

Schwippert, Knut

Schulentwicklung und der Blick von außen. Erfahrungen mit der externen Evaluation des Osnabrücker Mathematikcurriculums

Die Deutsche Schule 96 (2004) 4, S. 426-437



Quellenangabe/ Reference:

Schwippert, Knut: Schulentwicklung und der Blick von außen. Erfahrungen mit der externen Evaluation des Osnabrücker Mathematikcurriculums - In: Die Deutsche Schule 96 (2004) 4, S. 426-437 - URN: urn:nbn:de:01111-pedocs-274296 - DOI: 10.25656/01:27429

<https://nbn-resolving.org/urn:nbn:de:01111-pedocs-274296>

<https://doi.org/10.25656/01:27429>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Mitglied der


Leibniz-Gemeinschaft

Knut Schwippert

Schulentwicklung und der Blick von außen

Erfahrungen mit der externen Evaluation des Osnabrücker Mathematikcurriculums

Ein Ziel von Evaluationen ist, Unterricht oder Unterrichtskonzepte zum Zwecke der Weiterentwicklung zu beurteilen. Die Evaluation ist hierbei nicht nur eine Bestandsaufnahme, sondern dient primär zur Überprüfung und Regulierung des evaluierten Objekts und ist somit aktions-initiiierend ausgerichtet.

Wenn eine Evaluation zielführend durchgeführt wird, müssen im Vorwege *Kriterien* festgelegt werden, anhand derer der Evaluationsgegenstand beurteilt werden soll. Rolff (2001) und Peek (2001) merken kritisch an, dass aufgrund der oftmals kollegialen und auch freundschaftlichen Verbindungen der Lehrkräfte im Rahmen einer internen Evaluation notwendige Kritik nicht geäußert wird. Aus diesem Grund kann sich nur eingeschränkt eine wirksame kritische Diskussion entwickeln. Hingegen zeigt sich der Vorteil einer internen Evaluation darin, dass hier die Evaluatoren über Insiderwissen verfügen, was bei der Interpretation des Evaluationsergebnisses hilfreich sein kann.

Eine *externe Evaluation* kann prinzipiell nach den gleichen Kriterien wie eine interne durchgeführt werden, wobei die Evaluatoren dann Außenstehende sind, die in keiner persönlichen Beziehung zu den beobachteten Personen stehen. Somit bietet eine externe Evaluation die Möglichkeit, Befunde und daraus abgeleitete Rückschlüsse durch eine größere persönliche Distanz zu objektivieren.

Wo bislang eine interne Evaluation etabliert war, kann es eine weitere Alternative sein, diese zu erweitern und *ex post* durch eine externe Evaluation an den Ergebnissen von zum Beispiel einer bundesweit durchgeführten Schulleistungsstudie zu verankern (Klieme/Baumert/Schwippert 2000; Köller/Trautwein 2003). Für diese Alternative haben sich auch die Autoren des in Osnabrück realisierten Mathematikcurriculums entschieden. Bei der gewählten Form der externen Evaluation werden die Leistungen der nach diesem Curriculum unterrichteten Schülerinnen und Schüler mit den im Rahmen der 1995 durchgeführten TIMSS-Untersuchung (der für die Jahrgangsstufe normorientiert interpretiert werden kann) verglichen. Voraussetzung einer so angelegten externen Evaluation ist, dass die Ziele, Kriterien und die sich aus der Untersuchung ergebenden Handlungsalternativen geklärt sind und dass geeignete Instrumente im Evaluationsbereich zur Verfügung stehen. Der Forschungsbereich „Erziehungswissenschaft und Bildungssysteme“ des Max-Planck-Instituts für Bildungsforschung in Berlin (MPI) wurde gebeten, diese Evaluation durchzuführen.

1. Das Osnabrücker Mathematik-Curriculum

In den vergangenen 15 Jahren haben Mathematikdidaktiker der Universität Osnabrück ein innovatives Mathematikcurriculum für die Jahrgangsstufen 7 bis

10 des Gymnasiums entwickelt und in mehreren Schulversuchen erprobt (Cohors-Fresenborg 2001; Cohors-Fresenborg/Kaune 1993; Sjuts, 1999). Algorithmische und axiomatische Denkweisen, die für das Selbstverständnis moderner Mathematik von zentraler Bedeutung sind, werden hier in den Unterricht integriert. Mathematik wird in dem Osnabrücker Curriculum als ‚Werkzeug zur Wissensrepräsentation‘ verstanden und genutzt. Nach Überlegungen der Curriculaautoren soll ein Mathematikunterricht, in dem diese Auffassung im Mittelpunkt steht, nicht nur zu einem besseren Verständnis mathematischer Konzepte und zu höheren Fähigkeiten im mathematischen Argumentieren führen, sondern auch die epistemologischen Überzeugungen der Schüler, d.h. deren „mathematisches Weltbild“ (Grigutsch, Raatz & Törner, 1998) beeinflussen. Als weiterer Effekt rechnen die Autoren damit, dass das Curriculum die Lernmotivation und das Selbstkonzept der mathematischen Begabung fördert.

Zum Zeitpunkt der Evaluation wurde das Curriculum an zwei niedersächsischen Gymnasien mit Genehmigung der Aufsichtsbehörden als Schulcurriculum genutzt. Als Bedingungen wurden genannt, dass zum einen die Fachlehrer der Schulen, aber auch Schulleiter und Eltern vom Nutzen des Curriculums überzeugt waren. Zum anderen sollte der zusätzliche Aufwand, der mit der Fortbildung von Lehrern, dem Einsatz spezieller Lehrbücher und Unterrichtsmaterialien usw. verknüpft war, bewältigt werden können. Schließlich wurde vorausgesetzt, dass, obwohl nur 75 Prozent der Unterrichtszeit dem Landescurriculum gewidmet sind, die Lernziele in vergleichbarem Umfang wie bei Gymnasien mit konventionellem Curriculum erreicht würden (Cohors-Fresenborg/Schwippert/Klieme 2002; Klieme u.a. 2000).

2. Planung und Durchführung der Evaluation

Im Rahmen der Vorbereitung der Evaluation zeigte sich, dass die beiden beteiligten Schulen unterschiedliche Fragestellungen bezüglich der zukünftigen Implementierung des Osnabrücker Curriculums hatten.

Während beispielsweise Fachlehrer der einen Schule mit hohem Engagement weiterhin die Implementierung des Curriculums wünschten, plädierten die anderen für den Wechsel zu Standardlehrbüchern und -lehrplänen, wenn sich kein lohnenswerter Erfolg belegen ließe. In beiden Schulen ging es somit um die Frage, ob das Curriculum beibehalten werden kann. An der einen Schule wollte man prüfen, ob der Mathematikunterricht dieser Schule die mathematische Kompetenz der Schüler und deren Lernmotivation – verglichen mit anderen Gymnasien – in ausreichendem Maße sichert. In der anderen Schule hingegen lautete die zentrale Frage, ob sich der erhöhte Aufwand lohnt. Die Befürworter des Osnabrücker Curriculums vertraten die These, dass Kompetenz und Motivation der Schüler durch den Lehrgang über das an Gymnasien mit konventionellem Curriculum übliche Maß hinaus gefördert würden (Klieme u.a. 2000).

Die Ergebnisse der Evaluation sollten im Rahmen eines Workshops, der von den beteiligten Wissenschaftlern organisiert und durchgeführt wurde, diskutiert werden. Eine der Schulen entschied sich, die klassenbezogenen Ergebnisse auch lehrkraftbezogen zu diskutieren, während die andere Schule keine Zuordnung von Klassenergebnissen und Lehrkräften wünschte. In diesem Beitrag werden die Ergebnisse durchgängig anonym vorgestellt.

Die in der Evaluation eingesetzten Leistungstest und Hintergrundfragebögen entsprachen den in der Third International Mathematics and Science Study (TIMSS) eingesetzten Instrumente (Baumert u.a. 1997). Auch die Testdurchführung im Frühsommer 1999 wurde genau nach den in TIMSS angewandten Richtlinien durchgeführt. Im Rahmen der Untersuchung wurden in den beiden Schulen jeweils vier Klassen mit insgesamt 185 Schülerinnen und Schülern befragt.

Die Befunde der Evaluation wurden für die Rückmeldung so visualisiert, dass sie ohne spezifisches statistisches Basiswissen nachvollziehbar sein sollten. Die klassenbezogenen Ergebnisse wurden jeweils in Relation zu einer aus der TIMSS-Untersuchung stammenden Teilstichprobe verglichen.¹ Diese Teilstichprobe bestand aus den Schülerinnen und Schülern von Gymnasien desselben Jahrgangs aus Bundesländern, die vergleichbare Expansionsraten wie Niedersachsen aufwiesen. Somit stehen für einen Vergleich folgende Stichproben zur Verfügung:

Tabelle 1: Gegenüberstellung der Vergleichs- und Evaluationsstichprobe

	Systemmonitoring: Vergleichsstichprobe aus TIMSS	Evaluationsstichprobe
Anzahl der Schulen	16	2
Anzahl der Schüler	376	185
Berücksichtigte Schulen	Schulen aus Bundesländern mit gleicher Expansionsrate	Versuchsschulen aus einem Bundesland

3. Der Workshop

Ziel des Workshops war es, einerseits die Ergebnisse der Untersuchung zurückzumelden und andererseits den beteiligten Lehrkräften Interpretationshilfen der Befunde und somit Entscheidungshilfen für die Beantwortung der unterschiedlichen Fragestellungen zu bieten. Daher wurden die Ergebnisse anhand der Moderationsmethode und der Metaplan-Technik präsentiert.² Zunächst wurden die Erwartungen der Lehrkräfte ermittelt und für alle Beteiligten als Poster visualisiert. Erst dann wurden die ermittelten Ergebnisse in Relation zu den Erwartungen vorgestellt. Erwartungs- und Ergebnisdivergenzen wurden anschließend gemeinsam von Evaluatoren und Lehrkräften interpretiert und mögliche Ursachen diskutiert.

Die Rolle der Wissenschaftler, welche die externe Evaluation durchgeführt und die Präsentation vorbereitet haben, beschränkte sich darauf, die Diskussion zu moderieren und bei der Interpretation von Befunden zu assistieren. Der Workshop gliederte sich in verschiedene Blöcke:

- 1 Zur statistischen Überprüfung der beobachteten Mittelwerte wurde das Programm HLM genutzt, welches die Gruppierungseffekte der Schüler in Klassen (Klumpungseffekt) angemessen berücksichtigt.
- 2 Der zum Zeitpunkt des Workshops am MPI tätigen Praktikantin Martina Diedrich (jetzt DIPF-Frankfurt) danke ich für die gewissenhafte Protokollführung und Dokumentation der Moderationen.

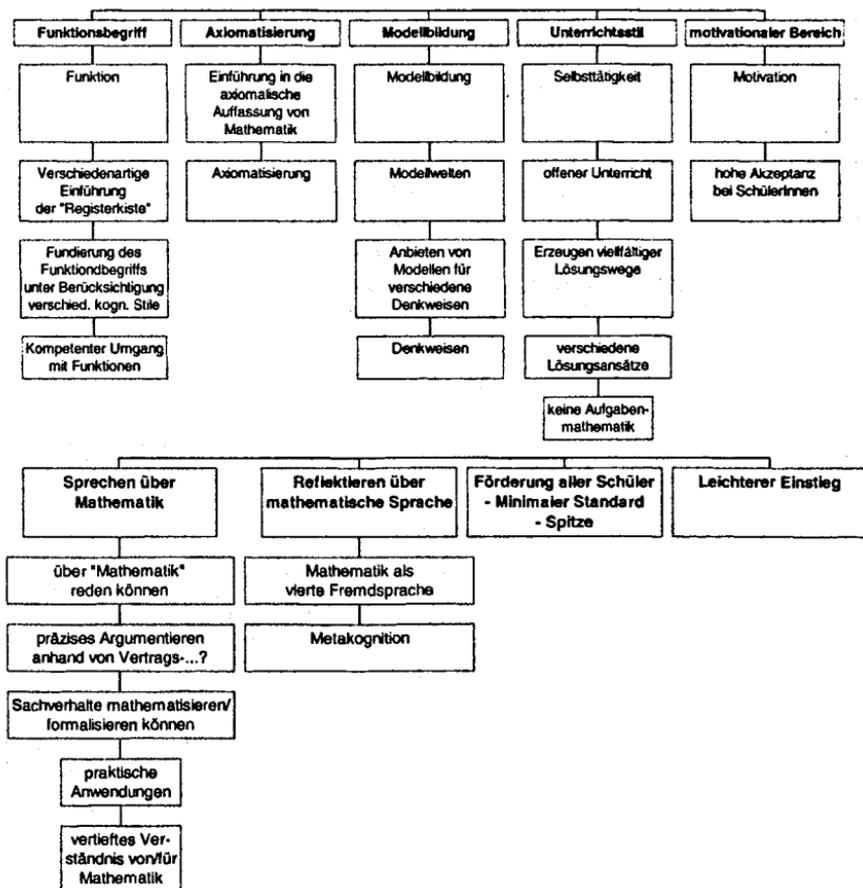
3.1 Typische Merkmale des Osnabrücker Curriculums

Um den Lehrkräften der beiden beteiligten Schulen einen gemeinsamen Bezugsrahmen für die folgenden Diskussionen zu geben, wurde zunächst anhand eines Brainstormings jeder Teilnehmer gebeten, Stichworte zu zentralen Inhalten, Zielen, Methoden/Prozessen oder Voraussetzungen der Unterrichtseinheiten auf Kärtchen zu notieren, die dann von einem Moderator an einer Stellwand angeheftet und dabei gruppiert wurden. Gruppen (hier: Konzepte höherer Ordnung) wurden nach kurzer Diskussion Überschriften zugeordnet.

Hierbei wurden folgende Kennzeichen des Curriculums herausgearbeitet: Übereinstimmend wurde das besondere *Verständnis des Funktionsbegriffs*, die axiomatische Auffassung von Mathematik, die Verwendung von Mathematik zur Modellierung von Problemsituationen (die zu einem vertieften Verständnis der Mathematik führt), ein besonders offener Unterrichtsstil und die verwendete Sprache als zentral kennzeichnend für das Osnabrücker Mathematikcurriculum beschreiben.

Als besondere Stärke wird von Schülern, die nach diesem Curriculum unterrichtet werden, erwartet, dass sie die *Mathematik als besondere Sprache* ver-

Abbildung 1: Metaplan



stehen. Darüber hinaus soll ihnen der Unterricht alternative Denkmodelle liefern, die ihnen auch in Bereichen außerhalb der Mathematik hilfreich sein sollen. Von den Beteiligten wird darüber hinaus damit gerechnet, dass die Schülerinnen und Schüler den Unterricht motivierter verfolgen, insbesondere potentiell leistungsschwächere Schüler Vorteile haben, und dass generell ein einfacherer Einstieg in die Mathematik gelingt.

Der als Resultat der Diskussion gemeinsam erarbeitete vollständige Metaplan ist in Abbildung 1 detailliert dargestellt.

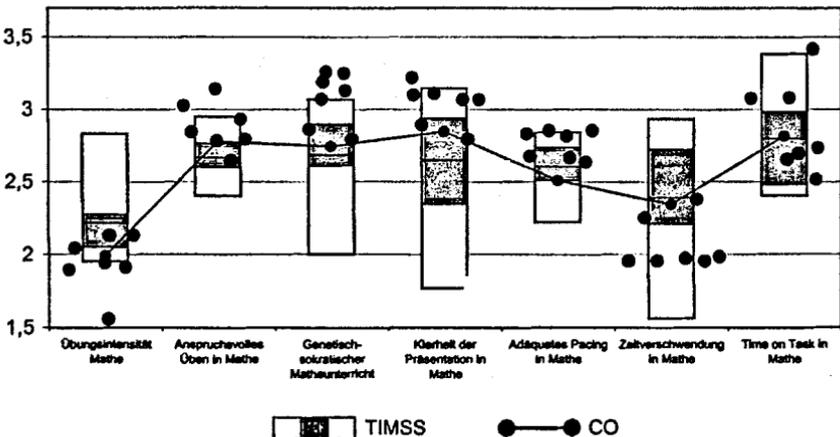
3.2 Erwartungen und Beobachtungen

Im Folgenden wurden die Lehrkräfte gebeten, ihre Erwartungen in Bezug auf von den Schülerinnen und Schülern wahrgenommene Unterrichtsmerkmale, Interessen, Einstellungen und Lernstrategien in Relation zur TIMSS-Vergleichsstichprobe zu kennzeichnen.

Hierzu wurde jede Lehrkraft gebeten, auf einer entsprechenden Abbildung (vgl. Abb. 2, 3 und 4) für jedes erfasste Merkmal einen Punkt zu machen, der ihre Erwartungen für die eigene Klasse in Relation zu den bereits eingetragenen Werten der TIMSS-Vergleichsstichprobe kennzeichnet. Die je Merkmal entstandenen Punkteverteilungen wurden anschließend im Plenum diskutiert. In einem nächsten Schritt wurden die im Rahmen der Evaluation ermittelten Schülerantworten in die Graphen übertragen. In den Abbildungen 2 und 3 sind die Ergebnisse ausgewählter Posterpräsentation dargestellt.

Die dargestellten Balken kennzeichnen die Verteilung von Schülerantworten in der TIMSS-Vergleichsstichprobe. Hierbei markiert der mittlere grau unterlegte Bereich die mittleren 50 Prozent der Schülerantworten und die jeweils oben und unten anschließenden Kästen die restliche um extreme Antworten (5% und 95% Perzentile) bereinigte Antwortspanne. Die freien Punkte entsprechen den Erwartungswerten der Lehrkräfte. Die mit Strichen verbundenen Punkte entsprechen den Mittelwerten der Schülerantworten in den Versuchsklassen. Als Stern (nur in Abb. 3) gekennzeichnete Klassenergebnisse kennzeichnen statistisch sig-

Abbildung 2: Schülerwahrnehmung von Unterrichtsmerkmalen, Vergleich der Versuchsklassen mit TIMSS-Verteilungen

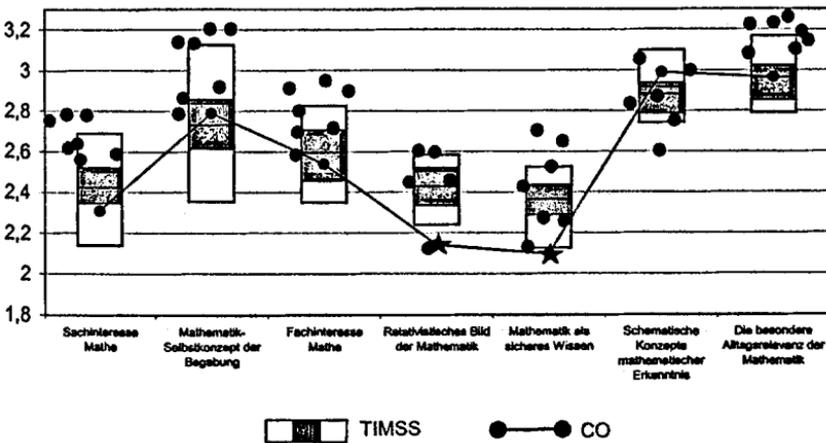


nifkante Abweichungen (auf 5%-Niveau) zwischen der TIMSS-Stichprobe und den nach dem Osnabrücker Curriculum unterrichteten Klassen.³

Abbildung 2 ist die Erwartung der Lehrkräfte zu entnehmen, die nach dem Osnabrücker Curriculum unterrichteten Schüler würden die Übungsintensität in Mathematik und eine Zeitverschwendung im Unterricht nicht so ausgeprägt wahrnehmen wie Schüler, die nach Standardcurricula unterrichtet werden. Hingegen sahen die Lehrkräfte die Wahrnehmung eines anspruchsvollen Übens, eines genetisch-sokratischen (entwickelnden/problemorientierten) Unterrichts, die Klarheit der Präsentation und eines angemessenen Unterrichtstempos (*pacing*) über dem Niveau der TIMSS-Vergleichsstichprobe. Uneinheitlich wurde die Wahrnehmung der effektiven Zeitznutzung im Unterricht (*time on task*) in Mathematik erwartet. Diese Erwartungen deckten sich auch mit den in der Besprechung der im Metaplan diskutierten Kennzeichen des Curriculums. Auch wenn – mit einer Ausnahme – die beobachteten Mittelwerte nominell in die antizipierten Richtungen von den TIMSS-Referenzwerten abweichen, lässt sich für keines der dargestellten Unterrichtsmerkmale ein signifikanter Unterschied zwischen den Versuchsklassen und der Vergleichsgruppe feststellen. Für Überraschung sorgte, dass die Schülerinnen und Schüler den genetisch-sokratischen Unterrichtsstil, die Klarheit der Präsentation und die angemessene Zeitznutzung weniger optimistisch einschätzten als die Lehrkräfte es erwarteten.

Die Befunde des bei den Schülerinnen und Schülern ermittelten Interesses, Selbstkonzepts und mathematischen Weltbilds ist in Abbildung 3 dargestellt. In Relation zu der TIMSS-Vergleichsgruppe schätzten die Lehrkräfte das relativistische Bild der Mathematik, die Mathematik als sicheres Wissen und schematische Konzepte mathematischer Erkenntnis ähnlich ein. Höher wurden das Sachinteresse, das mathematische Selbstkonzept der Begabung, das Fachinteresse und die Wahrnehmung der besonderen Alltagsrelevanz der Mathematik

Abbildung 3: Interesse, Selbstkonzept und Mathematische Weltbilder, Vergleich der Versuchsklassen mit TIMSS-Verteilungen



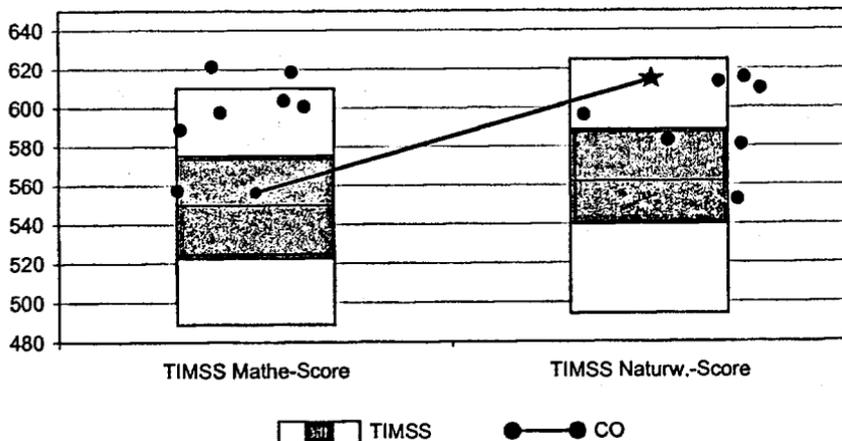
³ Zur Methode der zufallskritischen Absicherung siehe: Klieme/Baumert/Schwip-pert 2001.

erwartet. Die tatsächlich beobachteten Werte bei den Schülerinnen und Schülern unterschieden sich nur mit Ausnahme von zwei der vier mathematischen Weltbilder nicht signifikant von den in der TIMSS-Vergleichsgruppe ermittelten Durchschnittswerten. Nur die Vorstellungen des relativistischen Weltbilds und der Mathematik als sicheres Wissen wurde von den Schülerinnen und Schülern signifikant niedriger beurteilt als in der Vergleichsstichprobe.

Dass die Lehrkräfte hier höhere Erwartungen haben, ist auf eine abweichende *Interpretation der Skalen* zurückzuführen. Während die Lehrkräfte das Weltbild zum sicheren Wissen als die Überzeugung von Objektivität oder Gültigkeit mathematischer Erkenntnisse interpretierten, zielten die Einzelfragen, die den Schülerinnen und Schülern vorgelegt wurden und die zu dem Konstrukt des Weltbildes zusammengefasst wurden, darauf, dass Mathematik etwas Statisches sei bzw. aus fest vorgegebenen Formeln und Prozeduren bestehe. Entsprechend wichen auch die Vorstellungen der Lehrkräfte von denen der Schülerinnen und Schüler vorgelegten Einzelfragen ab, was in dem Konstrukt „Mathematik als sicheres Wissen“ subsumiert wurde. Hier vermuteten die Lehrkräfte, dass Algorithmen und Regeln in der Mathematik beherrscht werden sollten, um sie kreativ einsetzen zu können, während die Einzelfragen auf ein Verständnis von Mathematik als einer einfachen (schematischen) Anwendung von mathematischen Algorithmen zielten. Die signifikant niedrigeren Einschätzungen des relativistischen Bilds der Mathematik und der Mathematik als sicheren Wissens entspricht – obwohl von den Lehrkräften abweichend interpretiert und beurteilt – der Intention des Curriculums. Mathematik wird hier von den Schülerinnen und Schülern als kreative Sprache und weniger als eine Schemaorientierung verstanden.

Eine rege Diskussion ergab sich bei der Präsentation der Ergebnisse des Mathematik- bzw. Naturwissenschaftstests.⁴ Abbildung 4 zeigt diese Befunde.

Abbildung 4: Mathematische und naturwissenschaftliche Leistungen, Vergleich der Versuchsklassen mit TIMSS-Verteilungen



4 Abweichende Streuungsmaße, die in Klieme/Baumert/Schwippert (2000) bzw. Cohors-Fresenborg/Schwippert/Klieme (2002) berichtet werden, sind darauf zurückzuführen, dass in diesen Publikationen die Statistiken auf Individualebene und im Rahmen des Workshops zunächst auf Klassenebene basierten.

Erwartungsgemäß gaben die Lehrkräfte in der Regel an, dass die Schülerinnen und Schüler der von ihnen unterrichteten Klassen deutlich über den durchschnittlichen Leistungen der Vergleichsgruppe liegen würden. Eine ähnliche, jedoch insgesamt weniger optimistische Einschätzung wurde auch für die ebenfalls erfassten naturwissenschaftlichen Leistungen von den Lehrkräften auf dem Poster anhand der Klebepunkte gekennzeichnet. Umso größer war dann die Überraschung seitens der Lehrkräfte, als die tatsächlich ermittelten Ergebnisse präsentiert wurden. In Mathematik unterschieden sich die Versuchs- und Vergleichsklassen weder statistisch signifikant noch in für die Unterrichtspraxis relevanter Weise von der TIMSS-Vergleichsgruppe.

Mit Überraschung wurde auch das Ergebnis im Test zur Naturwissenschaft zur Kenntnis genommen. Hier zeigte sich, dass die nach dem Osnabrücker Curriculum unterrichteten Klassen signifikant besser abschnitten als die der Vergleichsgruppe. Zum einen wurde dies auf die allgemeine Verbesserung des logischen Denkens zurückgeführt, auch wenn sich dies nicht – wie eigentlich zu erwarten wäre – in ähnlicher Weise auch in der Mathematik hätte zeigen müssen. Als beruhigend wurde schließlich unisono festgestellt, dass das Ziel, mindestens das Mathematik-Niveau von vergleichbaren Klassen zu erreichen, als erfüllt angesehen werden konnte.

Neben den bis hierhin vorgestellten Ergebnissen der externen Evaluation wurden analog auch die Ergebnisse von spezifischen mathematischen Aufgaben vorgestellt und diskutiert. Eine differenzierte Darstellung auf Einzelfragenebene ist in Cohors-Fresenborg/Schwippert/Klieme (2002) dargestellt und erfolgt hier aus Platzgründen nicht.

3.3 Plenums- und Kleingruppendiskussion der Ergebnisse

Dass die nach dem Osnabrücker Curriculum unterrichteten Schüler in ihren Interessen bzw. Selbstkonzepten hinter den Erwartungen der Lehrkräfte liegen, wurde nach kurzer Diskussion auf eine fehlende Vergleichsnorm in der Beurteilung zurückgeführt. Die Schülerinnen und Schüler hätten keinen Vergleich, inwieweit sich der von ihnen erlebte Unterricht von demjenigen unterscheidet, der nach Standardcurricular gehalten wird. Daher werden die Antworten der Schüler nicht als Widerspruch zu den bei der Metaplan-Erstellung formulierten Erwartungen gesehen, und somit auch kein plausibler Grund abgelesen, das Curriculum aufgrund dieser Befunde abzulehnen. Das in den Wahrnehmungsdivergenzen ruhende Potenzial für die Curriculumentwicklung ist zum damaligen Zeitpunkt nicht aufgegriffen worden, da heute rezente Untersuchungen (vgl. z.B. Clausen 2002; Gruehn 2000) zum Zeitpunkt des Workshops noch nicht vorlagen. Die Befunde wären heute sicherlich anders zu interpretieren.

Das Hauptaugenmerk des Workshops ruhte auf den *Befunden aus dem Mathematik-Leistungstest*, da aufgrund dieser Ergebnisse die Diskussion zur Verbesserung des Curriculums in der einen Schule bzw. die Entscheidung für oder wider die Weiterführung des Curriculums in der anderen Schule geführt werden sollte. Abbildung 4 verdeutlicht diese Ergebnisse. Sowohl die erwartungswidrigen Befunde des Mathematik- als auch des Naturwissenschaftstests konnten in der folgenden Diskussion nicht zufriedenstellend erläutert werden. Durchgängig mit Erleichterung wurde aber zumindest seitens der Lehrkräfte

festgestellt, dass die nach dem Osnabrücker Curriculum unterrichteten Schülerinnen und Schüler nicht hinter die Leistungen vergleichbarer Schüler, die nach den üblichen Curricula beschult wurden, hinterherhinkten, obwohl nur rund 75 Prozent ihrer Unterrichtszeit dem Standardcurriculum gewidmet ist (Cohors-Fresenborg u.a. 2002). Praktisch ungeklärt ist der auch statistisch abgesicherte Vorsprung in den Naturwissenschaften geblieben. Die zeitweilig diskutierten Erklärungsansätze, die auf eine allgemeine Förderung des kognitiven Fähigkeitsniveaus basieren, mussten mit dem Blick auf die Befunde des Mathematiktests verworfen werden. Bis zum Abschluss des Workshops ist diese Frage ungeklärt geblieben.

Bei den Diskussionen der vorgestellten Ergebnisse zeigte sich eine methodische Schwierigkeit in der Vermittlung der Befunde. Selbst statistisch nicht abzusichernde Unterschiede zwischen dem Mittelwert der Versuchs- und der Vergleichsklassen wurden immer wieder von einigen Lehrkräften als bedeutsam interpretiert. Interventionen seitens der moderierenden Evaluatoren konnten immer nur für eine gewisse Zeit bewirken, dass marginale Unterschiede als für die Unterrichtspraxis irrelevant angesehen wurden. Hartnäckig hielt sich das nicht zutreffende Argument, dass die Klassen „immerhin etwas besser als die Vergleichsklassen“ seien.

Im Anschluss an diese Diskussion in der Gesamtgruppe wurden zwei *Kleingruppen aus den Lehrerinnen und Lehrern jeweils einer Schule* gebildet. Hier wurden die Ergebnisse nochmals differenziert nach einzelnen Klassen präsentiert. Eingangs wurde innerhalb der Gruppen geklärt, ob die Identität der Klassen offen gelegt werden soll oder nicht. Wie schon vor dem Workshop angekündigt, blieben die Lehrkräfte der einen Schule bei der anonymisierten Diskussion ihrer vier Klassen, während die Lehrkräfte der anderen die Ergebnisse der von ihnen persönlich unterrichteten Klasse erhielten. Im Laufe der Kleingruppenarbeit wurden die Lehrkräfte der einen Schule gebeten, ihre Klassen zu charakterisieren. Dies sollte mit Blick auf die Lerngeschichte, die allgemeinen Kompetenzen und die Implementierung des Curriculums – im Hinblick auf die Fragestellung, welche Schwerpunkte gesetzt wurden und was nicht realisiert werden konnte – erfolgen. Im Rahmen dieser Diskussion erhielten die Lehrerinnen und Lehrer die Gelegenheit, ihre Erwartungen für die eigene Klasse anhand der am Vormittag diskutierten Folien zu erläutern. Die Ergebnisse der Gruppendiskussion wurden zusammengefasst und in der abschließenden Plenumsdiskussion vorgestellt.

3.4 Abschlussdiskussion im Plenum

In der abschließenden Diskussion wurde nochmals rekapituliert, was Ziel der externen Evaluation war. Hierbei wurden folgende drei Aspekte in das Zentrum der Aussprache gerückt:

3.4.1 Stärken und Schwächen des Osnabrücker Curriculums

Die erwarteten Stärken des Curriculums in den mathematischen Leistungen konnten anhand der eingesetzten TIMSS-Instrumente nicht belegt werden. Neben den eingesetzten TIMSS-Fragebögen hatten die Lehrkräfte ergänzende Testfragen entwickelt, die aus ihrer Sicht curricular valide sind und somit sicher

von den Schülerinnen und Schülern beantwortet werden sollten. Anhand dieser Fragen sollten die besonderen Stärken der Schülerinnen und Schüler identifiziert werden. Allerdings zeigte sich bei den anschließenden Analysen dieser Fragen, dass diese nicht in das edumetrische Messmodell von TIMSS integriert werden konnten und somit keine spezifischen Informationen über die Stärken der Schülerinnen und Schüler identifiziert werden konnten. Bei den Einzelfragen zeigte sich zudem, dass die von den Schülerinnen und Schülern gegebenen Antworten hinter den Lehrererwartungen zurückblieben. Sicher konnte jedoch anhand des TIMSS-Tests festgestellt werden, dass die Leistungen der nach dem Osnabrücker Curriculum unterrichteten Schülerinnen und Schüler nicht hinter den nach dem Standardcurriculum unterrichteten fallen. Die besondere Stärke in den Naturwissenschaften wurde als Phänomen gewürdigt, konnte jedoch nicht zufriedenstellend geklärt werden. Zusammenfassend konnte festgestellt werden, dass keine besonderen Stärken erkannt, aber auch keine Schwächen identifiziert wurden.

3.4.2 Stärken und Schwächen der Implementierung an den beiden Schulen

Bei der Implementierung des Curriculums zeigte sich, dass die Kollegen das Curriculum unterschiedlich stringent in den Klassen implementieren. Diese Diskrepanzen wurden unter anderem auf unterschiedliche Zusammensetzungen und deutlich unterschiedliche Charakteristika der Klassen zurückgeführt. Die Frage, ob eine stringenterere Unterrichtsführung nach dem Osnabrücker Curriculum auch bessere Befunde in den mathematischen Kompetenzen nach sich zieht, ist in der Evaluation methodisch nicht angelegt gewesen und muss daher unbeantwortet bleiben. Mit Bezug auf die Wahrnehmung der Schülerinnen und Schüler des Unterrichts bzw. deren Interessen und Motivationslage wurde festgehalten, dass diesen aufgrund des ihnen fehlenden Vergleichsmaßstabs die ‚Güte‘ des Unterrichts nicht bewusst sei. Einhellig wurde jedoch von den Lehrkräften bestätigt, dass sich die Auswirkungen des Unterrichts wie erwartet von konventionell unterrichteten Klassen unterscheiden würde.

3.4.3 Konsequenzen für die Weiterentwicklung des Curriculums und für die Arbeit der Schulen

Die Befürworter des Curriculums, welche die Evaluation zur Weiterentwicklung nutzen wollen, sehen ihre Arbeit in den Ergebnissen bestätigt. Sie werden das Curriculum weiterführen und -entwickeln. In der Schule, in der die Weiterführung zur Entscheidung anstand, wurden die Ergebnisse ebenso positiv aufgenommen, jedoch wollte und konnte im Rahmen des Workshops noch keine abschließende Entscheidung über die Weiterführung getroffen werden. Die Befunde sollten zunächst noch mit anderen beteiligten Lehrkräften der Schule besprochen werden.

4. Kritische Würdigung

In einer professionell geplanten und durchgeführten Schulentwicklung, in deren Zentrum traditionellerweise der Unterricht steht, wird „keine Maßnahme ohne vorherige Diagnose“ (Horster/Rolff 2001) durchgeführt. Schulen entwickeln sich, ob sie wollen oder nicht. Daher ist es sinnvoll, sich die ständig wandelnden Anforderungen bewusst zu machen und zielgerichtet darauf zu rea-

gieren. Die im Rahmen der externen Evaluation des Osnabrücker Curriculums ermittelten Ergebnisse haben offensichtlich zur Klärung der Frage beigetragen, ob bzw. wie das Curriculum an den beteiligten Schulen beigetragen soll.

Bei der Rückmeldung der Ergebnisse ist seitens der Evaluatoren viel Wert darauf gelegt worden, vor der eigentlichen Ergebnispräsentation die *Erwartungen der Lehrkräfte* in Bezug auf die Schülerantworten zu erfragen. Dies wurde als wichtig angesehen, da Erwartungsanpassungen vermieden werden sollten. Nur so konnte gewährleistet werden, dass die Lehrkräfte sich nochmals bewusst machen, welche Charakteristika ihrer Meinung nach der von ihnen erteilte Unterricht hat und welche Stärken bzw. Schwächen nach ihren Erwartungen die Schülerinnen und Schüler entwickeln. Die zum Teil erwartungswidrigen Befunde konnten kontrastiert, interpretiert und analysiert werden. Die ermittelten Erkenntnisse konnten für die zukünftige Arbeit im Unterricht bzw. für die weitere Entwicklung des Curriculums genutzt werden. Positiv wurde seitens der Organisatoren des Workshops bemerkt, dass sowohl Anlage und Durchführung als auch die Befunde der externen Evaluation von den Lehrkräften durchgängig akzeptiert werden. Diese Akzeptanz war eine wichtige Voraussetzung für die erfolgreiche Durchführung der Evaluation, da eine ablehnende Haltung bei der Interpretation von erwartungswidrigen Befunden eher zur externen Attribuierung denn zur kritischen Selbstreflexion seitens der Lehrkräfte hätte führen können.

Evaluationen werden immer mit dem Ziel durchgeführt, spätere Handlungen auf Basis der ermittelten Befunde zu initiieren. Ziel der vorgestellten Evaluation war, Erkenntnisse über die Wirksamkeit des neu entwickelten Mathematikcurriculums zu gewinnen und so Steuerungswissen für die weitere Entwicklung zu erhalten.

Neben der Beantwortung der Evaluationsfragen zeigte sich auch für die beteiligten Evaluatoren der *Workshop als Feld neuer Erfahrungen*. So konnten sie feststellen, in welcher Weise unterschiedliche Präsentationsformen von nicht statistisch ausgebildeten Lehrkräften interpretiert werden. Besonders auffällig war bei den Diskussionen der ständige Rückgriff auf nur marginale und statistisch nicht abzusichernde Unterschiede sowohl in Einstellungs- als auch Leistungsskalen. Kleinste und aus pragmatischer Sicht irrelevante ‚Unterschiede‘ wurden immer wieder aufgegriffen und auf unzulässige Weise interpretiert. Diese Beobachtung hatte für die Evaluatoren, die sich insbesondere mit der Rückmeldung von Ergebnissen aus Schulleistungsstudien beschäftigen, einen hohen Erkenntniswert. Als Konsequenz wurde in nachfolgenden Untersuchungen bei der Ergebnisrückmeldung besonderes Augenmerk auf die Präsentation der Ergebnisse gelegt und es wurde versucht, anhand von veränderten Präsentationsformen Fehlinterpretationen vorzugreifen. So wurden im Rahmen der Internationalen Grundschul-Lese-Untersuchung (Bos u.a. 2003) nur noch dann Ergebnisse vergleichend zurückgemeldet, wenn diese sich statistisch signifikant unterschieden. Bei indifferenten Befunden wurde jeweils nur der klassenspezifische Mittelwert berichtet (Schwippert 2004). Ein vergleichbares Vorgehen wurde auch bei der Rückmeldung von Ergebnissen der PISA-Untersuchung gewählt (Watermann/Stanat/Kunter/Klieme/Baumert 2003).

Literatur

- Baumert, Jürgen; Rainer H. Lehmann, Manfred Lehrke, Bernd Schmitz, Marten Clausen, Ingmar Hosenfeld, Olaf Köller, Johanna Neubrand 1997: TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske + Budrich
- Bos, Wilfried; Eva-Maria Lankes, Manfred Prenzel, Knut Schwippert, Gerd Walther, Renate Valtin (Hg.) 2003: Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Münster: Waxmann
- Clausen, Marten 2002: Unterrichtsqualität: Eine Frage der Perspektive? Münster: Waxmann
- Cohors-Fresenborg, Elmar 2001: Mathematik als Werkzeug zur Wissensrepräsentation: das Osnabrücker Curriculum. In: Der Mathematikunterricht, 47, 1, S. 5-13
- Cohors-Fresenborg, Elmar; Christa Kaune 1993: Zur Konzeption eines gymnasialen mathematischen Anfangsunterrichts unter kognitionstheoretischem Aspekt. In: Der Mathematikunterricht, 39, 3, S. 4-11
- Cohors-Fresenborg, Elmar; Knut Schwippert, Eckhard Klieme 2002: The Osnabrueck Curriculum: Mathematics as a Tool for the Representation of Knowledge – An Evaluation Study on the Basis of TIMSS-Instruments. In: Weigand, Hans-Georg; Neill, Neville; Peter-Koop, Andrea; Reiss, Kristina; Turner, Günter; Wollring, Bernd (Hg.): Developments in Mathematics Education in German-speaking Countries – Selected Papers from the Annual Conference on Didactics of Mathematics, Potsdam 2000. Hildesheim: Franzbecker, S. 44-55
- Grigutsch, Stefan; Ulrich Raatz, Günter Törner 1998: Einstellungen gegenüber Mathematik bei Mathematiklehrern. In: Journal für Mathematik-Didaktik, 19, S. 3-45
- Gruehn, Sabine 2000: Unterricht und schulisches Lernen. Münster: Waxmann
- Helmke, Andreas 2001: Internationale Schulleistungsvergleichsforschung. Schlüsselprobleme und Perspektiven. In: Zeitschrift für Pädagogik, 47, 2, S. 155-160
- Horster, Leonhard; Hans-Günter Rolff 2001: Unterrichtsentwicklung. Weinheim: Beltz
- Klieme, Eckhard; Jürgen Baumert, Knut Schwippert 2000: Schulbezogene Evaluation und Schulleistungsvergleiche. Eine Studie im Anschluss an TIMSS. In: Rolff u.a. (Hg.): Jahrbuch der Schulentwicklung. Band 11. Weinheim: Juventa, S. 387-419
- Köller, Olaf; Ulrich Trautwein 2003: Schulqualität und Schülerleistung. Evaluationsstudie über innovative Schulentwicklungsprozesse an fünf hessischen Gesamtschulen. Weinheim: Juventa
- Roux, Susanna; Bernhard Wolf, Petra Becker 2001: Zur Verbindung von qualitativen und quantitativen Strategien in der Externen Empirischen Evaluation des Modellprojekts Kindersituation. In: Empirische Pädagogik, 15, 3, S. 483-499
- Schwippert, Knut 2004: Leistungsrückmeldungen an Grundschulen im Rahmen der Internationalen Grundschul-Lese-Untersuchung (IGLU). In: Empirische Pädagogik, 18, 1, S. 62-81
- Sjuts, Johann 1999: Mathematik als Werkzeug zur Wissensrepräsentation. Osnabrück: Forschungsinstitut für Mathematikdidaktik
- Watermann, Rainer; Petra Stanat, Mareike Kunter, Eckhard Klieme, Jürgen Baumert 2003: Schulrückmeldungen im Rahmen von Schulleistungsuntersuchungen. Das Disseminationskonzept von PISA. In: Zeitschrift für Pädagogik, 49, S. 92-111
- Knut Schwippert*, geb. 1965, PD Dr.; Gewerbelehrer (Lehramt berufliche Schulen/ Oberstufe); 1998 Promotion an der Humboldt Universität zu Berlin; 2004 Habilitation an der Universität Hamburg für das Fach International und Interkulturell Vergleichende Erziehungswissenschaft; seit März 2004 Privatdozentur an der Universität Hamburg.; Anschrift: Universität Hamburg, Fachbereich Erziehungswissenschaft, Institut für International und Interkulturell Vergleichende Erziehungswissenschaft, Sedanstraße 19, 20146 Hamburg
E-Mail: schwippert@erzwiss.uni-hamburg.de