

Zeitschrift: Schweizerische Zeitschrift für Bildungswissenschaften = Revue suisse des sciences de l'éducation = Rivista svizzera di scienze dell'educazione

Herausgeber: Schweizerische Gesellschaft für Bildungsforschung

Band: 30 (2008)

Heft: 3

Vorwort: Die Urteilsbildung im Unterricht = La formation du jugement en évaluation

Autor: Baeriswyl, Franz / Mottier Lopez, Lucie

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

Download PDF: 09.03.2025

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Editorial: Die Urteilsbildung im Unterricht

Franz Baeriswyl und Lucie Mottier Lopez

In den thematischen Beiträgen dieser Nummer wird die Genese der Urteilsbildung im Unterricht untersucht. Gegenstand und Funktion der Beurteilung im Unterricht sind je mannigfaltig. So wird die fachspezifische Leistungsfähigkeit, das Wissen und Können aber auch überfachliche Kompetenzen wie Arbeitshaltung, Engagement oder Kommunikationskompetenz beurteilt. Die Beurteilungsfunktion umfasst die Zielsetzung und die Beurteilungsform. Die lernprozesssteuernde (formative) Beurteilung orientiert die lernende Person über Fortschritte und zeigt das Gekonnte und die Lücken auf. Sie liefert der Lehrperson aber auch wichtige Erkenntnisse über die Bedürfnisse der Lernenden und hilft ihr den Unterricht zu differenzieren. Die lernsteuernde Beurteilung ist also kein abschliessendes Urteil, wie der Fall der summativen oder zertifizierten Beurteilung ist. Letztere bescheinigt den Abschluss eines Ausbildungsganges oder eines zeitlich strukturierten Abschlusses wie beispielweise eines Schuljahres. Bei Schulübergängen wird von den Lehrpersonen oft eine holistische Schulzuweisungs- oder Schulgangsempfehlung abgefordert, die einen speziell prognostischen Charakter hat.

Mit der Übertragung der psychometrischen Qualitätsanforderungen auf die Beurteilung im Schulkontext entstand in den siebziger und achziger Jahren des letzten Jahrhunderts eine Flut von empirischen Untersuchungen, welche die Unzulänglichkeiten der schulischen Beurteilung aufzeigten. Nur selten waren hingegen die Stimmen, die auf den diesen Studien zugrunde liegenden paradigmatischen Irrtum aufmerksam machten, dass die Gütekriterien der Psychometrie nicht für die Beurteilung im Klassenzimmer gelten dürfen, wenn der pädagogische Auftrag erfüllt werden soll. Trotzdem haben diese oft schuldzuweisenden Forschungsergebnisse in der Bildungslandschaft Wesentliches bewirkt, indem eine eigenständige Forschung mit einer angepassten Forschungsmethodik ausgelöst wurde.

Aspekte dieser erziehungswissenschaftlichen Forschung erscheinen in den thematischen Beiträgen. Dabei wird ein neuer Forschungsfokus gesetzt: Es interessiert die Genese der Urteilsbildung einer Lehrperson über verschiedene Beur-

teilungsgegenstände. Um diesen neuartigen Gesichtspunkt zu klären, wird im Folgenden eine begriffliche Diskussion publiziert.

Alle Beiträge dieser Ausgabe befassen sich mit der Urteilsbildung der Lehrperson in Situationen zertifizierender oder prognostischer Beurteilung, in Situationen von grosser Bedeutung für den Ausbildungsgang der Schülerinnen und Schüler. Solche Übergangssituationen sind nicht nur besonders empfindlich, insbesondere, wenn die Schulsysteme nur schwer Möglichkeiten von Neuorientierungen erlauben (Baeriswyl et al; Maaz et al., in dieser Ausgabe), sondern sie rufen darüber hinaus auch prognostische Beurteilungen hervor, die immer ein gewisses Mass an Unsicherheit beinhalten, welches niemals vollständig ausgeschaltet werden kann. Dies sind Formen von «Wetten auf die Zukunft», deren sich die Lehrpersonen bewusst sind, wie Marcoux und Crahay in dieser Ausgabe festhalten.

Die im ersten Artikel von *Lucie Mottier Lopez* und *Linda Allal* vorgestellte Forschung untersucht die Merkmale der professionellen Schülerbeurteilung von Genfer Lehrpersonen am Ende der Primarschule. Die Ergebnisse der Forschung beschreiben die Verfahren ihrer Notengebung und die Art der von ihnen vorgenommenen Beurteilungsanpassungen, für jene Schüler die auf der Kippe stehen. Die professionelle Schülerbeurteilung wird in diesem Artikel als ein kognitiver Prozess und eine an verschiedene Kontexte geknüpfte soziale Praxis beschrieben, die die Autorinnen dazu führen, die individuellen und sozialen Dimensionen der professionellen Urteilsbildung zu beschreiben.

Dany Laveault schafft einen Literaturüberblick über den Begriff der «professionellen Schülerbeurteilung», indem er sowohl nordamerikanische (französisch- und englischsprachige) als auch europäische Beiträge auswertet. Der Autor beleuchtet verschiedene Beurteilungskategorien und -indikatoren und diskutiert diese im Hinblick auf professionelle Praxis und Reformanforderungen in der Bildung. Die Komplexität der zertifizierenden und prognostischen Beurteilung wird hervorgehoben (z.B. wichtige Herausforderungen, multiple Ziele, mögliche Loyalitätskonflikte), sowie die «Synergien», die entwickelt werden müssten, um die Qualität der professionellen Urteilsbildung der Lehrpersonen zu verbessern.

Der Beitrag von *Géry Marcoux* und *Marcel Crahay* hat zum Ziel, in einem psychologischen Ansatz die Komplexität der Maßnahme zu analysieren, die die Lehrperson zur Entscheidung der Klassenrepetition bewegt- und dies trotz der zahlreichen Forschungsergebnisse, die die Unwirksamkeit dieser Maßnahme aufzeigen. Durch Interviews versuchen die Forscher, das Lehrerurteil sowohl im Falle einer Promotionsentscheidung als auch im Falle einer Klassenrepetition zu verstehen. Auf der Grundlage ihrer Forschungserkenntnisse diskutieren die Autoren vor allem die Professionalität und nicht die Urteile der Lehrer, wenn diese trotz allem ihre Schüler weiterhin wiederholen lassen.

Kai Maaz, Marko Neumann, Ulrich Trautwein, Wolfgang Wendt, Rainer Lehmann und Juergen Baumert untersuchen die Lernkompetenzeinschätzung der Lehrpersonen beim Übergang nach der sechsten Klasse der Grundschule in die gegliederte Sekundarstufe I in Berlin. Die Autoren diskutieren die verschiedenen «Quellen» oder Elemente, welche für die Bildung der Lernkompetenzeinschätzung herangezogen werden. Dabei haben sie einen systematischen Geschlechtereffekt und einen Effekt des sozioökonomischen Status festgestellt. Ebenso beeinflusst die Leistungsstärke der Klasse die Urteilsbildung der Lehrperson, was als Referenzgruppeneffekt bezeichnet werden kann.

Eine ähnliche Fragestellung bearbeiten *Franz Baeriswyl, Christian Wandeler und Oliver Christ*, indem sie die Stabilität der Zusammensetzung der Lehrerempfehlung beim Übertritt in die Sekundarstufe I untersuchen. Obwohl ein Grossteil der Lehrpersonen von einem Jahr zum anderen variiert, stellen sie keine statistisch interpretierbare Veränderung der Lehrerempfehlung über sieben Jahre fest. Von besonderem Interesse ist der Vergleich dieser Lehrerempfehlung mit der Lernkompetenzeinschätzung von Maaz et al., die einen rigorosen Geschlechtereffekt konstatieren, der bei Baeriswyl et al. nicht systematisch feststellbar ist. Die zeitliche Stabilität der Lehrerempfehlung wird als positives Merkmal für Gerechtigkeit interpretiert.

Der dritte deutschsprachige Beitrag von *Anke Treutlein, Jeanette Roos und Hermann Schöler* untersucht den Referenzgruppeneffekt einer Klasse bezüglich der individuellen Notengebung in Rechtschreiben, Lesen und der Gesamtbeurteilung Deutsch. Entgegen der Annahme und auch der Ergebnisse von Maaz et al. zeigte sich ein nur geringer oder kein Referenzgruppeneffekt. Dieses eher erwartungswidrige Ergebnis diskutierten sie ausführlich und zeigen mehrere Erklärungsmöglichkeiten auf.

Die sechs thematischen Beiträge zur Urteilsbildung im Unterricht fokussieren verschiedene Beurteilungsformen und -arten. So beginnt sich ein Puzzle zur Genese der Urteilsbildung zu formen, das allein mit diesen Beiträgen aber noch kein differenziertes Bild erkennen lässt. Die Artikel dieser Ausgabe erlauben jedoch, die Konzeptualisierung und das Verständnis der Prozesse in der pädagogischen Beurteilung weiterzuverfolgen, indem vor allem die schwierige Frage des Verhältnisses zwischen Beurteilung und Bewertung gestellt wird. Von Interesse ist unter anderem, die Bewertungsprozesse in der regulären Beurteilungspraxis zu untersuchen, insbesondere in Situationen, die sich, wie eingangs erwähnt wurde, durch hohe Einsätze auf individueller und kollektiver Ebene auszeichnen.

Die Problematik der Urteilsbildung ist erheblich, besonders unter Berücksichtigung von aktuellen bildungspolitischen Tendenzen, die versuchen, den Spielraum und die Professionalität der Lehrpersonen zu verringern, indem mechanische Verfahren zur Entscheidungsfindung eingeführt werden. Wir denken dabei an Verfahren, die sich im Wesentlichen auf Berechnungen der Durch-

schnittsnoten basieren, und die das Problem der ungenügenden Berücksichtigung anderer Informationen aufwerfen, von denen die (in dieser Ausgabe) im Folgenden dargestellten Forschungsarbeiten zeigen, dass diese Informationen für die Entscheidungen zertifizierender und prognostischer Bewertung unabdinglich und informativ sind.

Éditorial: La formation du jugement en évaluation

Lucie Mottier Lopez et Franz Baeriswyl

La genèse de la formation du jugement en évaluation dans l'enseignement est analysée dans les contributions thématiques de ce numéro. D'une façon générale, les objets et les fonctions de l'évaluation dans l'enseignement sont multiples. Ainsi, les performances disciplinaires spécifiques, les savoirs et savoir-faire, mais aussi les compétences transversales telles que l'attitude au travail, l'engagement ou la compétence communicative sont soumises à l'évaluation des enseignants. Lorsque l'évaluation sert à la régulation des processus d'apprentissage (évaluation formative), elle informe l'apprenant de ses progrès, de ses acquis et de ce qu'il doit encore apprendre ou améliorer. Elle fournit aussi des informations à l'enseignant pour qu'il adapte et différencie ses dispositifs de formation par rapport aux besoins des apprenants. L'évaluation formative n'est donc pas une évaluation finale comme c'est le cas de l'évaluation sommative ou certificative qui, quant à elle, atteste de l'aboutissement d'un parcours de formation ou d'une entité temporellement structurée, telle que par exemple une année scolaire. Pour poursuivre la scolarité, le passage d'un degré à l'autre nécessite une recommandation, souvent holistique et de nature pronostique.

Avec le transfert des exigences psychométriques de qualité à l'évaluation en contexte scolaire, une masse de recherches empiriques s'est développée dans les années 1970 et 1980. Elles ont mis en évidence les insuffisances de l'évaluation scolaire. On a, par contre, rarement entendu les voix de ceux qui mettaient l'accent sur l'erreur paradigmatique à la base de ces études, à savoir que les critères de qualité de la psychométrie ne pouvaient être pris en compte pour l'évaluation en classe qui est liée à un contrat pédagogique et didactique. Paradoxalement, les résultats des recherches pointant les «coupables» dans le paysage éducatif ont eu un effet considérable sur le développement d'une recherche propre, pourvue d'une méthodologie adaptée au contexte scolaire. Des aspects de cette recherche propre aux sciences de l'éducation apparaissent dans les contributions thématiques publiées ici. Une nouvelle focalisation s'y observe: la genèse de la formation du jugement en évaluation de l'enseignant à propos d'objets d'évaluation différents.

Pour clarifier ce point de vue original, une discussion notionnelle est publiée ci-après à la suite de l'éditorial.

Les contributions de ce numéro portent toutes sur la formation du jugement de l'enseignant dans des situations d'évaluation certificative ou pronostique, autrement dit dans des situations comportant des enjeux importants pour les parcours de formation des élèves et de leur orientation. Non seulement ces situations de transition sont particulièrement sensibles notamment quand les systèmes scolaires ne permettent que difficilement des réorientations de parcours (Baeriswyl, Wandeler & Christ; Maaz, Neumann, Trautwein, Wendt, Lehmann & Baumert, dans ce numéro), mais de plus elles convoquent des jugements pronostiques qui comportent une part d'incertitude que l'on ne peut jamais totalement éliminer. Ce sont des formes de «paris sur l'avenir» dont les enseignants ont conscience notent Marcoux et Crahay dans ce numéro de la Revue.

La recherche présentée par *Lucie Mottier Lopez* et *Linda Allal*, dans le premier article, examine les caractéristiques du jugement professionnel en évaluation d'enseignants genevois à la fin de l'école primaire. Les résultats de la recherche décrivent leurs pratiques de notation et la nature des ajustements qu'ils effectuent dans le cas de situations d'incertitude pour certains élèves en particulier. Le jugement professionnel, dans cet article, est appréhendé comme un processus co-gnitif et une pratique sociale liée à différents contextes, amenant les auteures à décrire les dimensions individuelles et sociales du jugement en évaluation.

Le deuxième article, de *Dany Laveault*, propose une revue de littérature sur la notion de «jugement professionnel d'évaluation», en puisant dans des apports nord-américains, (francophones et anglophones) ainsi qu'euro-péens. L'auteur met en lumière différentes catégories et indicateurs de jugement et les discute au regard des gestes professionnels et des exigences des réformes en éducation. La complexité de l'évaluation certificative et pronostique est soulignée (e.g, forts enjeux, buts multiples, conflits possibles de loyauté), ainsi que les «synergies» qu'il s'agirait de développer pour améliorer la qualité des jugements professionnels des enseignants.

La contribution de *Géry Marcoux* et *Marcel Crahay* a pour but d'analyser, dans une approche psychologique, la complexité de l'action qui conduit l'enseignant à des décisions de redoublement malgré les nombreux résultats de recherche qui montrent l'inefficacité de cette mesure. Par le moyen d'entretiens, les chercheurs tentent de dégager la trame décisionnelle du jugement de l'enseignant lorsque celui-ci décide pour un élève une promotion ou un redoublement pour un autre. Sur la base de leurs constats de recherche, les auteurs débattent notamment de la professionnalité ou non des jugements des enseignants lorsque ceux-ci continuent, malgré tout, à faire redoubler leurs élèves.

Kai Maaz, Marko Neumann, Ulrich Trautwein, Wolfgang Wendt, Rainer Lehmann et Juergen Baumert étudient les appréciations des enseignants à propos des compétences d'apprentissage lors du passage en fin d'école primaire vers les filières du secondaire I à Berlin. Les auteurs discutent les différentes «sources» ou éléments qui doivent être pris en compte pour former une appréciation des compétences d'apprentissage. En l'occurrence, ils ont observé un effet systématique à propos du genre (fille-garçon) et un effet lié au statut socioéconomique. De plus, le niveau de la classe influence le jugement évaluatif de l'enseignant, ce qui peut être caractérisé comme un effet du groupe de référence.

Un même questionnement est au cœur de la contribution de *Franz Baeriswyl, Christian Wandeler* et *Oliver Christ*: ils examinent la stabilité de la composition des recommandations de l'enseignant au moment du passage au secondaire I. Bien qu'une grande partie des enseignants changent d'une année à l'autre, sur les sept ans étudiés, les auteurs ne constatent aucun changement de recommandations statistiquement interprétables. La comparaison de ces recommandations est d'un grand intérêt: les évaluations des compétences d'apprentissage de Maaz et al. mettent en évidence un effet significatif lié au genre qui n'est pas systématiquement repérable chez Baeriswyl et al. Par ailleurs, la stabilité temporelle des recommandations des enseignants est interprétée comme une caractéristique positive pour l'équité du jugement.

La troisième contribution en allemand de *Anke Treutlein, Jeanette Roos* et *Hermann Schöler* étudie l'effet du groupe de référence d'une classe à propos de la notation individuelle en orthographe, en lecture et, prise globalement, en allemand. Contrairement à l'hypothèse et aux résultats de Maaz et al., l'effet du groupe de référence est moindre, voire absent. Le résultat contraire aux attentes initiales est discuté de façon détaillée. Les auteurs mettent en évidence plusieurs explications possibles. Nous y renvoyons le lecteur.

Les six contributions thématiques sur la formation du jugement en évaluation focalisent différents modes et formes d'évaluation. Un puzzle de la genèse de la formation du jugement en évaluation s'amorce ainsi, qui, sur la base de ces seules contributions, n'esquisse toutefois pas encore une image bien différenciée. Les contributions de ce numéro permettent la poursuite de la conceptualisation et de la compréhension des processus en jeu dans l'évaluation pédagogique, en posant notamment la question difficile du rapport entre jugement et évaluation. Un des intérêts est d'étudier les processus de jugement dans des *pratiques ordinaires* d'évaluation, plus particulièrement dans des situations qui, comme on l'a dit, se caractérisent par de forts enjeux au niveau individuel et collectif. La problématique de la formation du jugement en évaluation apparaît importante compte tenu aussi de certaines politiques éducatives actuelles qui tendent à réduire les marges de souplesse et de professionnalité des enseignants en instituant des procédures

mécaniques de prise de décision. Nous pensons notamment aux procédures qui se fondent essentiellement sur des calculs de moyennes de notes, soulevant le problème de la prise en compte des informations autres que les résultats chiffrés et dont les recherches dans ce numéro montrent qu'elles sont capitales pour des décisions d'évaluation certificative et pronostique bien fondées.