilung in Leistungsgruppen auf der Sekundarstufe I. Eine Schulleitung äussert in diesem Zusammenhang die Hoffnung, dass mit dem neuen Schwerpunkt C auch Tools zur Verfügung gestellt werden – im Sinne eines «Kompetenzkastens» –, die klare Zuordnungen erlauben.

Mit Blick auf den Weiterbildungsstart ab Sommer 2019 werden folgende Aspekte genannt:

- Sprache sei schwieriger zu beurteilen als andere Bereiche, hier sei ein Schwerpunkt zu setzen.
- Aha-Erlebnisse für Lehrpersonen seien zentral.
- Die Gestaltung von Zeugnissen sei zu thematisieren, da es zu zeitintensiv sei, eigene Normen zu erarbeiten.
- Mehr Informationen zur Zusammenarbeit aller Beteiligten.
- Die bisherigen QUIMS-Weiterbildungen hätten sich sehr bewährt, vor allem der zweijährige Ansatz (inkl. Umsetzungen), den es so nur in QUIMS gäbe.
- Die WB müsse klar sagen, wie Sprache sein müsse, wie man formulieren müsse, sodass es alle Schüler/-innen verstehen würden.
- Die WB dürfe kein Standardprogramm sein.
- Es brauche kompetente Referenten und Referentinnen, die Theorie vermitteln könnten, damit man das Zentrale verstehe, die aber auch Praxis vermitteln

könnten, damit es die Lehrpersonen umsetzen könnten.

- Die Lehrpersonen dürften bei einer WB zum Thema Beurteilen nicht den Eindruck erhalten, sie hätten bisher alles falsch gemacht.
- Eine WB solle auch nicht belehren, sondern sie müsse Angebote für Lehrpersonen schaffen (das habe sich sehr bewährt).

9.3 Fazit

Mit Blick auf die in Kapitel 1 formulierten Empfehlungen und Themenfelder lässt sich auf Basis der vereinfachten Auswertung der Interviews Folgendes festhalten:

- In den Aussagen der Lehrpersonen finden sich Bezüge zu nahezu allen Empfehlungen. Empfehlung 9 wurde aufgrund dieser Interviews ergänzt.
- Besonders zu beachten ist mit Blick auf SchilW, dass – bezogen auf E7, E9 sowie E10 – auch die Gewichtung der einzelnen Sprachlernbereiche anzugehen ist.
- Im Rahmen von E5 sind vorausplanenden Beobachtungen und Beurteilungen ein hohes Gewicht einzuräumen. Dabei muss allenfalls auch geklärt werden, was schulalltagsnahe Formen sind, was standardisierte Verfahren sind, was sie leisten können, was auch nicht.
- Mit Blick auf E8 (Wortschatz) ist einer Sichtweise vorzubeugen, die die sprachbewusste Gestaltung von Unterricht auf «Oberflächenmerkmale» bzw. sog. textseitige Massnahmen reduziert.
- Die Rolle der Schulleitung ist – gerade auch wenn man die Aussagen der Lehrpersonen und Schulleitungen aufeinander bezieht – zu klären und vor allem zu stärken (E3). Mit Blick auf SchilW ist hier noch genauer zu überlegen, wie sichergestellt werden kann, dass die Schulleitungen

beim Thema «Beobachten und Fördern mit Fokus auf Sprache» aktiv steuern.

- Bei der Frage, inwiefern mögliche Verzerrungen bei Beobachtungen bzw. Beurteilungen vermieden werden können, wirkten die Befragten weitgehend ratlos. Zudem zeigte sich, dass sie mögliche Verzerrungen vor allem so interpretieren, dass Schüler/-innen nicht die Leistung erbringen, die erwartbar wäre. Dies müsste bei den kommenden schulinternen Weiterbildungen auf jeden Fall berücksichtigt werden.

Bei der kommunikativen Validierung, die am 29. November 2018 stattfand, wurden die Empfehlungen und Themenfelder, wie sie in Kapitel 1 erläutert sind, begrüsst. Insbesondere wurde den Empfehlungen E2 (forschende Haltung aufbauen) und E8 (Wortschatz fördern und beurteilen) sowie dem Themenfeld T4 (in allen Fächern sprachbewusst fördern und beurteilen) ein hohes Gewicht für den QUIMS-Schwerpunkt C eingeräumt.

Kritisch angemerkt wurde, dass prognostisches Beurteilen in den Empfehlungen und Themenfeldern nicht vorkomme. Gerade für DaZ-Lehrpersonen sei dies jedoch ein wichtiger Aspekt, so etwa wenn es darum gehe zu entscheiden, ob ein Schüler bzw. eine Schülerin von der Aufnahme- in die Regelklasse umgeteilt werden könne. Zudem würden oft Schüler/-innen mit Deutsch als Zweitsprache beim Übertritt in die Sekundarstufe I eher in Niveau B statt A eingeteilt, obwohl sie das Potenzial für Niveau A hätten. (Anmerkung: Beides sind Aspekte, die im Themenfeld T1 «Beobachtungen und Beurteilungen planen und durchführen» thematisiert werden können.)

Kritisch angemerkt wurde, dass für Sekundarschulen geprüft werden müsse, inwiefern sich T1–T3 auch für alle Fächer eignen, inwiefern sämtliche Lehrpersonen – also nicht nur Deutsch-Lehrpersonen – in entsprechende Weiterbildungen integriert werden können. Dies wurde in Kapitel 1.2 entsprechend berücksichtigt.

10 Anhang

10.1 Gruppeninterviews mit Lehrpersonen – Leitfaden

Ebene Unterricht
Thema ist das Beurteilen: Was ist aus eurer Sicht eine gute Beurteilungspraxis? Achtung: Hier geht es noch nicht um die spezifischen Herausforderungen für QUIMS-Schulen.
Wir haben euch zur Vorbereitung die Ziele des neuen Schwerpunkts auf der Ebene des Unterrichts geschickt. Das kompetenzorientierte Unterrichten und Beurteilen hat mit dem Lehrplan 21 ein neues Gewicht erhalten. Inwiefern stellen sich für QUIMS-Schulen aus eurer Sicht spezifische Herausforderungen? Welcher Art? Wo seht ihr die grössten Herausforderungen für euch? Wo seht ihr Chancen / Potenziale?
Der Lehrplan 21 unterscheidet verschiedene Lernbereiche: Hören, Sprechen, Lesen, Schreiben sowie Sprache(n) im Fokus und Literatur im Fokus. Seht ihr mit Blick auf euren Unterricht, für QUIMS-Schulen für einzelne Bereiche bezogen auf den neuen Schwerpunkt grössere Herausforderungen und wenn ja warum?
Es ist oft die Rede davon, dass Beobachtungen zum Lernen der SuS bzw. Beurteilungen von Leistungen verzerrt sein können. Könnt ihr 1–2 Beispiele nennen, die solche Verzerrungen sind? Wie können diese Verzerrungen aus eurer Sicht vermieden werden?
Das fachliche Lernen setzt auch sprachliche Fähigkeiten voraus. Zudem muss beim fachlichen Lernen auch fachspezifisches literales Wissen erworben werden (Wortschatz, Textsorten u.a.). Bei den Zielen zum QUIMS-Schwerpunkt «Beurteilen und Fördern» heisst es hierzu, dass man als Lehrperson über das dazu notwendige sprachdidaktische Wissen verfügen müsste. Wie gut fühlt ihr euch dazu gerüstet?
Vorhandene Instrumente – das können auch Prüfungen sein – oder Materialien sollen bei Bedarf von Lehrpersonen auch sprachbewusst adaptiert werden können. Wenn ihr an euren Unterricht denkt: Wie adaptiert ihr? Was hat sich bewährt, was weniger?
Im Rahmen des Schwerpunkts «Schreiben auf allen Schulstufen» wurden zahlreiche Musteraufgaben erarbeitet, die nicht nur Aufträge für die SuS enthielten, sondern auch den Lehrpersonen aufzeigen sollten, wie bestimmte Förderansätze funktionieren und umgesetzt werden können. Wenn ihr jetzt an den neuen Schwerpunkt denkt und dass wiederum Materialien erarbeitet werden sollen: Was würdet ihr euch zuallerst wünschen?
Ebene der Schule
Welche Form von Austausch habt ihr in eurer Schule zu Beurteilungen?
In welcher Hinsicht habt ihr als Schule, als Kollegium ein gemeinsames Verständnis rund ums Beurteilen (mit Fokus auf Sprache)?
Gibt es in eurem Team, eurem Schulhaus bereits Abmachungen zum «Beurteilen und Fördern mit Fokus auf Sprache»? Wenn ja, welche? Wie geht ihr damit um?
Schulinterne Weiterbildung
Ab Sommer 2019 können die ersten Schulen mit einer schulinternen Weiterbildung zu diesem Schwerpunkt beginnen. Welche Erwartungen hättet ihr an diese WB? ... wenn ihr an euch selbst denkt, ... wenn ihr an eure Schule als Ganzes denkt?
Abschluss
Gibt es etwas, was wir bisher nicht angesprochen oder diskutiert haben, was aber mit Blick auf die schulinternen WBs in den QUIMS-Schulen noch wichtig wäre?

10.2 Einzelinterviews mit Schulleitungen – Leitfaden

Ebene Schule
Thema ist das Beurteilen: Was ist aus deiner Sicht eine gute Beurteilungspraxis bezogen auf den Unterricht? (Achtung: Hier geht es noch nicht um die spezifischen Herausforderungen für QUIMS-Schulen.)
Welche Form von Austausch oder Abmachungen habt ihr in deiner Schule zu Beurteilungen?
Das kompetenzorientierte Unterrichten und Beurteilen hat mit dem Lehrplan 21 ein neues Gewicht erhalten. Inwiefern stellen sich für QUIMS-Schulen aus deiner Sicht spezifische Herausforderungen? Welcher Art? Wo siehst du die grössten Herausforderungen für deine Schule und deine Lehrpersonen?
In welcher Hinsicht habt ihr als Schule, als Kollegium ein gemeinsames Verständnis rund ums Beurteilen (mit Fokus auf Sprache)?
Gibt es ein Thema, eine Frage, einen Aspekt, der euch als Schule mit Blick auf den neuen Schwerpunkt der Beurteilung besonders beschäftigt? Warum?
Wie geht ihr mit SuS um, die die Grundansprüche nicht erreichen? Inwiefern ist das für deine Schule ein Problem, das die Schule lösen kann und soll?
Thema ist das Beurteilen: Was ist aus deiner Sicht eine gute Beurteilungspraxis bezogen auf den Unterricht? (Achtung: Hier geht es noch nicht um die spezifischen Herausforderungen für QUIMS-Schulen.)
Ebene des Unterrichts
Mit Blick auf das Team: Welche Unterstützung wäre für dein Team wichtig / sinnvoll, um die Herausforderungen rund ums Beurteilen besser meistern zu können?
Mit Blick auf dich selbst als SL: Was ist für dich wichtig, um die Herausforderungen rund ums Beurteilen mit deinem Team und deiner Schule (inkl. Eltern) meistern zu können? Welche Unterstützung wäre für dich wichtig?
Inwiefern werden Beobachtungen / Beurteilungen von den Lehrpersonen auch genutzt, um daraus Konsequenzen für den Unterricht abzuleiten? (Allenfalls erläutern: andere Fördermethode ...)
Schulinterne Weiterbildung
Ab Sommer 2019 können die ersten Schulen mit einer schulinternen WB zu diesem Schwerpunkt der Beurteilung beginnen. Welche Erwartungen hättest du in Bezug an diese WB?
Abschluss
Gibt es etwas, was wir bisher nicht angesprochen oder diskutiert haben, was aber mit Blick auf die schulinternen WBs in den QUIMS-Schulen noch wichtig wäre?

10.3 Ziele des Schwerpunkts C

Ziele auf Ebene der Schule (Schulprogramm, Praxis der gesamten Schule):

1. Die QUIMS-Schulen erfüllen grundlegende Standards¹ einer sowohl formativen wie auch summativen Beurteilung und Förderung, die sich auf die Sprache fokussieren und die Herausforderungen im multikulturellen Kontext berücksichtigen. Zu den spezifischen Herausforderungen gehört insbesondere der grosse Anteil an Lernenden aus sozial benachteiligten, mehrsprachigen und zugewanderten Familien.
2. Der Schwerpunkt «Beurteilen und Fördern mit Fokus auf Sprache» ist im Schulprogramm verankert.
3. Die Lehrpersonen des Schulkollegiums haben ein gemeinsames Verständnis der Beurteilung und Förderung und arbeiten zu diesen Themen zusammen.
4. Das Konzept der Beurteilung und Förderung ist im Schulalltag sowie in den Fächern, insbesondere in Deutsch, sichtbar.

Ziele auf Ebene des Unterrichts (alle Lehrpersonen, inklusive Fach- und DaZ-Lehrpersonen):

5. Die Lehrpersonen wissen, was kompetenzorientiertes Unterrichten und Beurteilen heisst, und wenden dieses Wissen unter Berücksichtigung der spezifischen Herausforderungen in einem mehrsprachigen und sozial benachteiligten Umfeld an.
Das heisst:
 - a. Die Lehrpersonen aller Schulstufen und Fächer wissen, wie sie gemäss kantonalen Vorgaben kompetenzorientiert beurteilen sollen. Sie kennen bezüglich der Beurteilung insbesondere die unterschiedlichen Funktionen (formativ, summativ, prognostisch), Bezugsnormen (Sach-, Individual- und Sozialnorm) und Formen (Selbsteinschätzung, Dialog / Feedback, Bericht, Noten usw.).
 - b. Sie kennen Methoden, mit denen sie Verzerrungen von Beurteilungen minimieren können, und wenden diese an.
 - c. Sie verfügen über das für ihr Fach notwendige sprachdidaktische Wissen, um das fachliche Lernen sprachbewusst beurteilen und fördern zu können.
 - d. Sie wenden dieses Wissen sowohl bei formativen wie auch summativen Rückmeldungen an.
 - e. Sie können Lernaufgaben und Prüf-/Testaufgaben sprachbewusst konzipieren.
 - f. Sie können vorhandene Instrumente – Lehrmittel, Diagnoseinstrumente, Lernstandserfassungen u.Ä. – in Bezug auf Sprache analysieren und bei Bedarf sprachbewusst adaptieren.
- 6) Sie können ihre formativen und summativen Beurteilungen für die Schülerinnen und Schüler sprachlich so formulieren, dass die Beurteilungen deren Lernen wirksam unterstützen.

11 Literatur

- Adamina, M. (2010). Lernen begleiten, begutachten und beurteilen. In P. Labudde (Hrsg.), *Fachdidaktik Naturwissenschaft. 1.–9. Schuljahr* (S. 181–196). Bern: Haupt.
- Akkus, R., Gunel, M. & Hand, B. (2007). Comparing an inquiry-based approach known as the science writing heuristic to traditional science teaching practices: Are there differences? *International Journal of Science Education*, 29 (14), 1745–1765.
- Bangert-Drowns, R. L., Hurley, M. M. & Wilkinson, B. (2004). The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement: A Meta-Analysis. *Review of Educational Research*, 74 (1), 29–58.
- Beck, I. L., McKeown, M. G. & Kucan, L. (2013). *Bringing words to life* (2nd Auflage). New York/London: Guilford Press.
- Beerenwinkel, A. & Sturm, A. (2018). Schreiben als konstitutiver Bestandteil im naturwissenschaftlichen Unterricht: Grenzen und Möglichkeiten. Bochum.
- Behrmann, L. & Glogger-Frey, I. (2017). Produkt- und Prozessindikatoren diagnostischer Kompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 134–142). Münster, New York: Waxmann.
- Behrmann, L. & Souvignier, E. (2013). The Relation Between Teachers' Diagnostic Sensitivity, their Instructional Activities, and their Students' Achievement Gains in Reading*. *Zeitschrift für Pädagogische Psychologie*, 27 (4), 283–293.
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18 (1), 5–25.
- Bildungsdirektion Kanton Zürich (Hrsg.). (2009). *Qualitätsmerkmale zu den Handlungsfeldern «Förderung des Schulerfolgs», «Förderung der Integration»*. Ein Arbeitsinstrument für die Schulentwicklung (3.). Zürich: Bildungsdirektion Kanton Zürich.
- Bildungsdirektion Kanton Zürich. (2018). *Beurteilung und Schullaufbahnentscheide. Über das Fördern, Notengeben und Zuteilen*. Zürich: Lehrmittelverlag des Kantons Zürich.
- Bildungsdirektion Kanton Zürich (Hrsg.). (2017). *Natur, Mensch, Gesellschaft. Lehrplan für die Volksschule des Kantons Zürich auf der Grundlage des Lehrplans 21*. Zürich: Bildungsdirektion Kanton Zürich. Zugriff am 1.2.2017. Verfügbar unter: zh.lehrplan.ch
- Birkel, P. (2003). Aufsatzbeurteilung - ein altes Problem neu untersucht. *Didaktik Deutsch*, 9 (15), 46–63.
- Birkel, P. (2009). Rechtschreibleistung im Diktat - eine objektiv beurteilbare Leistung? *Didaktik Deutsch*, 15 (27), 5–32.
- Bitchener, J., Young, S. & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, (14), 191–205.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21 (1), 5–31.
- Bochnik, K. & Ufer, S. (2017). Fachsprachliche Kompetenzen im sprachsensiblen Mathematikunterricht der Grundschule. Implikationen einer Studie zur sprachbezogenen Analyse mathematischer Leistungsunterschiede zwischen Lernenden mit deutscher und nicht-deutscher Familiensprache (Sprachliche Bildung). In D. Leiss, M. Hagen, A. Neumann & K. Schwippert (Hrsg.), *Mathematik und Sprache. Empirischer Forschungsstand und unterrichtliche Herausforderungen* (S. 81–98). Münster, New York: Waxmann.
- Bohl, T. (2001). Theoretische Strukturierung - Begründung neuer Leistungsbeurteilung. In H.U. Grunder & T. Bohl (Hrsg.), *Neue Formen der Leistungsbeurteilung in den Sekundarstufen I und II* (S. 9–49). Baltmannsweiler: Schneider Verlag Hohengehren.
- Bölsterli, J., Niederbacher, E., Stotz, T. & Neuenchwander, M. P. (2018, Juni 28). Bildungschancen in sozial heterogenen Schulklassen fördern. SGBF Zürich.
- Bong, M. & Skaalvik, E. M. (2003). Academic Self-Concept and Self-Efficacy: How Different Are They Really? *Educational Psychology Review*, 15 (1), 1–40.
- Brühwiler, C. (2014). *Adaptive Lehrkompetenz und schulisches Lernen: Effekte handlungssteuernder Kognitionen von Lehrpersonen auf Unterrichtsprozesse und Lernergebnisse der Schülerinnen und Schüler*. Waxmann.

- Brühwiler, C. (2017). Diagnostische und didaktische Kompetenz als Kern adaptiver Lehrkompetenz. In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 123–134). Münster, New York: Waxmann.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Burns, R. B. (1984). The process and context of teaching: A conceptual framework. *Evaluation in Education*, (8), 95–112.
- D-EDK (2016). *Lehrplan 21. Natur, Mensch, Gesellschaft*. https://v-fe.lehrplan.ch/container/V_FE_DE_Fachbereich_NMG.pdf (15.07.2018).
- Deno, S. L. (1985). Curriculum-Based Measurement: The Emerging Alternative. *Exceptional Children*, 52 (3), 219–32.
- Dittmar, M., Schmellentin, C., Gilg, E. & Schneider, H. (2017). Kohärenzaufbau aus Text-Bild-Gefügen: Wissenserwerb mit schulischen Fachtexten. *Leseforum*, (1), 1–19.
- Duarte, J., Gogolin, I. & Kaiser, G. (2011). Sprachlich bedingte Schwierigkeiten von mehrsprachigen Schülerinnen und Schülern bei Textaufgaben (Mehrsprachigkeit). In S. Prediger & E. Özdil (Hrsg.), *Mathematiklernen unter Bedingungen der Mehrsprachigkeit. Stand und Perspektiven der Forschung und Entwicklung in Deutschland* (S. 35–53). Münster New York: Waxmann.
- Ehl, B., Paul, M., Bruns, G., Fleischhauer, E., Vock, M., Gronostaj, A. et al. (2018). Testgütekriterien der „Profilanalyse nach Grießhaber“. Evaluation eines Verfahrens zur Erfassung grammatischer Fähigkeiten von ein- und mehrsprachigen Grundschulkindern. *Zeitschrift für Erziehungswissenschaft*, 1–21. x
- Ennemoser, M., Kuhl, J. & Pepouna, S. (2013). Evaluation des Dialogischen Lesens zur Sprachförderung bei Kindern mit Migrationshintergrund. *Zeitschrift für Pädagogische Psychologie*, 27 (4), 229–239.
- Feser, M. S., & Höttecke, D. (2017). Klassenarbeiten kriteriengeleitet korrigieren – Wie beurteile ich eine Schülererklärung? *Unterricht Physik*, 158, 15–18.
- Feser, M.S., Höttecke, D. & Ehmke, T. (2016). Testitems zur qualitativen Untersuchung der Ressourcen von Physiklehrkräften beim Bewerten schriftlicher Schülerleistungen. *PhyDid B - Didaktik der Physik - Beiträge zur DPG-Frühjahrs-tagung*. <http://www.phydid.de/index.php/phydid-b/issue/view/281> (15.07.2018).
- Fischer, H. E., Labudde, P., Neumann, K. & Viiri, J. (Hrsg.). (2014). *Quality of instruction in physics*. Münster, New York: Waxmann.
- Glogger-Frey, I. & Herppich, S. (2017). Formative Diagnostik als Teilaspekt diagnostischer Kompetenz (Pädagogische Psychologie und Entwicklungspsychologie). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (Band 94, S. 42–46). Münster, New York: Waxmann.
- Graham, S. & Harris, K. R. (2012). The role of strategies, knowledge, will, and skills in a 30-year program of writing research (with a homage to Hayes, Fayol, and Boscolo). In V.W. Berninger (Hrsg.), *Past, present, and future contributions of cognitive writing research to cognitive psychology* (S. 177–196). New York: Psychology Press/Taylor & Francis Group.
- Graham, S., Harris, K. R. & Hebert, M. (2011). *Informing Writing. The benefits of Formative Assessment*. New York: Carnegie Corporation of New York.
- Graham, S., Hebert, M. & Harris, K. R. (2015). Formative Assessment and Writing: A Meta-Analysis. *The Elementary School Journal*, 115 (4), 523–547.
- Groeben, N. (2009). Möglichkeiten und Grenzen des Testens von kultureller Bildung (Leseozialisation und Medien). In A. Bertschi-Kaufmann & C. Rosebrock (Hrsg.), *Literalität. Bildungsaufgabe und Forschungsfeld* (S. 107–124). Weinheim / München: Juventa.
- Hand, B., Norton-Meier, I.A., Gunel, M. & Akkus, R. (2016). Aligning Teaching to Learning: A 3-Year Study Examining the Embedding of Language and Argumentation into Elementary Science Classrooms. *International Journal of Science and Mathematics Education* 14. 847–863.
- Hand, B., Wallace, C. W. & Yang, E.-M. (2004). Using a Science Writing Heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: quantitative and qualitative aspects. *International Journal of Science Education*, 26 (2), 131–149.
- Härtig, H., Bernholt, S., Prechtel, H., & Retelsdorf, J. (2015). Unterrichtssprache im Fachunterricht – Stand der Forschung und Forschungsperspektiven am Beispiel des Textverständnisses.

- Zeitschrift für Didaktik der Naturwissenschaften* 21. 55–67.
- Härtig, H., Pehlke, C., Fischer, H.E. & Schmeck, A. (2012). Sind Fachsprache und Fachwissen bezogen auf Physik unterscheidbar? *Zeitschrift für Didaktik der Naturwissenschaften* 18. 381–390.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement* (1. Auflage). London; New York: Routledge.
- Hattie, J. (2012). *Visible Learning for Teachers: Maximizing Impact on Learning*. London/New York: Routledge Chapman & Hall.
- Hattie, J. & Gan, M. (2011). Instruction based on Feedback. In R.E. Mayer & P.A. Alexander (Hrsg.), *Handbook of Research on Learning and Instruction* (S. 249–295). New York: Taylor & Francis.
- Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77 (1), 81–112.
- Hebbecker, K. & Souvignier, E. (2018). Formatives Assessments im Leseunterricht der Grundschule – Implementation und Wirksamkeit eines modularen, materialgestützten Konzepts. *Zeitschrift für Erziehungswissenschaft*, 1–31. <https://doi.org/10.1007/s11618-018-0834-y>
- Heitzmann, A. (2010). Von der Alltagssprache zur Fachsprache gelangen. In P. Labudde (Hrsg.), *Fachdidaktik Naturwissenschaft. 1.–9. Schuljahr* (73–86). Bern: Haupt.
- Helmke, A. (2009). *Unterrichtsqualität und Lehrprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Seelze: Kallmeyer.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulmanagement und Schulentwicklung* (S. 119–144). Baltmannsweiler: Schneider Verlag Hohengehen.
- Heppt, B., Stanat, P., Dragon, N., Berendes, K. & Weinert, S. (2014). Bildungssprachliche Anforderungen und Hörverstehen bei Kindern mit deutscher und nicht-deutscher Familiensprache. *Zeitschrift für Pädagogische Psychologie*, 28 (3), 139–149.
- Hosp, J. L. (2012). Using assessment data to make decisions about teaching and learning. In K.R. Harris, S. Graham, T.C. Urdan, A.G. Bus, S. Major & H.L. Swanson (Hrsg.), *APA educational psychology handbook, Vol. 3: Application to learning and teaching* (S. 87–110). Washington: American Psychological Association.
- Isler, D., Künzli, S. & Wiesner, E. (2014). Alltagsgespräche im Kindergarten – Gelegenheitsstrukturen für den Erwerb bildungssprachlicher Fähigkeiten. *Schweizerische Zeitschrift für Bildungswissenschaften*, 36 (3), 459–479.
- Jussim, L., Robustelli, S. L. & Cain, T. R. (2009). Teacher expectations and self-fulfilling prophecies. In K.R. Wentzel & A. Wigfield (Hrsg.), *Handbook of motivation at school* (S. 349–380). New York/London: Routledge.
- Karing, C., Pfof, M. & Artelt, C. (2011). Is there a relationship between lower secondary school teacher judgment accuracy and the development of students' reading and mathematical competence? *Journal for Educational Research Online / Journal für Bildungsforschung Online*, 3 (2), 119–147.
- Karst, K. (2017). Akkurate Urteile - die Ansätze von Schrader (1989) und McElvany et al. (2009). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 21–25). Münster, New York: Waxmann.
- Karst, K. & Förster, N. (2017). Ansätze zur Modellierung diagnostischer Kompetenz (Pädagogische Psychologie und Entwicklungspsychologie). In A. Südkamp & A.-K. Praetorius (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (Band 94, S. 19–29). Münster, New York: Waxmann.
- Kingston, N. & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30 (4), 28–37.
- Klug, J. (2017). Ein Prozessmodell zur Diagnostik und Förderung von selbstreguliertem Lernen. In A. Südkamp (Hrsg.), *Diagnostische Kompetenz von Lehrkräften* (S. 54–59). Münster, New York: Waxmann.
- Koch, H., von Grünigen Mota Campos, S., Diener, M., Gürber, E., Suter, B., Blanc, B. et al. (2017). *Kompetenzorientiert beurteilen*. Zürich: Bildungsdirektion Kanton Zürich.
- Koch, H., Rüttimann, D., Fiechter, M., Keller, M., Hefti, K. & Hollenweger, J. (2017). *Kompetenzorientiert unterrichten. Einblicke. 2., leicht überarbeitete Auflage*. Zürich: Bildungsdirektion Kanton Zürich.

- Köller, O. (2004). *Konsequenzen von Leistungsgruppierungen*. Münster, New York: Waxmann.
- Konsortium HarmoS Naturwissenschaften+ (2008). *HarmoS Naturwissenschaften+. Kompetenzmodell und Vorschläge für Bildungsstandards. Wissenschaftlicher Schlussbericht*. Unveröffentlicht.
- Koster, M., Tribushinina, E., de Jong, P. F. & van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research*, 7 (2), 249–274.
- Kunz, P., Colberg, C., Bäuml, E., Beerenwinkel, A., Bernhard, F., Labudde, P., Ludwig-Petsch, K., Metzger, S., Sieber-Suter, B., Stübi, C., Wagner, U. & Wilhelm, M. (2016). SWiSE-Kompetenzrahmen: Was müssen Lehrpersonen für gutes naturwissenschaftlich-technisches Unterrichten können? In S. Metzger, C. Colberg & P. Kunz (Hrsg.). *Naturwissenschaftsdidaktische Perspektiven* (217-233). Bern: Haupt.
- Leisen, J. (2008). Lesen ist schon schwer genug, dann auch noch Schreiben? *Naturwissenschaften im Unterricht Physik* 19 (104). 4–10.
- Leisen, J. (2011). Praktische Ansätze schulischer Sprachförderung – Der sprachensible Fachunterricht. www.hss.de/download/111027_RM_Leisen.pdf (15.07.2018).
- Lindauer, T., Schmellentin, C., Beerenwinkel, A., Hefti, C. & Furger, J. (2013). *Sprachbewusst unterrichten. Eine Unterrichtshilfe für den Fachunterricht* (Bildungsraum Nordwestschweiz). Brugg: Pädagogische Hochschule FHNW.
- Lipowsky, F. & Rzejak, D. (2012). Lehrerinnen und Lehrer als Lerner – Wann gelingt der Rollentausch? Merkmale und Wirkungen wirksamer Lehrerfortbildungen. *Schulpädagogik heute*, 3 (5).
- Lötscher, H., Tanner Merlo, S. & Joller-Graf, K. (2017). *Beurteilung in integrativen Schulen. Kompetenzfördernd unterrichten mit dem Lehrplan 21*. Luzern: Pädagogische Hochschule Luzern.
- Mächler, S. (2015). Die QUIMS-Schulen in der Schreibförderung unterstützen – der Beitrag der zentralen Bildungsverwaltung. *Leseforum*, (2). <https://doi.org/www.leseforum.ch>
- McKeown, M. G. & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly*, 29 (4), 520–530.
- McKeown, M. G., Deane, P. D., Scott, J. A., Krovetz, R. & Lawless, R. R. (2017). *Vocabulary Assessment to Support Instruction: Building Rich Word-Learning Experiences*. New York London: Guilford Publications.
- Meyer, M. & Tiedemann, K. (2017). *Sprache im Fach Mathematik* (Mathematik im Fokus). Berlin: Springer.
- Möller, J., Machts, N. & Retelsdorf, J. (2016). Diagnostische Kompetenz von Lehrkräften. *Schulmanagement online*, (4), 14–17.
- Morek, M. & Heller, V. (2012). Bildungssprache – Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs. *zfal*, 57 (1), 67–101.
- Moser, U., Angelone, D., Hollenweger, J. & Buff, A. (2011). *Nach sechs Jahren Primarschule: Deutsch, Mathematik und motivational-emotionales Befinden am Ende der 6. Klasse*. Zürich: Bildungsdirektion.
- Moser, U., Oostlander, J. & Tomasik, M. J. (2017). Soziale Ungleichheiten im Leistungszuwachs und bei Bildungsübergängen. In M.P. Neuenchwander & Nägele, Christof (Hrsg.), *Bildungsverläufe von der Einschulung bis in den ersten Arbeitsmarkt* (S. 59–77). Wiesbaden: Springer VS.
- Moser, U. & Rhy, H. (2000). *Lernerfolge in der Primarschule: Eine Evaluation der Leistungen am Ende der Primarschule*. Aarau: Sauerländer.
- Naumann, J., Artelt, C., Schneider, W. & Stanat, P. (2010). Lesekompetenz von PISA 2000 bis PISA 2009. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel et al. (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 23–71).
- Neuenchwander, M. P. & Nägele, C. (Hrsg.). (2017). *Bildungsverläufe von der Einschulung bis in den ersten Arbeitsmarkt*. Wiesbaden: Springer VS.
- Neugebauer, C. (2015). Schulinterne Weiterbildungen «Schreiben auf allen Schulstufen»: Erste Erfahrungen aus QUIMS-Schulen. *Leseforum*, (2), 1–14.
- Neugebauer, C. (2017). *Begleitmassnahmen für die Schreibförderung an QUIMS-Schulen (Massnahmenplan 2013–2017): Teilauftrag C «Schulinterne Weiterbildungen»*. Auswertungsbericht. Brugg: Pädagogische Hochschule Zürich.

- Oester, K., Fiechter, U. & Kappus, E. N. (2008). *Schulen in transnationalen Lebenswelten: Integrations- und Segregationsprozesse am Beispiel von Bern West*. Zürich: Seismo.
- Philipp, M. (2017). Förderung hierarchiehoher Leseprozesse. In M. Philipp (Hrsg.), *Handbuch Schriftspracherwerb und weiterführendes Lesen und Schreiben* (S. 252–265). Weinheim: Beltz Juventa.
- Ralle, B. (2015). Sprachliche Heterogenität und fachdidaktische Forschung. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im Naturwissenschaftlichen Unterricht* (4–18).
- Rautenstrauch, H. & Busker, M. (2015). Möglichkeiten der (Fach-)Sprachstandserhebung bei Studienanfängern. In S. Bernholt (Hrsg.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im Naturwissenschaftlichen Unterricht* (471–473).
- Renkl, A. (Hrsg.). (2008). *Lehrbuch Pädagogische Psychologie*. Huber, Bern.
- Rheinberg, F. (2006). Bezugsnormorientierung. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarbeitete und erweiterte Auflage., S. 55–62). Weinheim, Basel, Berlin: Beltz.
- Riebling, L. (2013). *Sprachbildung im naturwissenschaftlichen Unterricht. Eine Studie im Kontext migrationsbedingter sprachlicher Heterogenität* (Interkulturelle Bildungsforschung) (Band 20). Münster: Waxmann.
- Rincke, K. (2010). Alltagssprache, Fachsprache und ihre besonderen Bedeutungen für das Lernen. *Zeitschrift für Didaktik der Naturwissenschaften* 16. 235–260.
- Roos, M. (2017). *Entwicklungen von 2014 bis 2016 im Programm QUIMS mit Fokus auf den obligatorischen Schwerpunkten*. Bericht im Auftrag des Volksschulamts des Kantons Zürich. Baar: spectrum GmbH.
- Roos, M., Wandeler, E. & Mosimann, M. (2013). *Das Übertrittsverfahren Primarschule - Sekundarstufe I des Kantons Luzern: Schlussbericht zur externen Evaluation*. Baar: spectrum3.
- Russell, J. & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar. A meta-analysis of the research. In J.M. Norris & L. Ortega (Hrsg.), *Synthesizing research on language learning and teaching* (S. 133–164). Amsterdam: John Benjamins Publishing.
- Sander, A., Ohle, A., McElvany, N., Zander, L. & Hannover, B. (2018). Stereotypenbedrohung als Ursache für geringeren Wortschatzzuwachs bei Grundschulkindern mit Migrationshintergrund. *Zeitschrift für Erziehungswissenschaft*, 21 (1), 177–197.
- Schilcher, A., Röhl, S. & Krauss, S. (2017). Sprache im Mathematikunterricht - eine Bestandesaufnahme des aktuellen didaktischen Diskurses (Sprachliche Bildung). In D. Leiss, M. Hagena, A. Neumann & S. Schwippert (Hrsg.), *Mathematik und Sprache. Empirischer Forschungsstand und unterrichtliche Herausforderungen* (S. 11–42). Münster, New York: Waxmann.
- Schmiemann, P. (2011). Fachsprache in biologischen Testaufgaben. *Zeitschrift für Didaktik der Naturwissenschaften* 11. 115–136.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Lang.
- Schrader, F.-W. (2013). Diagnostische Kompetenz von Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 31, 154–165.
- Schuler, H. (2006). Noten als Prädiktor von Studien- und Berufserfolg. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarbeitete und erweiterte Auflage., S. 535–541). Weinheim, Basel, Berlin: Beltz.
- Senn, W. (2009). Mit HarmoS zu einer neuen Aufgabekultur? Schreibaufgaben im Vergleich: Aufgaben zum Testen – Aufgaben zum Lernen. *Informationen zur Deutschdidaktik*, 3 (33), 88–101.
- Senn, W. (2018). *Schreibmotivation und Schreibziel: Eine Untersuchung zur Konzeption und Modellierung motivationaler Variablen und Schreibziele in Schreibprozess und Schreibprodukt*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Senn, W., Lötscher, H. & Malti, T. (2005). *Selbst- und Fremdbeurteilungsprozesse bei gemeinsam lernenden Schülerinnen und Schülern. Eine Studie zum kooperativen Lernen und Beurteilen in Schreibkonferenzen*. Forschungsbericht No. 2. Luzern: Pädagogische Hochschule Zentralschweiz.
- Shintani, N. & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, (22), 286–306.

- Stephany, S. (2017). Textkohärenz als Einflussfaktor beim Lösen mathematischer Textaufgaben (Sprachliche Bildung). In D. Leiss, M. Hagena, A. Neumann & S. Schwippert (Hrsg.), *Mathematik und Sprache. Empirischer Forschungsstand und unterrichtliche Herausforderungen* (S. 43–62). Münster, New York: Waxmann.
- Sturm, A. (2017). *Von der fachdidaktischen Orientierung zum fachdidaktischen Support. Ein Auswertungsbericht zu «QUIMS: Schreibförderung auf allen Schulstufen (2014–2017)»*. Brugg: Pädagogische Hochschule FHNW.
- Sturm, A. (2018). *Empfehlungen zur Sprachförderung im Pilotprojekt ALLE*. Brugg: Pädagogische Hochschule FHNW.
- Sturm, A., Lindauer, N. & Sommer, T. (i. Vorb.). Schreibunterricht im Spannungsfeld von schreibbezogenem Wissen und Überzeugungen seitens Lehrpersonen.
- Sturm, A. & Weder, M. (2016). *Schreibkompetenz, Schreibmotivation, Schreibförderung. Grundlagen und Modelle zum Schreiben als soziale Praxis*. Seelze: Kallmeyer.
- Südkamp, A., Kaiser, J. & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104 (3), 743–762.
- Südkamp, A. & Praetorius, A.-K. (2017). *Diagnostische Kompetenz von Lehrkräften* (Pädagogische Psychologie und Entwicklungspsychologie) (Band 94). Münster, New York: Waxmann.
- Tajmel, T. (2010). DaZ-Förderung im naturwissenschaftlichen Fachunterricht. In B. Ahrenholz (Hrsg.), *Fachunterricht und Deutsch als Zweitsprache* (167–184). Tübingen: Narr Francke Attempto.
- Tent, L. (2006). Zensuren. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. überarbeitete und erweiterte Auflage., S. 873–880). Weinheim, Basel, Berlin: Beltz.
- Thürmann, E. (2012). Lernen durch Schreiben? Thesen zur Unterstützung sprachlicher Risikogruppen im Sachfachunterricht. *dieS-online*, (1).
- Timperley, H. (2011). *Realizing the power of professional learning*. Berkshire: Open University Press.
- Timperley, H. & Parr, J. (Hrsg.). (2010). *Weaving evidence, inquiry and standards to build better schools*. Wellington: NZCER Press.
- Timperley, H., Wilson, A., Barrar, H. & Fung, I. (2007). *Teacher professional learning and development*. Wellington: Ministry of Education.
- Troia, G. A. & Maddox, M. E. (2004). Writing Instruction in Middle Schools: Special and General Education Teachers Share Their Views and Voice Their Concerns. *Exceptionality*, 12 (1), 19–37.
- Truscott, J. (1999). The Case for «The Case Against Grammar Correction in L2 Writing Classes»: A Response to Ferris. *Journal of Second Language Writing*, 9 (2), 111–122.
- Tschannen-Moran, M., Salloum, S. J. & Goddard, R. D. (2015). Context Matters: The Influence of Collective Beliefs and Shared Nouns. In H. Fives & M.G. Gill (Hrsg.), *International Handbook of Research on Teachers' Beliefs* (S. 301–316). New York/London: Routledge.
- Walter, J. (2009). *LDL – Lernfortschrittsdiagnostik Lesen. Ein curriculumbasiertes Verfahren*. Göttingen: Hogrefe.
- Walzebug, A. (2015). *Sprachlich bedingte soziale Ungleichheit* (Empirische Erziehungswissenschaft) (Band 56). Münster, New York: Waxmann.
- Weinert, F. E. (1999). Die fünf Irrtümer der Schulreformer. *Psychologie heute*, 7 (26), 28–34.
- Wright, T. S. & Neuman, S. B. (2014). Paucity and Disparity in Kindergarten Oral Vocabulary Instruction. *Journal of Literacy Research*, 46 (3), 330–357.
- Zeinz, H. & Köller, O. (2006). Noten, soziale Vergleiche und Selbstkonzepte in der Grundschule. In A. Schröder-Lenzen (Hrsg.), *Risikofaktoren kindlicher Entwicklung: Migration, Leistungsangst und Schulübergang* (S. 177–190). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-90075-9_9
- Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, 41 (2), 64–70.