

## Editorial: Leistungsmessung und Evaluation der Bildungssysteme – Grundannahmen und Kontexte

**Claudia Crotti und Fritz Osterwalder**

Die unübersehbare Präsenz von international vergleichenden Evaluations- und Leistungsstudien über die nationalen Bildungssysteme wie Third International Mathematics and Science Study (TIMSS), Programme für International Student Assessment (PISA) und Adult Literacy and Life Skills Survey (ALL) in der nationalen und zunehmend auch internationalisierten Öffentlichkeit ist mehr als ein Sturm im Wasserglas der Medien. Die ausgedehnten Diskussionen über Ranglisten der Bildungssysteme, über das Leistungsniveau lokaler oder nationaler Bildungssysteme, die Besonderheiten der leistungsstarken und der leistungsschwachen «Modelle» weisen auf bedeutsame Entwicklungen in den Bildungssystemen und in der Bildungspolitik selbst hin:

Zum einen ist unübersehbar, dass in den Ländern und politischen Systemen, wo die institutionelle und die inhaltliche Einbindung des Bildungssystems in lokale Gegebenheiten, wo der Bildungsföderalismus bislang noch als politisch unhintergehbare Konstante gehandelt wurde, wie in den USA oder auch in der Schweiz, die Vereinheitlichung oder vielmehr die Einheitlichkeit oder die Vereinheitlichung des Bildungswesens zu einem allgemein akzeptierten Gestaltungsrahmen geworden ist.<sup>1</sup>

Darüber hinaus ist aber auch festzustellen, dass sowohl in Ländern, die über eine längere Tradition des national bestimmten Bildungszentralismus und der Vereinheitlichung verfügen wie Frankreich, wie auch in jenen, die erst seit kurzem ihre Bildungspolitik in diesem Rahmen neu zu bestimmen beginnen, wie in der Schweiz, zunehmend die *internationalen* Messungen, ihre Messgrößen und die entsprechenden Rankings als bildungspolitische Referenzgrößen betrachtet werden. Die jahrzehntelange Tradition der national-kulturellen Besonderheit und Einbindung der Bildungssysteme wird abgelöst durch eine gezielte internationale Angleichung. Die durch internationale Gremien und Organisationen festgelegten Kriterien der Leistungsmessungen werden für die jeweils nationale Bildungspolitik fraglos als Richtlinien akzeptiert. Hinter PISA scheint sich sogar ein «Weltcurriculum» abzuzeichnen (Fuchs, 2003, S. 161).

Dabei handelt es sich keineswegs um eine blossе Koinzidenz von Vereinheit-

lichungsbestrebungen und internationaler Evaluation bzw. Leistungsmessung. Vielmehr wird diesen Studien selbst ein ganz bestimmter Platz, eine bestimmte Aufgabe in diesem Prozess der internationalen Vereinheitlichung der Bildung zugeordnet. Sie sollen «Ertragsindikatoren für Bildungsprozesse» zur Verfügung stellen, die «für politisch administrative Entscheidungen zur Verbesserung der nationalen Bildungssysteme» (Baumert et al., 2001, S. 285-310) nutzbar oder zu nutzen sind. Die Leistungsmessungen und Evaluationsstudien haben einen vorrangigen, politischen Adressaten, sie richten sich vorweg an politisch oder administrativ ausgewiesene Akteure oder gehören wesentlich zu deren Instrumentarium. Es wird davon ausgegangen oder nahegelegt, dass diese Akteure die Massstäbe der Leistungsmessung auch zu den Richtgrössen ihrer Politik machen und damit zur Vereinheitlichung des Bildungssystems beitragen. Und umgekehrt darf auch davon ausgegangen werden, dass international ausgewiesene Akteure der Bildungspolitik, die seit längerer Zeit die Vereinheitlichung des Bildungswesens zu ihrem wesentlichen Programm machten, sich des Instrumentes der Leistungsmessungen und der vergleichenden Evaluation der Bildungssysteme bedienen.

Während in der Zeit nach dem Zweiten Weltkrieg mit der United Nations Educational, Scientific and Cultural Organization (UNESCO) und dem Europarat verschiedene institutionelle und öffentlich diskutierte Versuche unternommen wurden, die Bildungspolitik der einzelnen Staaten programmatisch anzugleichen, setzte die Organisation for Economic Co-operation and Development (OECD) als Forum und Dienstleistungsanbieterin für die Wirtschaftspolitik der Regierungen schon sehr früh auf eine informelle Vereinheitlichung der Bildungspolitik über prozedurale Fragen. Bereits 1968 gründete die OECD das *Centre for Educational Research and Innovation* (CERI), das die frühere, festgefahrene Initiative der Organisation, die Naturwissenschaften in den Schulen der Mitgliedstaaten zum Zentrum zu machen, dynamisch umsetzen sollte. In der erneuerten Initiative wurde nicht direkt das Ziel, die Vereinheitlichung der Schulprogramme, anvisiert, sondern mit der «wissenschaftlichen Curriculumentwicklung» wurde in einer Reihe von internationalen Seminarien für Bildungspolitiker und Bildungspolitikerinnen, -verwalter und -verwalterinnen sowie -forscher und -forscherinnen ein Verfahren entwickelt, das beitragen sollte, die Curricula aus ihrer Einbindung in nationale politische und verwaltungstechnische Zusammenhänge herauszulösen und in einen internationalen Angleichungsprozess zu überführen. Dieses Programm scheiterte trotz der Bereitschaft zur Zusammenarbeit eines grossen Teiles der europäischen Erziehungswissenschaftler und Erziehungswissenschaftlerinnen und der Bildungsverwaltenden vor allem am damals noch äusserst starken Willen der Aufrechterhaltung der traditionellen lokalen und nationalen Eigenheiten und Einbindungen der Bildungsinstitutionen.<sup>2</sup>

Zum Scheitern der Vereinheitlichung durch wissenschaftliche Curriculumreform trug aber auch die Komplexität des Prozesses selbst bei, der durch die Re-

former und Reformerrinnen in Gang gesetzt werden sollte. Dabei wurden zu viele Vorannahmen getroffen, auf Voraussetzungen gebaut, über welche die Akteure in keiner Art und Weise verfügten.

Das Projekt der international vergleichenden Leistungsmessung, das in den vergangenen Jahren im gleichen institutionellen Rahmen wie die «wissenschaftliche Curriculumsentwicklung» vorangetrieben wurde, ist demgegenüber viel einfacher gestaltet. Es geht aus von der Datenerhebung und Auswertung und setzt dann auf den Willen und die Aktivitäten der Beteiligten der nationalen Bildungspolitik und -verwaltung, sich in einem internationalen Bildungswettbewerb an der Spitzenposition auszurichten. Dabei sind aber gleichfalls die jeweiligen spezifischen Vorannahmen und die spezifischen Voraussetzungen der Leistungsmessung und des Leistungsvergleichs sowie die programmatischen Erwartungen entscheidend.

Es geht darum, die Vorannahmen, Voraussetzungen und die unterschiedlichen Einbindungen, die den vergleichenden Leistungsmessungen und Evaluationssystemen zugrunde liegen, nachzuweisen und zu untersuchen. Dabei werden zwei unterschiedliche Untersuchungsstrategien angewandt.

In den ersten beiden Beiträgen des Thementeils der Zeitschrift wird nach den Programmen der internationalen Leistungsmessungs- und Evaluationsstudien selbst gefragt. Rolf Becker prüft die Frage, ob sich aus den PISA-Studien, wie sie international durchgeführt wurden, überhaupt wissenschaftliche Programme ableiten lassen, wie es bildungspolitisch in den letzten Jahren üblich geworden ist. Dabei geht es nicht um die in vielen zusätzlichen nationalen Studien an PISA angebotenen Untersuchungen und deren Schlussfolgerungen<sup>3</sup>, sondern um die in der bildungspolitischen Auseinandersetzung allgemein präsente Annahme, PISA und ähnliche Studien könnten ein bildungspolitisches Programm begründen und rücke damit den bildungspolitischen Entscheid in den Rahmen von Wissenschaft und objektiver Erhebung. Désirée Donzallaz fragt demgegenüber nach den Wirkungsannahmen, die in den Qualitätssicherungs- und Evaluationsinstrumenten selbst – implizit – getroffen werden.

Im zweiten Abschnitt des Thementeils dieser Zeitschrift wird nach den Vorannahmen und Erwartungen, den Verwendungsweisen und Effekten gefragt, die mit Leistungsmessungen und Evaluationen in historisch und kulturell unterschiedlichen Kontexten getroffen werden. Claudia Crotti und Katharina Kellerhals wenden sich dabei den pädagogischen Rekrutenprüfungen zu, die im ausgehenden 19. Jahrhundert durch die liberalen Kräfte in der Schweiz politisch gezielt, verbunden mit dem Programm der liberalen Erneuerung, eingesetzt wurden. Dabei wird gefragt, auf welche Vorannahmen sich die liberalen Akteure stützen konnten, um mit diesem Instrument einen Vereinheitlichungsprozess der unterschiedlichen kantonalen Bildungssysteme der Schweiz einzuleiten. Jacqueline Levasseur und Jean-Claude Emin, die beide in verantwortlicher Stelle in der französischen Evaluations- und Leistungsmessungsinstitution beteiligt sind, beschreiben deren Programm und Einbindung in die Entwicklung des französi-

schen Bildungssystem. Adam Gamoran schliesslich zeigt die Nutzung eines Leistungs- und Qualitätsmessinstrumentes zur nationalen Vereinheitlichung bzw. Einbindung des hochgradig föderalistischen Bildungssystems der USA.

Es wird hier weder danach gefragt, wie die Instrumente der Leistungsmessung gestaltet sind, noch welche Ergebnisse sie zeitigen und was diese bedeuten. Auch die Vereinheitlichung des Bildungswesens durch gemeinsame Standards und entsprechende Leistungsmessungen steht hier nicht zur Diskussion. Vielmehr soll geklärt werden, welche Vorannahmen der Anwendung dieser Instrumente zugrunde liegen und in welche spezifischen Kontexte ihre Anwendungen eingepasst sind.

### Anmerkungen

- 1 Zu diesem Wandel in der Schweiz siehe Osterwalder, Weber (2004). Unterdessen ist im Mai 2006 die Schweizerische Bundesverfassung ohne namhaften Widerstand im Parlament oder in der Stimmbevölkerung dahingehend geändert worden, dass die Vereinheitlichung der kantonalen Bildungssysteme gemeinsame Aufgabe von Kantonen *und* Bund ist.
- 2 Dieses Vorgehen wurde damals ohne jede Einschränkung offen dargelegt und erläutert durch den damaligen Bildungsverantwortlichen der OECD, David C. Thomas (1972). Das Verfahren der wissenschaftlichen Curriculumsentwicklung sollte zu einem internationalen Schulreform- und Vereinheitlichungsprozess führen. Eingepasst war dieses Projekt in die Diagnose der bevorstehenden technologischen Erneuerung und des Mangels an technischen Kadern. Die expandierende Erziehungswissenschaft in Europa, die den Prozess in den USA nachvollziehen sollte, war dabei ein nicht zu übersehender strategischer Angelpunkt.
- 3 Obwohl die Wissenschaftler und Wissenschaftlerinnen in der frühen Phase der Diskussion über PISA noch regelmässig auf die Grenzen der Untersuchung und auf die Notwendigkeit von Zusatzstudien hinwiesen (Baumert, Stanat & Demmrich, 2001), entwickelte sich alsbald eine normative Auseinandersetzung über die Ergebnisse von PISA, an der sich auch viele Wissenschaftler und Wissenschaftlerinnen uneingeschränkt beteiligten.

### Literatur

- Baumert, J., Stanat, P. & Demmrich, A. (2001). Untersuchungsgegenstand, Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000* (S. 15-68). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). Programme für International Student Assessments. In E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 285-315). Weinheim, Basel: Beltz.
- Fuchs, H.-W. (2003). Auf dem Weg zu einem Weltcurriculum? Zum Grundbildungskonzept von PISA und der Aufgabenzuweisung an die Schule. *Zeitschrift für Pädagogik* 49, 161-179
- Osterwalder, F., Weber, K. (2004). Die Internationalisierung der föderalistischen Bildungspolitik. *Schweizerische Zeitschrift für Bildungswissenschaften*, 26, 11-32.
- Thomas, D. C. (1972). Die Arbeit der OECD an der Curriculumentwicklung. In *Schulreform durch Curriculumrevision. Dokumentation des XIV. Gesprächs zwischen Wirtschaft und Wissenschaft über Schulreform durch Lehrplanrevision* (S. 61-68). Hrsg. Stifterverband für die Deutsche Wirtschaft. Stuttgart: Klett Verlag.

# Éditorial: Mesure des performances et évaluation des systèmes éducatifs – présupposés et contextes

**Claudia Crotti et Fritz Osterwalder**

Loin d'être un effet de mode médiatique, les études comparatives internationales d'évaluation et de performances des systèmes éducatifs (Third International Mathematics dans Science Study (TIMSS), le Programme for International Student Assessment (PISA) ou l'Adult Literacy and Life Skills Survey (ALL)) se sont imposées tant au niveau national qu'international. Les longues discussions sur le positionnement des systèmes éducatifs, sur le niveau de compétences des systèmes éducatifs locaux ou nationaux, sur les particularités des « modèles » très ou peu performants représentent une évolution significative des systèmes éducatifs et de la politique de l'éducation même.

Dans les pays et les systèmes politiques où l'implantation du système éducatif, institutionnelle et relativement aux contenus, était traitée au plan des réalités locales – alors que le fédéralisme était considéré comme une constante incontournable, comme c'était le cas aux Etats-Unis et en Suisse aussi –, la standardisation ou l'unité, voir l'uniformisation de l'enseignement est désormais un programme-cadre universellement accepté.<sup>1</sup>

Non seulement dans les pays comme la France, disposant d'une longue tradition de centralisme et d'uniformisation de la politique de l'éducation nationale, mais aussi dans les pays qui commencent seulement à définir leur politique dans ce cadre comme la Suisse, on constate que les épreuves internationales, les indicateurs de mesure et le positionnement y relatif sont de plus en plus considérés comme des unités de références en matière de politique éducative. La longue tradition des spécificités culturelles nationales ancrées dans les systèmes éducatifs est remplacée par une politique d'alignement international. Les critères de mesures des performances établis par les instances et organisations internationales sont acceptés comme ligne de conduite pour la politique de l'éducation sans discussion. Derrière PISA, un « curriculum mondial » pourrait même s'esquisser (Fuchs, 2003, p. 161).

Il ne s'agit pourtant pas d'une simple coïncidence conjuguant divers efforts d'uniformisation, d'évaluation internationale et de mesures de performances. Ces études ont bien plus pour tâche précise d'harmoniser internationalement

l'éducation. Elles ont la fonction de fournir des «indicateurs de compétences pour les processus de formation», à utiliser «pour des décisions politiques et administratives visant l'amélioration des systèmes éducatifs nationaux» (Baumert *et al.*, 2001, p. 285-310). Les mesures de performances et les études d'évaluation ont un destinataire politique prioritaire: elles s'adressent en particulier à des responsables influents de la scène politique ou administrative, et constituent un élément essentiel de leurs instruments d'action. On postule que ces responsables adoptent les mêmes critères de mesures des performances pour orienter leur politique et, de cette façon, contribuer à l'uniformisation du système éducatif. Dans le même mouvement, on suppose que les responsables internationaux de la politique de l'éducation qui, depuis longtemps, ont fait de l'harmonisation de l'éducation leur programme principal, utilisent aussi ces mesures des performances et l'évaluation comparative des systèmes éducatifs pour leur politique.

Après la deuxième guerre mondiale, avec l'*United Nations Educational, Scientific and Cultural Organization* (UNESCO) et le *Conseil de l'Europe*, diverses expériences institutionnelles, discutées publiquement, ont été tentées pour harmoniser la politique de l'éducation des différents états du point de vue des programmes. En qualité de forum et de prestataire de services pour la politique économique des gouvernements, l'OCDE avait mis très tôt sur une uniformisation informelle de la politique de l'éducation grâce à des procédures. L'OCDE avait fondé, déjà en 1968, le *Centre for Educational Research and Innovation* (CERI). Ce centre proposait de redynamiser une ancienne initiative, pourtant enlisée, de l'organisation. Il s'agissait de mettre les sciences naturelles au centre de l'éducation dans les écoles des états membres. Avec cette nouvelle initiative, l'objectif ne visait plus directement l'uniformisation des programmes scolaires. Grâce au «développement du programme d'enseignement scientifique», une série de séminaires internationaux ont été mis sur pied pour les politiques, les administrateurs et les chercheurs dans le domaine de l'éducation. Par ce moyen, les programmes d'enseignement devaient s'abstraire de leur contexte politique et administratif national et s'engager dans un processus d'harmonisation international. Malgré la collaboration d'une grande partie de pédagogues et administrateurs de l'éducation européens, ce programme a avorté, surtout à cause de la très forte volonté, manifeste à cette époque, de maintenir les particularités traditionnelles, locales et nationales, des institutions éducatives.<sup>2</sup>

La complexité du processus que les réformateurs et les réformatrices auraient dû enclencher a aussi contribué à l'échec de l'harmonisation par la réforme scientifique des programmes d'enseignement. Trop de présupposés s'appuyaient sur des conditions dont les responsables ne disposaient absolument pas.

Le projet de mesure comparative des performances internationales, déjà activé dans les années passées dans le même cadre institutionnel que le «développement scientifique des programmes d'enseignement», est par contre structuré de façon plus simple. Il s'appuie sur la saisie et l'exploitation de données et se fonde sur la volonté et l'engagement des politiques et des administrateurs dans le

domaine de l'éducation nationale d'atteindre une position de pointe dans la compétition internationale. Les présupposés spécifiques, les conditions de mesure et de comparaison des performances, ainsi que les attentes programmatiques sont cruciales dans le contexte actuel.

En conséquence, il s'agit de montrer et d'examiner les présupposés, les conditions et les différentes implantations qui sont à la base des mesures comparatives de performances et des systèmes d'évaluation. Dans la partie thématique de la présente *Revue*, deux différentes stratégies d'analyse sont utilisées pour y parvenir.

Dans les deux premiers articles, la discussion porte sur les programmes d'études des performances et d'évaluation internationales. *Rolf Becker* examine la possibilité même de s'appuyer sur les études PISA, réalisées au niveau international, pour concevoir des programmes scientifiques comme c'est devenu l'habitude ces dernières années en matière de politique de l'éducation. Cet examen ne concerne pas les nombreuses recherches et conclusions<sup>3</sup> en relation avec de nouvelles études nationales sur PISA, mais plutôt l'opinion répandue dans le débat politique que PISA et des études analogues pourraient fonder un programme en matière de politique de l'éducation, déplaçant ainsi le processus de décision de la politique éducative dans le cadre scientifique et d'une enquête objective. De son côté, *Désirée Donzallaz* interroge l'effet des hypothèses qui, de façon implicite, doivent être posées dans le cadre des instruments pour l'assurance de la qualité et l'évaluation.

Dans une deuxième partie sont étudiés les fondements et les attentes, les modes d'utilisation et les effets des contextes historiques et culturels différents qui interviennent dans les questions de mesures de performances et les évaluations. *Claudia Crotti* et *Katharina Kellerhals* examinent ici les examens pédagogiques des recrues militaires. Vers la fin du 19<sup>e</sup> siècle, ces examens avaient été introduits en Suisse par les forces libérales, avec une visée politique, parallèlement à leur programme de renouveau libéral. Dans ce contexte rénovateur, la question des présupposés des responsables libéraux se pose : qu'est-ce qui, grâce à cet instrument, pouvait fonder un processus d'uniformisation des différents systèmes éducatifs cantonaux en Suisse? *Jacqueline Levasseur* et *Jean-Claude Emin*, qui occupent tous les deux un poste à responsabilité dans l'institution française pour l'évaluation et la mesure des performances, décrivent le programme et son implémentation dans le développement du système d'éducation français. Pour terminer, *Adam Gamoran* montre l'utilisation d'un instrument de mesures des performances et de la qualité visant, par son intégration, l'harmonisation nationale du système éducatif hautement fédéraliste des Etats-Unis.

La discussion ne porte donc ici ni sur la façon dont sont structurés les instruments de mesure des performances, ni sur les résultats produits, et pas non plus sur leur interprétation. L'uniformisation de l'éducation par des standards communs et des mesures des performances n'est même pas débattue ici. Il s'agit bien plus de clarifier quels sont les présupposés au cœur de l'utilisation de ces instruments et dans quels contextes ils sont utilisés.

## Notes

- 1 Pour d'autres renseignements sur la transformation en Suisse, voir Osterwalder et Weber (2004). Entre-temps, en mai 2006 la constitution fédérale suisse a été modifiée sans opposition notable du parlement ou des électeurs: l'uniformisation des systèmes éducatifs cantonaux représente aujourd'hui une tâche commune des cantons et de la Confédération.
- 2 À cette époque, cette procédure avait été ouvertement exposée et commentée sans aucune restriction de la part du responsable de l'éducation de cette époque de l'OCDE, David C. Thomas (1972). Le développement scientifique des programmes d'enseignement devait mener à un processus international de réforme et d'uniformisation scolaires. Ce projet se fondait sur le diagnostic d'un renouveau technologique imminent et d'un manque de cadres techniques. L'expansion de la pédagogie en Europe, qui devait imiter celle des États-Unis, a représenté un argument stratégique.
- 3 Bien que, pendant la première phase des discussions sur PISA, les scientifiques fassent régulièrement remarquer les limites de la recherche ainsi que la nécessité d'études additionnelles (Baumert, Stanat & Demmrich, 2001), peu après il s'est développé un débat normatif à propos des résultats de PISA, auquel de nombreux scientifiques ont pris part sans aucune réserve.

## Bibliographie

- Baumert, J., Stanat, P. & Demmrich, A. (2001). Untersuchungsgegenstand, Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Ed.), *PISA 2000* (pp.15-68). Opladen: Leske + Budrich.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). Programme für International Student Assessments.
- In E. Weinert (Ed.), *Leistungsmessungen in Schulen* (p. 285-315). Weinheim, Basel: Beltz.
- Fuchs, H.-W. (2003). Auf dem Weg zu einem Weltcurriculum? Zum Grundbildungskonzept von PISA und der Aufgabenzuweisung an die Schule. *Zeitschrift für Pädagogik* 49, 161-179.
- Osterwalder, F./Weber, K. (2004). Die Internationalisierung der föderalistischen Bildungspolitik. *Schweizerische Zeitschrift für Bildungswissenschaften*, 26, 11-32.
- Thomas, D. C. (1972). Die Arbeit der OECD an der Curriculumentwicklung. In *Schulreform durch Curriculumsrevision. Dokumentation des XIV. Gesprächs zwischen Wirtschaft und Wissenschaft über Schulreform durch Lehrplanrevision* (Stifterverband für die Deutsche Wirtschaft, pp. 61-68). Stuttgart: Klett Verlag.