

Manfred Prenzel/Claus H. Carstensen/Martin Senkbeil/
Christian Ostermeier/Tina Seidel

Wie schneiden SINUS-Schulen bei PISA ab?

Ergebnisse der Evaluation eines Modellversuchsprogramms

Zusammenfassung

In Reaktion auf die Ergebnisse der TIMS-Studie wurde das BLK-Modellversuchsprogramm SINUS konzipiert und 1998 mit 180 Schulen über fünf Jahre betrieben. Im vorliegenden Beitrag berichten wir über zentrale Ergebnisse der summativen Evaluation des Programms. Ziel des Programms war es, (1) die Professionalisierung der Lehrkräfte zu unterstützen, (2) die Qualität des mathematisch-naturwissenschaftlichen Unterrichts zu verbessern und (3) die Lernprozesse und Lernergebnisse der Schülerinnen und Schüler zu fördern. Zur Überprüfung der Ausgangslage und der Wirkungen des Programms diente ein Vergleich mit einer repräsentativen Schulstichprobe. Dazu wurden an den SINUS-Schulen Erhebungen mit nationalen PISA-Instrumenten durchgeführt. Die Ergebnisse der Abschlusserhebung im Jahr 2003 zeigen, dass SINUS im Verlauf der Programmzeit auf allen untersuchten Ebenen Wirkungen entfaltet hat. Dies betrifft die erfolgreiche Umsetzung der Projektinhalte auf Seiten der Lehrkräfte, die positive Wahrnehmung des Unterrichts auf Seiten der Schülerschaft sowie die Interessen, Haltungen und Kompetenzen der Schülerinnen und Schüler an SINUS-Schulen. Die schulartspezifischen Analysen zeigen jedoch, dass SINUS nicht in allen Schularten die gleiche Wirksamkeit erzielt hat. In erster Linie scheinen die Hauptschulen, die Schulen mit mehreren Bildungsgängen und die Integrierten Gesamtschulen profitiert zu haben.

Summary

The performance of SINUS schools in the PISA test. Evaluation results of a pilot programme.

In reaction to the German results of the TIMS-study the "BLK-Modellversuchsprogramm" SINUS has been conceptualised and started in 1998. Overall, 180 schools have been participating over a period of five years. In this article the results of the summative evaluation of the programme are presented. The goal of the programme SINUS was (1) to encourage processes of teacher professional development, (2) to foster the quality of mathematics and science instruction, and (3) to enhance the students' attitudes and interests as well as their competencies in mathematics and science. To evaluate the starting conditions as well as the effects of the programme a comparison with a representative German school sample was chosen. In all participating SINUS schools data has been collected on the basis of national PISA instruments. The results of the final data collection in 2003 show effects of SINUS on all dependent variables: the successful implementation of SINUS by the teachers, the positive student perception of the quality of mathematics and science instruction, as well as higher attitudes, interests and competencies compared to the representative sample. The school form specific analyses show, however, that the programme did not effect different school forms in the same way: lower and medium tracked school forms (Hauptschulen, Schulen mit mehreren Bildungsgängen, Integrierte Gesamtschulen) seem to have profited most by the programme.

Schlüsselwörter: Qualitätsentwicklung; Qualitätssicherung; Evaluation; mathematisch-naturwissenschaftlicher Unterricht

Keywords: Quality development; Evaluation; Mathematics and Science Instruction

1 Einleitung

Als die Ergebnisse der *Third International Mathematics and Science Study* (TIMSS) publiziert wurden (vgl. BAUMERT u.a. 1997), löste dies in Deutschland eine breite öffentliche Debatte aus. Diskutiert wurden mögliche Ursachen für die mittelmäßigen Ergebnisse, aber auch Wege, wie mit den erkannten Defiziten umgegangen werden soll. In einer länderübergreifenden Initiative wurde ein BLK-Modellversuchsprogramm zur „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ konzipiert und 1998 mit 180 Schulen in 15 Ländern gestartet (vgl. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung 1997). In dem unter der Abkürzung SINUS bekannt gewordenen Programm entwickeln Lehrkräfte ihren eigenen Unterricht innerhalb eines inhaltlichen und organisatorischen Unterstützungsrahmens weiter. Im vorliegenden Beitrag berichten wir über zentrale Ergebnisse der summativen Evaluation des SINUS-Programms.

2 Das BLK-Programm SINUS: Ein unterrichts- und schulnaher Implementationsansatz

Der Ansatz des SINUS-Programms wird durch folgende Prinzipien bestimmt (vgl. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung 1997; PRENZEL 2000b):

- *Bearbeitung von Problembereichen durch Module.* In einer Expertise wurden zentrale Problembereiche des mathematischen und naturwissenschaftlichen Unterrichts in Deutschland beschrieben (vgl. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung 1997). Zur Bearbeitung dieser Problemzonen wurde ein System von elf so genannten Modulen entworfen und durch Handreichungen für die Weiterentwicklung des Unterrichts konkretisiert. Die Unterrichtsentwicklung setzt auf der Ebene der Schule an. Die Fachgruppen wählen die für sie vordringlichen Module. Sie können Module kombinieren, Schwerpunkte setzen und Stück für Stück umfangreichere Veränderungen vornehmen. Entscheidend ist, dass sie die Arbeit an den Modulen als wirksam für das Lehren und Lernen erleben, in absehbarer Zeit zu Verbesserungen gelangen und ein gemeinsames Verständnis von Unterricht entwickeln.
- *Kooperation und Netzwerke.* Die unterrichtsbezogene Zusammenarbeit stellt ein Leitprinzip des Programms dar, denn sie ist Voraussetzung für die Professionalisierung. Als kleinste Grundeinheit der Kooperation gilt die Fachgruppe an der beteiligten Schule. Die schulübergreifende Zusammenarbeit findet in kleinen Schulnetzwerken, so genannten Schulsets, statt. Der Austausch zwischen den Lehrkräften wird auf Schulset- und Landesebene sowie bundesweit durch Koordinatorinnen und Koordinatoren unterstützt.
- *Qualitätsentwicklung und Qualitätssicherung.* Die beteiligten Schulen sollen mit Verfahren der Qualitätsentwicklung und Qualitätssicherung vertraut gemacht werden. Dies

bedeutet beispielsweise, dass Kolleginnen und Kollegen im gemeinsamen Austausch unterrichtliche Problembereiche identifizieren, zu bearbeitende Ziele festlegen und mit einfachen Verfahren der Selbstevaluation überprüfen, ob sie ihre Ziele erreicht haben. Prozesse einer professionellen Qualitätskontrolle sollen für professionelles Handeln selbstverständlich werden.

- *Wissenschaftliche Anregung und Unterstützung.* Die Arbeit der Lehrkräfte wird durch die konzeptionellen Leitlinien strukturiert und durch Materialien, Fortbildungen und Beratungen extern angeregt und unterstützt. Die wissenschaftliche Unterstützung orientiert sich an dem von den Modulen vorgezeichneten Rahmen. Ein zentraler Programmträger (federführend das Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) in Kiel, in Kooperation mit dem Staatsinstitut für Bildungsforschung (ISB) in München sowie dem Zentrum zur Förderung des mathematisch-naturwissenschaftlichen Unterrichts (Z-MNU) an der Universität Bayreuth) koordiniert die Unterstützungsmaßnahmen und stellt Informationen und Materialien unter anderem über den Internetserver des Programms bereit.
- *Begleitforschung.* Ein weiteres wichtiges Element im Programm ist die Begleitforschung. Sie befasst sich mit konzeptionellen Fragen und ist für die formative sowie summative Evaluation zuständig. Weiterhin untersucht sie im Rahmen von Implementationsforschung Bedingungen einer erfolgreichen Umsetzung des Programms.

3 Ziele des BLK-Programms SINUS

Mit dem BLK-Programm SINUS werden Entwicklungen und Verbesserungen angestrebt, die auf unterschiedlichen Ebenen liegen: (a) Die Professionalisierung der Lehrkräfte, (b) die Gestaltung des Unterrichts und (c) die Lernprozesse und Lernergebnisse der Schülerinnen und Schüler.

- a) *Professionalisierung der Lehrkräfte.* Das Programm zielt auf die professionelle Weiterentwicklung der beteiligten Lehrkräfte und Schulen. Dazu muss die unterrichtsbezogene Kooperation zwischen Lehrkräften angeregt und unterstützt werden, denn Kooperation ist, wie die Schuleffektivitätsforschung zeigt (vgl. SAMMONS 1999), ein zentrales Merkmal effektiver Schulen. Jedoch trifft man in deutschen Schulen häufig noch einen Lehrerindividualismus an, der einer professionellen Weiterentwicklung entgegensteht (vgl. TERHART 2001). Da insbesondere die Kooperation in der Fachgruppe förderlich zu sein scheint (vgl. LUYTEN 1994), regt das Programm vor allem die unterrichtsbezogene Zusammenarbeit zwischen Fachkolleginnen und -kollegen an. Kooperation im Sinne des Programms bedeutet, dass die Lehrkräfte beispielsweise Unterrichtsmaterialien und Prüfungsaufgaben austauschen, wechselseitig hospitieren und gemeinsam Unterricht im eigenen Fach oder auch mit dem anderer Fächer abstimmen. Lehrerfortbildung wird gemeinsam organisiert, auf der Ebene der Fachgruppe(n) oder der gesamten Schule. Ein weiteres Merkmal von Professionalität stellt die unterrichtsbezogene Selbstvergewisserung der Lehrkräfte dar. Das Programm regt die Lehrkräfte an, die Wirkung des eigenen Unterrichtshandelns zu hinterfragen und sensibilisiert für didaktische Problemfelder des Unterrichts. Langfristig sollen sich so in den Programmschulen entsprechende Prozesse der kooperativen Qualitätssicherung und Qualitätsentwicklung etablieren und sich in festen Strukturen niederschlagen.

- b) *Gestaltung des Unterrichts*. Ein zentrales Anliegen des Programms ist die Verbesserung des Unterrichts. Die Lehrkräfte nehmen einzelne Unterrichtsaspekte in den Blick und entwickeln diese modulbezogen Stück um Stück weiter. Im Mathematikunterricht, aber auch in den naturwissenschaftlichen Fächern, nehmen zum Beispiel Aufgaben eine zentrale Rolle ein. In Modul 1 „Weiterentwicklung der Aufgabenkultur“ werden deshalb die Lehrkräfte angeregt, Aufgaben nach ihrer didaktischen Funktion im Unterricht zu betrachten und das Aufgabenspektrum zu erweitern. Bei der Erarbeitung neuen Stoffs, beim Durcharbeiten und Üben oder beim systematischen Wiederholen zurückliegender Inhalte können komplexere, kognitiv herausfordernde Aufgaben das verständnisvolle Lernen unterstützen (vgl. LAMPERT 1990; REUSSER/STEBLER 1997). In Modul 2 „Naturwissenschaftliches Arbeiten“ wird beispielsweise die Rolle des Experiments und dessen Bedeutung für die Förderung naturwissenschaftlichen Denkens thematisiert (vgl. HOFSTEIN/LUNETTA 2004; WHITE/FREDERIKSEN 1998). Fehler als Lerngelegenheiten zu betrachten und diese systematisch für Lernprozesse zu nutzen, steht im Zentrum von Modul 3 „Lernen aus Fehlern“ (vgl. OSER/HASCHER/SPYCHIGER 1999). Die Module 8 „Kooperatives Arbeiten von Schülerinnen und Schülern“ und 9 „Verantwortung für das eigene Lernen stärken“ beispielsweise greifen wichtige Merkmale selbstregulierten und kooperativen Arbeitens auf (vgl. RENKL 1997) und nehmen ebenfalls domänenspezifische Lernstrategien in den Blick (vgl. PERELS/GÜRTLER/SCHMITZ 2005).
- c) *Lernprozesse und Lernergebnisse der Schülerinnen und Schüler*. Das BLK-Programm SINUS strebt insgesamt eine veränderte Unterrichtskultur an, die sich auf der Ebene der beteiligten Schulen etablieren soll. Die Module dienen dazu als inhaltlicher Orientierungsrahmen, und alle Module zielen darauf ab, verständnisvolles Lernen und positive Haltungen sowie Interesse in Bezug auf Mathematik und Naturwissenschaften zu fördern.

Eine veränderte Aufgabenkultur, sinnvolle Experimente sowie kumulatives und selbstreguliertes Lernen stellen Bausteine für einen Unterricht dar, der für Schülerinnen und Schüler eine andere Qualität von Lernprozessen ermöglicht: Lernprozesse, die sich dadurch auszeichnen, dass Unterricht als kognitiv aktivierend und selbstbestimmt motivierend erlebt wird (vgl. SEIDEL 2003). Diese veränderte Qualität von Lernprozessen wiederum führt mit einer höheren Wahrscheinlichkeit zu verbesserten Kompetenzen und Interessen in der Mathematik und den Naturwissenschaften (vgl. KLIEME/RAKOCZY 2003; WHITE/FREDERIKSEN 1998).

Basierend auf dem aktuellen Stand der domänenspezifischen und allgemeinen Lehr-Lern-Forschung zielen die elf Module in SINUS also insgesamt auf eine veränderte Lehr-Lern-Kultur, deren wichtigstes Ziel das vertiefte, verständnisorientierte Lernen der Schülerinnen und Schüler ist (vgl. PRENZEL/OSTERMEIER 2003). Die an den Modulen orientierte Weiterentwicklung des Unterrichts findet in der Zusammenarbeit der Kolleginnen und Kollegen in den Fachgruppen statt. Positive Ergebnisse dieser Arbeit werden demnach nicht nur bei einzelnen Klassen, sondern langfristig auf der Ebene der Schule erwartet. Im Laufe der Zeit sollte die kooperative Unterrichtsentwicklung anhand der Module das Kompetenzniveau, die Motivation und das Interesse der Schülerinnen und Schüler in Mathematik und den Naturwissenschaften verbessern.

4 Evaluation des Programms

Die Evaluation spielt eine zentrale Rolle im Programm und betrifft folgende Bereiche (vgl. OSTERMEIER u.a. 2004; OSTERMEIER/PRENZEL 2005): (a) Schulinterne Evaluation und Qualitätssicherung, (b) Formative Evaluation, (c) Implementationsforschung und (d) Summative Evaluation.

- a) *Schulinterne Evaluation und Qualitätssicherung.* Das Programm strebt an, Prozesse der Qualitätsentwicklung und Qualitätssicherung auf Schulebene einzurichten (vgl. PRENZEL 2000a). Die beteiligten Lehrkräfte werden angeregt und unterstützt, sich auf gemeinsame Ziele zu verständigen und einfache Verfahren der Selbstevaluation (kollegiale Evaluation, Hospitationen, kurze Schülerfragebögen o.ä.) zu verwenden.
- b) *Formative Evaluation.* In der prozessbegleitenden formativen Evaluation wird das programminterne Unterstützungsangebot überprüft und dessen Nutzung untersucht. Diese Aspekte wurden in dem Programm mit Hilfe zweier Befragungen der beteiligten Lehrkräfte beleuchtet. Im Blickpunkt steht, wie die Lehrerinnen und Lehrer den Programmansatz akzeptieren und wie die Arbeit an den Schulen in Gang gekommen ist (vgl. OSTERMEIER 2004). Ergänzende Informationen liefern die Befragungen der Schulleitungen und kleine Stichproben von Eltern sowie von Schülerinnen und Schülern.
- c) *Implementationsforschung.* Die unterschiedlichen Erhebungsstränge im Programm erlauben es weiterhin, Fragen der Implementationsforschung nachzugehen: Welche Bedingungen wirken sich förderlich, welche hinderlich für die Umsetzung der kooperativen Qualitätsentwicklung aus? In den Blick genommen werden beispielsweise wichtige Ausgangsbedingungen an den beteiligten Programmschulen (vgl. OSTERMEIER u.a. 2004). Zudem sind die durch das Programm bereitgestellten Unterstützungsangebote Gegenstand der Untersuchung. Die Angaben zur Bewertung und Nutzung dieser Angebote geben wichtige Hinweise auf die Umsetzung der Programmkonzeption und auf Bedingungen erfolgreicher Unterrichtsentwicklung (vgl. OSTERMEIER/PRENZEL 2005).
- d) *Summative Evaluation.* Bezugspunkt für die summative Evaluation sind die Ziele des Programms, nämlich die Professionalisierung der Lehrkräfte, die Weiterentwicklung des Unterrichts und die Verbesserung der Lernprozesse und Lernergebnisse der Schülerinnen und Schüler. Eine systematische Evaluation von groß angelegten und weit reichenden Interventionen im Praxisfeld steht vor großen Herausforderungen. Um Wirkungen des Programms kontrollieren zu können, muss zu Beginn des Programms geklärt werden, unter welchen Ausgangsbedingungen die Schulen in die Arbeit eintreten. Eine solche Überprüfung der Ausgangslage auf Schulebene (z.B. Kompetenzen, motivationale Orientierungen, familiärer Hintergrund der Schülerinnen und Schüler) und deren Vergleich mit einer repräsentativen Schulstichprobe dienen als Bezugspunkte für eine summative Evaluation der Ergebnisse nach Abschluss des Programms. Eine Erhebung an den SINUS-Schulen im Jahr 2000 unter Verwendung nationaler PISA-Instrumente nahm deshalb die Frage in den Blick, inwieweit die Programmschulen hinsichtlich ihrer Ausgangslagen mit einer repräsentativen Stichprobe deutscher Schulen der PISA-Erweiterung 2000 (vgl. BAUMERT u.a. 2002) vergleichbar sind. Untersucht wurden die Bereiche familiärer Hintergrund, Schulklima, motivationale Orientierungen und Präferenzen im mathematisch-naturwissenschaftlichen Bereich sowie die Kompetenzen der Schülerinnen und Schüler in Mathematik und Naturwissenschaften. Insgesamt belegen die Ergebnisse der Eingangserhebung, dass die SINUS-Schulen sich

nicht systematisch von den PISA-Schulen unterscheiden (vgl. OSTERMEIER u.a. 2004). Im Zentrum des vorliegenden Beitrags steht nun die summative Abschlusserhebung des Programms und die Frage, inwieweit die SINUS-Schulen sich auch – drei Jahre später – in der nachfolgenden PISA-Erhebung (vgl. PRENZEL u.a. 2005) von einer repräsentativen Stichprobe von Schulen in Deutschland unterscheiden.

5 Fragestellungen

Ausgehend von den Zielen der Programmkonzeption werden im Folgenden vier Fragen untersucht, die Aufschluss über den Erfolg von SINUS geben:

- 1) Inwieweit unterscheiden sich Lehrkräfte aus SINUS-Schulen von Kolleginnen und Kollegen aus der PISA-Stichprobe in der praktizierten Kooperation und Qualitätsentwicklung auf Schulebene? Da das Programm systematisch die Kooperation zwischen Lehrkräften und die Qualitätsentwicklung an den Schulen angeregt hat, wird angenommen, dass sich hier deutliche Effekte nachweisen lassen. An den SINUS-Schulen sollte also mehr Kooperation und mehr Qualitätssicherung betrieben werden als an der für Deutschland repräsentativen Schulstichprobe der PISA-Erweiterung 2003.
- 2) Wie nehmen Schülerinnen und Schüler an SINUS-Schulen am Ende des Programms den Mathematikunterricht in ihrer Klasse wahr? Es wird angenommen, dass sich der Unterricht an SINUS-Schulen im Verlauf des Programms systematisch in Richtung eines stärker kognitiv aktivierenden und motivierenden Unterrichts verändert hat und dies auch von den Schülerinnen und Schülern wahrgenommen wird. SINUS-Schulen unterscheiden sich deshalb systematisch von PISA-Schulen in der Wahrnehmung des Unterrichts aus Schülersicht.
- 3) Welche Ergebnisse zeigen sich an SINUS-Schulen am Ende des Programms in Bezug auf motivationale Orientierungen und das Interesse an Mathematik und den Naturwissenschaften? Befunde aus der Lehr-Lern-Forschung verweisen auf die bedeutende Rolle individueller Voraussetzungen für die Lernergebnisse von Schülerinnen und Schülern. Es wird deshalb angenommen, dass sich Effekte von SINUS auf Interessen und Haltungen in begrenztem Maße nachweisen lassen. Dennoch wird vermutet, dass sich SINUS-Schulen systematisch in motivationalen Merkmalen von PISA-Schulen unterscheiden.
- 4) Welche Kompetenzen weisen Schülerinnen und Schüler an SINUS-Schulen am Ende des Programms im Vergleich zu Schülerinnen und Schülern aus der PISA-Stichprobe 2003 auf? Ähnlich zur vorherigen Fragestellung wird vermutet, dass sich Effekte von SINUS auf die Kompetenzen der Lernenden in einem eingeschränkten Maße nachweisen lassen, da die Ergebnismessungen bei Schülerinnen und Schülern einer Alters- und Klassenstichprobe vorgenommen wird, die an der Schule möglicherweise nur zu einem Teil einen an SINUS orientierten Unterricht erhalten hatte. Es wird aber wiederum postuliert, dass sich ein Unterschied zwischen SINUS-Schulen im Vergleich zu PISA-Schulen in der mathematischen und naturwissenschaftlichen Kompetenz nachweisen lässt.

6 Design und Methoden der summativen Evaluation

6.1 Design und Erhebungsinstrumente

Für die summative Evaluation des BLK-Programms SINUS wurde ein Untersuchungsdesign gewählt, das die Schulen der PISA-Erweiterung als Kontrollgruppe nutzt. Ein erster Vergleich mit dieser Kontrollgruppe wurde bei PISA 2000 vorgenommen (vgl. OSTERMEIER u.a. 2004). Während die Gruppe der SINUS-Schulen weitgehend gleich blieb, diente 2003 eine neue repräsentative Stichprobe von Schulen als Kontrollgruppe, um Entwicklungen auf der Schulebene erfassen zu können. Aus diesen Gründen wurden Erhebungsinstrumente eingesetzt, die einen systematischen Vergleich mit repräsentativen Schulstichproben erlauben. Auch für den zweiten Erhebungszyklus der PISA-Studie im Jahr 2003 wurden von deutschen Expertengruppen zusätzliche Leistungstests in Mathematik und den Naturwissenschaften, abgestimmt auf die Altersgruppe der Fünfzehnjährigen bzw. auf Schülerinnen und Schüler der neunten Jahrgangsstufe, konzipiert, entwickelt und erprobt (vgl. BLUM u.a. 2004; ROST u.a. 2004; SENKBEIL u.a. 2005). Diese auf nationaler Ebene entwickelten Instrumente korrelieren sehr hoch mit den internationalen Tests und wurden im Rahmen der SINUS-Abschlusserhebung im Jahr 2003 eingesetzt (vgl. PRENZEL u.a. 2004a, 2005).

Neben den Kompetenzen in Mathematik und Naturwissenschaften wurden motivationale Orientierungen und Interessen der Schülerinnen und Schüler erhoben sowie ihre Einschätzung des Mathematikunterrichts. Die Lehrkräfte wurden unter anderem zur Kooperation im Kollegium sowie zu den Bereichen schulinterner Fortbildung und Evaluation befragt. Bei den Skalen griffen wir auf Verfahren zurück, die auch in der PISA-Erhebung 2003 angewendet wurden (Tabelle 1). Die vollständigen Skalen sind bei RAMM u.a. (vgl. 1995) zu finden.

6.2 Durchführung der Erhebung

Die Abschlusserhebung zur summativen Evaluation des SINUS-Programms fand in der ersten Hälfte des Jahres 2003 statt. Es waren alle 180 SINUS-Schulen für eine Teilnahme vorgesehen. Allerdings wurden 42 SINUS-Schulen in die Stichprobe der PISA-2003-Erhebung gezogen; sie beteiligten sich deshalb nicht an der SINUS-Erhebung. Aus Gründen des Datenschutzes konnten ferner die sächsischen SINUS-Schulen nicht in die Abschlussuntersuchung einbezogen werden. Insgesamt beteiligten sich 102 SINUS-Schulen an der Abschlusserhebung zur summativen Evaluation; weitere 42 nahmen regulär an PISA teil.

In der Eingangserhebung im Jahr 2000 wurden an den SINUS-Schulen alle Schülerinnen und Schüler des neunten Jahrgangs getestet (vgl. OSTERMEIER u.a. 2004). Überprüft man die Messgenauigkeit der Schulkenntwerte, so zeigt sich, dass eine Erhöhung der Anzahl getesteter Klassen ab zwei Klassen pro Schule nicht mehr zu niedrigeren Messfehlern in der Schätzung des Schulmittelwerts führt. Aufgrund dieses Befundes und aufgrund ökonomischer Überlegungen wurde entschieden, in der Abschlusserhebung nicht den kompletten neunten Jahrgang einer Schule, sondern zwei zufällig ausgewählte neunte Klassen pro Schule zu testen.

Tabelle 1: Übersicht zu den im vorliegenden Beitrag verwendeten Skalen
(Fragebogen für Schülerinnen und Schüler, Fragebogen für Lehrkräfte)

Bereich/Skala	Anzahl Items	Cronbach's Alpha	Beispielitem
<i>Lehrkräfte</i>			
Häufigkeit der Kooperation ^{1, 2, 3}	3	0.69	Wie oft treffen Sie sich mit Mathematiklehrkräften, um Lehrstoff oder Unterrichtsmethoden zu besprechen oder zu planen?
Integration von Kollegen ^{1, 3}	7	0.86	Als Ansprechpartner für neue Lehrkräfte steht hier immer jemand zur Verfügung.
Schulinterne Abstimmungen ^{1, 3}	9	0.79	Die Schulleitung hat immer ein offenes Ohr für die Sorgen und Probleme der Lehrerinnen und Lehrer.
Umgang mit Fortbildungsmaßnahmen ^{1, 3}	10	0.83	Lehrerinnen und Lehrer stellen ihre Erfahrungen und neue Ideen aus Weiterbildungsveranstaltungen im Kollegium in der Regel vor.
Nutzung lernunterstützender Innovationen ^{1, 3}	5	0.67	Wie hoch ist bei Ihnen der Anteil von unkonventionellen Wegen für die Lösung der Aufgaben des Lehrplanes?
Evaluative Maßnahmen ^{1, 3}	12	0.80	Wir arbeiten an unserer Schule mit Methoden der Selbstevaluation der Schülerinnen und Schüler (z.B. Lerntagebücher, Portfolios).
<i>Wahrnehmung des Mathematikunterrichts (Schülerinnen und Schüler)</i>			
Alltagsbezug ⁴	3	0.75	Unsere Mathematiklehrkraft zeigt uns an Beispielen aus dem täglichen Leben, wozu man Mathematik brauchen kann.
Kognitiv aktivierende Aufgaben ^{4, 5}	8	0.79	Unsere Mathematiklehrkraft stellt oft Fragen, die man nicht spontan beantworten kann, sondern die zum Nachdenken zwingen.
Kognitive Selbständigkeit ⁴	8	0.76	Unsere Mathematiklehrkraft stellt auch Aufgaben, bei denen man mehrere Lösungswege zeigen muss.
Diskussion von multiplen Lösungen im Mathematikunterricht ⁴	5	0.71	Im Mathematikunterricht diskutieren wir häufiger Lösungsvorschläge, die wir zuvor in Gruppen erarbeitet haben.
Anspruchsvolles Üben ⁴	3	0.66	Wir üben im Mathematikunterricht mit Aufgaben, durch die man die Sache besser versteht.
Wahrgenommene Kooperationskompetenzen der Lehrkraft im Mathematikunterricht ⁴	5	0.86	In Mathematik lernen wir, wie man zusammenarbeiten kann, dass es allen etwas bringt.
<i>Motivationale Orientierungen (Schülerinnen und Schüler)</i>			
Interessen und Freude an Mathematik ⁶	4	0.90	Mich interessiert das, was ich in Mathematik lerne.
Selbstkonzept in Mathematik ⁶	5	0.91	Im Mathematikunterricht verstehe ich sogar die schwierigsten Aufgaben.

1 Den theoretischen Hintergrund dieser Skalen bilden die von SCHEERENS/BOSKER (vgl. 1997) anhand von Metaanalysen abgeleiteten effective-enhancing factors.

2 Skala aus dem Projekt „QuaSUM“ (vgl. LEHMANN u.a. 2000)

3 vgl. PISA 2003 (Nationaler Lehrerfragebogen) (vgl. RAMM u.a. im Druck)

4 PISA 2003 (Nationaler Schülerfragebogen) (vgl. RAMM u.a. im Druck)

5 BAUMERT u.a. (vgl. 2004)

6 PISA 2003 (Internationaler Schülerfragebogen) (vgl. RAMM u.a. im Druck)

6.3 Beschreibung der Stichproben

Da Lernergebnisse stark mit der besuchten Schulart zusammenhängen, betrachten wir zunächst die Verteilung der getesteten Schülerinnen und Schüler auf unterschiedliche Schularten. Tabelle 2 zeigt die Verteilung der Schülerinnen und Schüler getrennt für die SINUS- und die PISA-Stichprobe. Wie bereits in der Erhebung der Ausgangslage unterscheiden sich auch in der Abschlusserhebung die Verteilungen deutlich (vgl. OSTERMEIER u.a. 2004). Während in der SINUS-Stichprobe im Vergleich zu PISA größere Anteile der Schülerinnen und Schüler ein Gymnasium oder eine Integrierte Gesamtschule besuchen, liegen die entsprechenden Anteile für Haupt- und Realschulen unter denen der Vergleichsgruppe.

Tabelle 2: Verteilung getesteter Schülerinnen und Schüler auf Schularten.
Darstellung für beide Stichproben (SINUS, PISA)

Schulform	SINUS (in Prozent)	PISA (in Prozent)
Hauptschule	10.8	22.9
Schule mit mehreren Bildungsgängen	14.7	11.1
Integrierte Gesamtschule	12.7	9.7
Realschule	19.6	25.9
Gymnasium	42.2	30.4

Im Folgenden werden wir aufgrund dieser ungleichen Verteilung die Ergebnisse jeweils getrennt nach Schularten berichten.

6.4 Auswertungsverfahren

Während in der Eingangserhebung 2000 die Testhefte vom zweiten Testtag der PISA-Studie unverändert übernommen wurden, bearbeiteten die Schülerinnen und Schüler in der Abschlusserhebung 2003 eine Auswahl von Aufgaben, die in einer eigenen Zusammenstellung zu Testheften vorgegeben wurden; damit wurde die Testzeit der Abschlusserhebung einschließlich Schülerfragebogen auf 120 Minuten begrenzt. In den PISA-Testheften kommt jeder Aufgabenblock an vier unterschiedlichen Positionen vor (zu Beginn und nach 30 oder 60 oder 90 Minuten), in den SINUS-Testheften gibt es dagegen nur zwei Positionen, zu Beginn und nach 30 Minuten. Die Verbindung der Leistungsskalen beider Studien erfolgte daher auf einer Auswahl der Itembearbeitungen, nämlich anhand der Aufgaben, die jeweils in den ersten 30 Minuten bearbeitet wurden. Die Skalierung der Leistungsdaten und Fragebogenskalen erfolgte mit den gleichen Verfahren wie in PISA 2003 (vgl. CARSTENSEN u.a. 2004), d.h mit einer Skalierung nach dem Rasch-Modell; zur genauen Bestimmung von Populationskennwerten werden als Personenmesswerte Multiple Imputationen (Plausible Values) mit einem so genanntem Hintergrundmodell berechnet.

7 Ergebnisse

Die Ergebnisse der SINUS-Schulen werden im Vergleich mit der für Deutschland repräsentativen Stichprobe der PISA-E-Schulen dargestellt. Da bei sehr großen Stichproben (wie den beiden vorliegenden) auch geringfügige Differenzen zu statistisch signifikanten Befunden führen können und Signifikanztests daher nur eine eingeschränkte Aussagekraft besitzen, werden in dieser Arbeit deskriptiv die Mittelwertsdifferenzen beider Stichproben in Form von Effektstärken berichtet. Effektstärken geben unmittelbar Auskunft über die Größe der Mittelwertsdifferenzen und können inhaltlich interpretiert werden. Aufgrund des strengen Evaluationsdesigns können Effektstärken von $d \geq 0.20$ nach der Terminologie von COHEN (vgl. 1988) zugunsten der SINUS-Schulen als inhaltlich relevante und bedeutsame Mittelwertsdifferenzen angesehen werden. Signifikante Unterschiede werden in den Tabellen durch Fettdruck der zugehörigen Effektstärken gekennzeichnet.

Da die Verteilung der SINUS-Schulen auf die Schularten deutlich von der diesbezüglichen Verteilung der Schulen in der PISA-Stichprobe abweicht (siehe auch Abschnitt 6.3), werden die Vergleiche schulartspezifisch durchgeführt. Zunächst werden die Befunde zu prozessbezogenen Merkmalen auf Seiten der Lehrkräfte (Einschätzung der Kooperation) und auf Seiten der Schülerinnen und Schüler (Wahrnehmung des Mathematikunterrichts) berichtet (Abschnitte 7.1 und 7.2). Anschließend werden als nicht-kognitive Ergebnismerkmale das auf Mathematik bezogene Interesse und Selbstkonzept der Schülerinnen und Schüler dargestellt (Abschnitt 7.3). Abschnitt 7.4 vergleicht die mathematische und naturwissenschaftliche Kompetenz der befragten Jugendlichen in SINUS- und PISA-Schulen.

7.1 Einschätzungen der Lehrkräfte zu Kooperation und Qualitätsentwicklung

Die Ergebnisse zu den Einschätzungen der Lehrkräfte sind in Tabelle 3 in Form von Skalenwerten und Effektstärken für jede Schulart wiedergegeben. Die dort angegebenen Skalenwerte sind so normiert, dass der Mittelwert über alle Einschätzungen 0 und die Standardabweichung 1 betragen („z-Standardisierung“). Positive Skalenwerte bedeuten demnach eine höhere Ausprägung, negative eine geringere Ausprägung als im Durchschnitt. Die Skalenwerte wie auch die standardisierten Mittelwertsdifferenzen (d) weisen dabei auf schulartspezifische Effekte hin (Effektstärken mit $d \geq 0.20$ sind durch Fettdruck hervorgehoben). In Abbildung 1 sind die Skalenwerte der SINUS- und PISA-Schulen für das Merkmal „Häufigkeit der Kooperationen“ in Abhängigkeit von der Schulart grafisch wiedergegeben. Wie die Tabelle 3 zeigt, unterscheiden sich die Schularten – generell und unabhängig von den beiden Stichproben – erheblich im Umfang ihrer Kooperationspraxis. Die schulartbezogene Spannweite in den Einschätzungen beträgt sowohl für SINUS- als auch für PISA-Schulen bis zu 0.8 Standardabweichungen (SINUS: Häufigkeit der Kooperation; PISA: Evaluative Maßnahmen).

Tabelle 3: Schulartspezifische Mittelwerte (z-standardisiert) und Effektstärken (d) für die Einschätzungen der Lehrkräfte

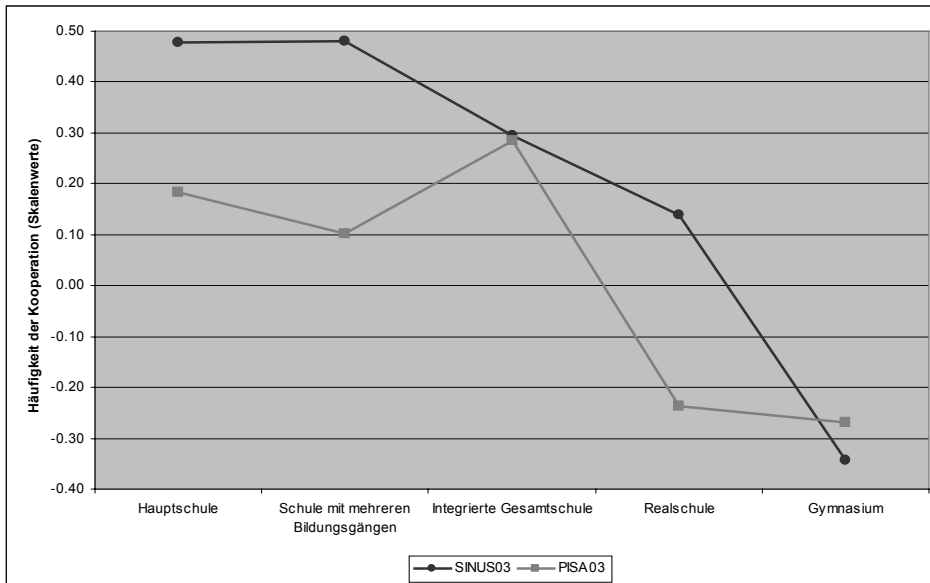
Skala	Hauptschule			MBG			IGS			Realschule			Gymnasium		
	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d
Häufigkeit															
Kooperation	0.48	0.18	0.29	0.48	0.10	0.42	0.29	0.28	0.01	0.14	-0.24	0.36	-0.34	-0.09	-0.27
Integration von Kollegen	0.33	-0.06	0.40	0.40	0.01	0.43	0.34	0.18	0.16	0.00	-0.12	0.12	-0.09	-0.05	-0.04
Schulinterne Abstimmung	0.31	0.25	0.06	0.51	0.09	0.41	0.53	0.46	0.09	0.21	-0.14	0.38	-0.38	-0.32	-0.07
Fortbildung	0.24	-0.24	0.45	0.51	0.28	0.27	0.17	0.16	0.01	0.12	-0.16	0.27	0.00	-0.04	0.03
Lernunterstützende Innovation	0.54	0.10	0.44	0.33	0.11	0.25	0.09	-0.13	0.22	0.15	-0.31	0.47	0.14	-0.08	0.21
Evaluation	0.21	-0.03	0.23	0.44	0.29	0.16	0.50	0.53	-0.03	-0.05	-0.22	0.18	-0.31	-0.09	-0.23

Anmerkung: MBG: Schule mit mehreren Bildungsgängen; IGS: Integrierte Gesamtschule; Bei signifikanten Unterschieden sind die Effektstärken fett gedruckt (zweiseitiger Test; Alpha = .05).

Besonders auffällig ist, dass der inhaltliche und fachliche Austausch zwischen den Lehrkräften in den Hauptschulen, in Schulen mit mehreren Bildungsgängen und in Integrierten Gesamtschulen vergleichsweise stark verbreitet ist, während jener in Realschulen und Gymnasien in sehr viel geringerem Ausmaß stattfindet (vgl. hierzu z.B. SENKBEIL u.a. 2004). Wie den Skalenwerten in Tabelle 3 zu entnehmen ist, zeigt sich dieses Muster übereinstimmend bei allen berücksichtigten Merkmalen sowohl bei den SINUS- als auch bei den PISA-Schulen.

Zum anderen zeigen sich in den Hauptschulen, Schulen mit mehreren Bildungsgängen und Realschulen für die Mehrzahl der Merkmale bedeutsame Effektstärken ($d \geq 0.20$) zugunsten der SINUS-Schulen, die zum Teil nahezu mittleren Effektgrößen von $d = 0.50$ entsprechen. Dies gilt insbesondere für die Hauptschulen und Schulen mit mehreren Bildungsgängen in Bezug auf die Kooperation (z.B. *Kooperationshäufigkeit*, *Integration von Kollegen*, *schulinterne Abstimmung*), aber auch hinsichtlich eines kreativen Umgangs mit Lehrplanvorgaben (Skala *Lernunterstützende Innovation*). Bei den Skalen zur *Fortbildung* und zu *evaluativen Maßnahmen* fallen die Mittelwertsunterschiede für die genannten Schularten etwas geringer aus. Gleichwohl zeigen sich auch bei diesen Merkmalen höhere Ausprägungen für die SINUS-Stichprobe mit Effektstärken, die nahezu durchgehend in der Größenordnung von $d \geq 0.20$ liegen.

Abbildung 1: Skalenwerte für die Häufigkeit von Kooperationen in SINUS- und PISA-Schulen in Abhängigkeit von der Schulart



Bei den Integrierten Gesamtschulen und den Gymnasien konnten – mit Ausnahme der Skala „Lernunterstützende Innovation“ – die intendierten Wirkungen hinsichtlich einer verbesserten und intensivierten Kooperationspraxis offenbar nicht erreicht werden. Für das Gymnasium ergeben sich zum Teil sogar bedeutsame Vorteile zugunsten der PISA-Schulen (*Häufigkeit der Kooperation, evaluative Maßnahmen*), wenngleich auf insgesamt sehr niedrigem Aktivitätsniveau. Bei Betrachtung der Mittelwerte für die Integrierten Gesamtschulen fällt auf, dass eine intensivere Nutzung der Kooperationsmöglichkeiten und die Durchführung evaluativer Maßnahmen erheblich häufiger zum Schulalltag gehören als in den anderen Schularten. Insofern kann für die Integrierten Gesamtschulen in gewisser Weise von einem „Deckeneffekt“ ausgegangen werden, der vergleichsweise wenig Spielraum für eine weitere Steigerung des Aktivitätsniveaus lässt. Immerhin lassen sich auch für diese Schulart noch kleine Effekte hinsichtlich der *Integration von Kolleginnen und Kollegen* und insbesondere bei der *Nutzung lernunterstützender Innovationen* feststellen. Bis auf vereinzelte Ausnahmen ergeben sich insgesamt auf Schulartebene relativ homogene Mittelwertsdifferenzen zwischen den SINUS- und PISA-Schulen. Dieser Befund weist darauf hin, dass das Programm nicht nur punktuell, sondern in die Breite gewirkt hat.

7.2 Veränderung des Unterrichts

Eine wichtige Frage bei der Evaluation von Programmen betrifft nicht nur die im vorangegangenen Abschnitt vorgenommene Überprüfung, wie erfolgreich die Lehrkräfte neue Orientierungen und Verhaltensweisen entwickeln, sondern auch deren mittelbare Effekte auf die letztlich entscheidende Zielgruppe der Schülerinnen und Schüler. Dementsprechend sollten sich die Maßnahmen zur Veränderung der Lehr-Lernkultur auch in einer veränderten Unterrichtswahrnehmung auf Seiten der Schülerinnen und Schüler widerspiegeln. Die Ergebnisse in *Tabelle 4* geben Auskunft darüber, inwieweit sich die Einschätzungen der Schülerschaft in den SINUS-Projektschulen von den Wahrnehmungen der Jugendlichen in den PISA-Schulen in projektrelevanten Unterrichtsmerkmalen unterscheiden.

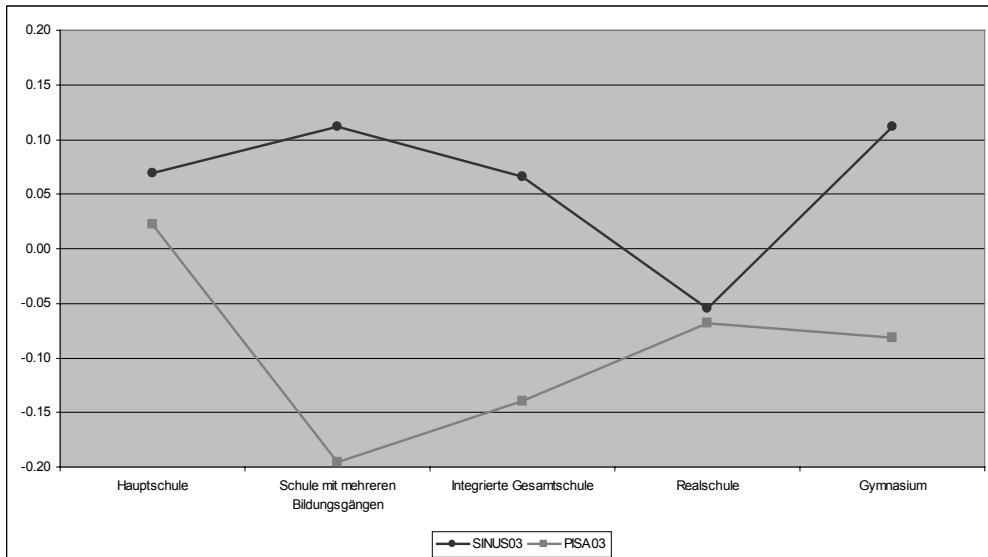
Tabelle 4: Schulartspezifische Mittelwerte (z-standardisiert) und Effektstärken (d) für die Unterrichtswahrnehmung in Mathematik durch die Schülerinnen und Schüler

Skala	Hauptschule			MBG			IGS			Realschule			Gymnasium		
	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d
Alltagsbezug	0.37	0.32	0.05	0.01	-0.03	0.05	0.09	-0.07	0.16	-0.13	-0.09	-0.03	0.00	-0.24	0.24
Kognitiv aktivierende Aufgaben	0.07	0.02	0.04	0.11	-0.20	0.32	0.07	-0.14	0.19	-0.05	-0.07	0.01	0.11	-0.08	0.20
Kognitive Selbstständigkeit	0.07	0.12	-0.04	0.05	-0.13	0.20	0.01	-0.16	0.16	-0.03	-0.08	0.05	0.06	-0.05	0.12
Diskussion	0.08	0.16	-0.08	0.03	-0.15	0.20	0.02	-0.01	0.03	-0.06	0.00	-0.06	-0.03	-0.05	0.02
Anspruchsvolles Üben	0.17	0.09	0.07	0.08	-0.08	0.16	-0.01	-0.12	0.11	-0.02	0.00	-0.02	0.01	-0.11	0.13
Kooperationskompetenz Mathematikunterricht	0.60	0.44	0.17	0.23	-0.09	0.34	0.16	-0.02	0.20	0.04	-0.13	0.17	-0.18	-0.40	0.23

Anmerkung: MBG: Schule mit mehreren Bildungsgängen; IGS: Integrierte Gesamtschule; Bei signifikanten Unterschieden sind die Effektstärken fett gedruckt (zweiseitiger Test; Alpha = .05).

Wie aus den Effektstärken in *Tabelle 4* ersichtlich wird, fallen die Unterschiede zwischen den SINUS- und PISA-Schulen über alle Schularten im Durchschnitt deutlich geringer aus, als es bei den Lehrkräften der Fall war. An den Skalenwerten (s. hierzu auch *Abbildung 2* für die Skala *Kognitiv aktivierende Aufgaben*) ist zusätzlich erkennbar, dass auch die Wahrnehmungen der Schülerinnen und Schüler vergleichsweise gering zwischen den Schularten variieren. Sie betragen maximal 0.3 Standardabweichungen. Ausnahmen bilden nur die Skalen zum *Alltagsbezug* und zur *Kooperationskompetenz der Lehrkräfte*.

Abbildung 2: Skalenwerte für die Wahrnehmung der Schülerschaft hinsichtlich kognitiv aktivierender Aufgaben in SINUS- und PISA-Schulen in Abhängigkeit von der Schulart



Trotz der generell kleinen Mittelwertsdifferenzen ergeben sich bedeutsame Wirkungen zugunsten der SINUS-Schulen bei den Schulen mit mehreren Bildungsgängen. Die Effektstärken erreichen bis auf eine Ausnahme (Skala *Anspruchsvolles Üben*) die Größenordnung von $d \geq 0.20$. Dieses Ergebnis stimmt mit den Einschätzungen der Lehrkräfte insofern überein, als dass auch dort die Vorteile der SINUS-Schulen bei dieser Schulart am homogensten und deutlichsten zu Tage treten. Bedeutsame Effekte zugunsten der SINUS-Schulen lassen sich ansonsten noch bei den Gymnasien (für die Skalen *Alltagsbezug*, *kognitiv aktivierende Aufgaben* und die *Kooperationskompetenz der Lehrkräfte*) und bei den Integrierten Gesamtschulen feststellen (für die Skalen *kognitiv aktivierende Aufgaben* und *Kooperationskompetenz der Lehrkräfte* und mit Einschränkung für die Skalen *Alltagsbezug* und *kognitive Selbständigkeit*). Die relativ gleichförmigen und zum Teil bedeutsamen Effekte bei den Integrierten Gesamtschulen unterstützen somit die Vermutung eines „Deckeneffekts“ bei den Einschätzungen der Lehrkräfte (siehe Abschnitt 7.1).

Für die Hauptschulen und Realschulen ergeben sich keinerlei bedeutsame Unterschiede. Angesichts der erheblichen Effekte, die für diese Schularten auf Seiten der Lehrkräfte ermittelt werden konnten, erstaunt dieser Befund. Entsprechend finden sich auf Schulebene auch nur relativ geringe Korrelationen zwischen den Angaben der Lehrkräfte und den Unterrichtswahrnehmungen der Schülerinnen und Schüler. Dies gilt für beide Schülergruppen in etwa gleichermaßen. Die einzige Ausnahme betrifft die *Kooperationskompetenz der Lehrkräfte* aus Sicht der Schülerschaft. Hier ergeben sich für die SINUS-Schulen erheblich straffere Zusammenhänge mit den Angaben der Lehrkräfte zu ihrer *Kooperationspraxis*. Während die Korrelationen für die SINUS-Schulen zwischen $r = 0.22$ und $r = 0.37$ liegen, betragen sie für die PISA-Schulen zwischen $r = -0.12$ und $r = 0.21$. Entsprechend wird die *Kooperationskompetenz der Lehrkräfte* in den SINUS-Schulen über alle Schularten erheblich höher eingeschätzt als in den PISA-Schulen (siehe auch Tabelle 4).

7.3 Das mathematikbezogene Interesse und Selbstkonzept der Schülerinnen und Schüler in SINUS und PISA

Eines der langfristigen Ziele des SINUS-Programms stellt die Steigerung des mathematik- und naturwissenschaftsbezogenen Interesses und Selbstkonzepts dar. Im Rahmen dieser Evaluationsstudie konnten für die SINUS-Schulen das Interesse und Selbstkonzept der Schülerinnen und Schüler in Bezug auf das Fach Mathematik erhoben und mit Daten der repräsentativen PISA-Stichprobe verglichen werden. Die Ergebnisse hierzu sind in Tabelle 5 dargestellt und zeigen ein mit den bisher berichteten Daten weitgehend übereinstimmendes Befundbild.

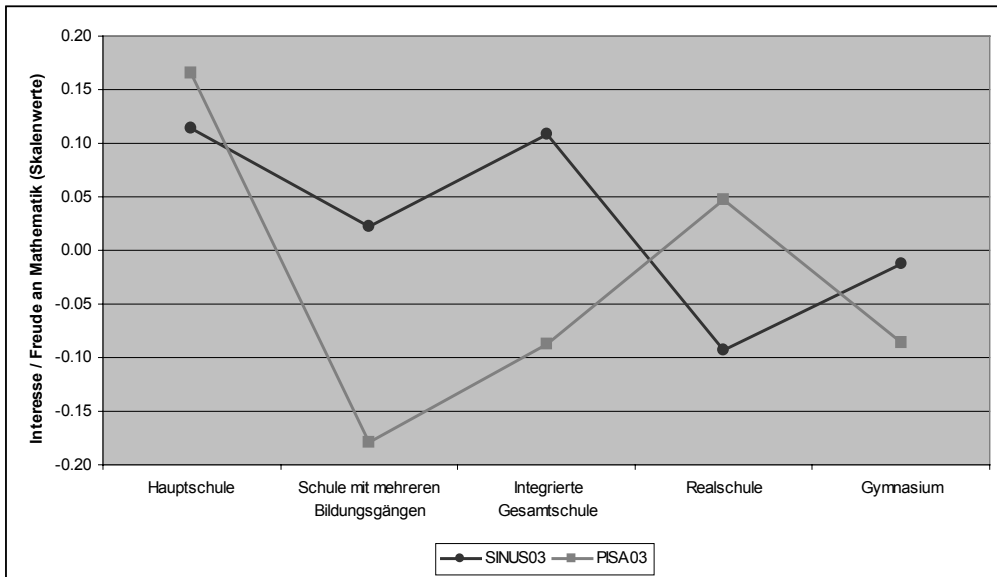
Tabelle 5: Schulartspezifische Mittelwerte (z-standardisiert) und Effektstärken (d) für das mathematikbezogene Interesse und Selbstkonzept der Schülerinnen und Schüler

Skala	Hauptschule			MBG			IGS			Realschule			Gymnasium		
	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d
Interesse und Freude an Mathematik	0.11	0.17	-0.05	0.02	-0.18	0.21	0.11	-0.09	0.21	-0.09	0.05	-0.14	-0.01	-0.09	0.08
Selbstkonzept Mathematik	0.05	0.06	-0.01	-0.03	-0.12	0.10	0.08	-0.08	0.17	-0.02	0.03	-0.05	0.02	-0.04	0.06

Anmerkung: MBG: Schule mit mehreren Bildungsgängen; IGS: Integrierte Gesamtschule; Bei signifikanten Unterschieden sind die Effektstärken fett gedruckt (zweiseitiger Test; Alpha = .05).

Bei Betrachtung der Skalenwerte (Mittelwerte) sind analog zu den Skalen zur Unterrichtswahrnehmung nur relativ geringfügige Unterschiede zwischen den Schularten festzustellen. Die schulartbezogene Spannbreite umfasst bei den PISA-Schulen etwa 0.4 Standardabweichungen, bei den SINUS-Schulen sogar nur 0.2 Standardabweichungen. Dies zeigt sich auch in *Abbildung 3*, in der die schulartspezifischen Skalenwerte (Mittelwerte) für das mathematikbezogene Interesse der SINUS- und PISA-Schülerschaft grafisch veranschaulicht werden.

Abbildung 3: Schulartspezifische Skalenwerte für das mathematikbezogene Interesse der Schülerinnen und Schüler in SINUS- und PISA-Schulen



Wie *Tabelle 5* zu entnehmen ist, zeigen sich – wiederum analog zu den Befunden zu den Merkmalen der Unterrichtswahrnehmung – bei den Jugendlichen der Schulen mit mehreren Bildungsgängen und der Integrierten Gesamtschulen bedeutsame Mittelwertsdifferenzen für das mathematikbezogene Interesse zugunsten der SINUS-Schulen ($d = 0.21$ für Schulen mit mehreren Bildungsgängen; $d = 0.21$ für Integrierte Gesamtschulen). Hinsichtlich des mathematikbezogenen Selbstkonzepts fallen die Effektstärken mit $d = 0.10$ (Schulen mit mehreren Bildungsgängen) und $d = 0.17$ etwas geringer aus, jedoch wiederum zugunsten der SINUS-Schulen. Zu beachten ist hierbei, dass das mathematikbezogene Interesse und Selbstkonzept eng miteinander verwobene Konstrukte darstellen (vgl. z.B. HELMKE 1992; HELMKE/VAN AKEN 1995) und hoch miteinander korrelieren (in der vorliegenden Stichprobe: $r = 0.71$). Für die anderen Schularten sind keine bedeutsamen Differenzen festzustellen. Als Ergebnis ist somit auch festzuhalten, dass die vom SINUS-Programm angestrebten Veränderungen in der Lehr-Lernkultur zu keinen negativen Effekten auf Seiten der Jugendlichen in Bezug auf das fachbezogene Interesse oder Selbstkonzept geführt haben, sondern insbesondere in bestimmten Teilgruppen (Schularten) diese stärker förderten.

7.4 Die mathematischen und naturwissenschaftlichen Kompetenzen der Schülerinnen und Schüler in SINUS und PISA

In diesem Abschnitt werden die mathematischen und naturwissenschaftlichen Kompetenzen der Jugendlichen in den SINUS- und PISA-Schulen gegenüber gestellt. Hierbei werden zwei getrennte Auswertungen verbunden. Separate Analyseschritte sind deshalb notwendig, weil ein Teil der Schulen ($N = 42$), die am SINUS-Programm teilnehmen, in der

regulären PISA-Stichprobe enthalten sind (vgl. Kap. 6.2). Die Ergebnisse dieser Schulen, die auf den Daten der internationalen Tests zur mathematischen und naturwissenschaftlichen Grundbildung beruhen, werden hier nicht gesondert dargestellt, sondern in die Befunde der eigentlichen Untersuchung zur SINUS-Evaluation integriert. An dieser hat die Mehrzahl der SINUS-Schulen ($N = 102$) teilgenommen, wobei die Kompetenzwerte auf den Daten der nationalen Testkonstruktionen in PISA 2003 beruhen. Für die im Folgenden dargestellten Kennwerte in Tabelle 6 und Abbildung 4 bedeutet dies, dass sich die schulspezifischen Mittelwerte auf die Daten der eigentlichen SINUS-Evaluation mit 102 Schulen im Vergleich mit den Schulen der PISA-Stichprobe beziehen. Bei den wiederum aufgeführten Effektstärken sind die oben erwähnten 42 SINUS-Schulen, die Bestandteil der PISA-E-Stichprobe waren, mit einbezogen. Da sich durch die Integration dieser zusätzlichen Schulen nur relativ geringfügige Änderungen in den Mittelwerten ergeben, können die in Tabelle 6 dargestellten Kennwerte als guter Indikator für die Gesamtstichprobe gelten. Die Kompetenzwerte wurden hierbei so normiert, dass der Mittelwert für die PISA-Stichprobe in beiden Fächern $M = 50$ und die Standardabweichung $SD = 10$ betragen.

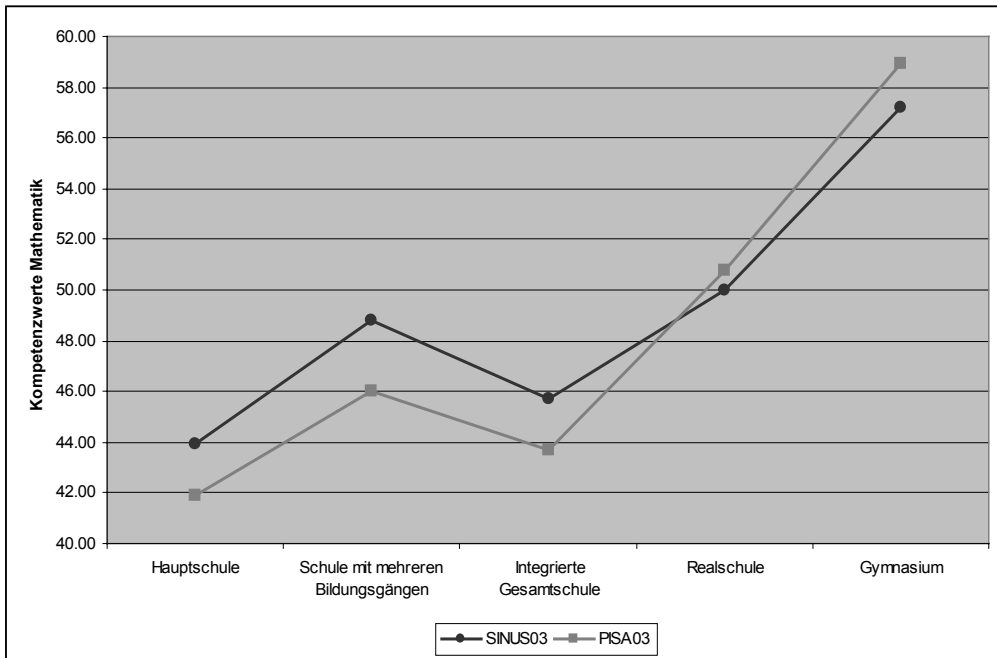
Tabelle 6: Schulartspezifische Kompetenzwerte in Mathematik und Naturwissenschaften ($N = 102$ Schulen) und Effektstärken d ($N = 144$ Schulen)

Kompetenzbereich	Hauptschule			MBG			IGS			Realschule			Gymnasium		
	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d	SINUS	PISA	d
Mathematik	43.90	41.90	0.18	48.80	46.00	0.26	45.70	43.70	0.18	50.00	50.80	-0.09	57.20	58.90	-0.17
Naturwissenschaften	45.10	41.00	0.36	49.70	46.80	0.27	49.10	45.00	0.37	51.50	51.00	0.04	59.20	58.70	0.06

Anmerkung: MBG: Schule mit mehreren Bildungsgängen; IGS: Integrierte Gesamtschule;
Bei signifikanten Unterschieden sind die Effektstärken fett gedruckt (zweiseitiger Test; $\alpha = .05$).

Bei Betrachtung der Kompetenzwerte für Mathematik ergibt sich für die Schulen mit mehreren Bildungsgängen und für die Integrierten Gesamtschulen ein mit den bisherigen Befunden weitgehend übereinstimmendes Bild. Danach erzielten die Schülerinnen und Schüler dieser SINUS-Schulen um 0.26 (Schulen mit mehreren Bildungsgängen) bzw. um 0.18 Standardabweichungen (Integrierte Gesamtschulen) höhere Werte als die Jugendlichen der PISA-Schulen. Die Effektstärken fallen damit ähnlich stark aus wie bei den Einschätzungen zum mathematikbezogenen Interesse und Selbstkonzept. Für die Hauptschulen ergibt sich eine vergleichbare Effektstärke ($d = 0.18$) zugunsten der SINUS-Schulen, obwohl für diese in Bezug auf die Unterrichtswahrnehmung und das mathematikbezogene Interesse und Selbstkonzept keine bedeutsamen Unterschiede gegenüber den PISA-Schulen festzustellen waren. Während analog zu den anderen schülerbezogenen Merkmalen für die Realschulen kein bedeutsamer Kompetenzunterschied zu beobachten ist, schneiden die Gymnasien der PISA-Stichprobe etwas besser ab als die der SINUS-Schulen ($d = -0.17$). Hinsichtlich der Gesamtstichprobe fällt auf, dass die Spannbreite zwischen den Schularten – wie auch Abbildung 6 veranschaulicht – für die SINUS-Schulen etwas geringer ausfällt als bei den PISA-Schulen (SINUS: 15 Punkte; PISA: 17 Punkte), was vornehmlich auf die höheren Mittelwerte der Hauptschulen und Schulen mit mehreren Bildungsgängen in SINUS zurückzuführen ist.

Abbildung 4: Schulartspezifische Kompetenzwerte in Mathematik in SINUS- und PISA-Schulen (N = 102)



Für die naturwissenschaftliche Kompetenz ergeben sich insoweit identische Ergebnisse, als wiederum die Hauptschulen, die Schulen mit mehreren Bildungsgängen und die Integrierten Gesamtschulen in SINUS höhere Kompetenzwerte erzielen als die entsprechenden Schulstichproben in PISA. Die Vorteile zugunsten der SINUS-Schulen fallen bei Effektstärken von $d = 0.36$ (Hauptschule), $d = 0.27$ (Schule mit mehreren Bildungsgängen) und $d = 0.37$ (Integrierte Gesamtschule) in der naturwissenschaftlichen Kompetenz erheblich größer aus als bei der mathematischen Kompetenz. Für die Realschulen und Gymnasien lassen sich keine bedeutsamen Differenzen ermitteln. Auch hier zeigt sich für die SINUS-Schulen eine geringere Kompetenzstreuung zwischen den Schularten (SINUS: 14.1 Punkte vs. PISA: 18.2 Punkte), die wiederum auf die vergleichsweise höheren Werte der Schülerinnen und Schüler in den Hauptschulen und den Schulen mit mehreren Bildungsgängen zurückzuführen ist. Geht man bei PISA nach einer groben Faustregel davon aus, dass eine Differenz von 0.4 Standardabweichungen in den Kompetenzwerten ungefähr einem Abstand von einem Schuljahr entspricht (vgl. PRENZEL u.a. 2004b), sind die hier ermittelten Unterschiede in der Größenordnung von 0.2 Standardabweichungen zugunsten der SINUS-Schulen als beträchtlich und von praktischer Relevanz zu bezeichnen.

8 Diskussion

Die Ergebnisse der Evaluationsstudie zeigen, dass SINUS im Verlauf der Programmzeit auf allen untersuchten Ebenen Wirkungen entfaltet hat. Dies betrifft die erfolgreiche Um-

setzung der Projektinhalte auf Seiten der Lehrkräfte, die positive Wahrnehmung des Unterrichts auf Seiten der Schülerschaft sowie die Interessen, Haltungen und Kompetenzen der Schülerinnen und Schüler an SINUS-Schulen. Trotz dieser im Allgemeinen als positiv einzuschätzenden Befunde verdeutlichen die schulartspezifischen Analysen jedoch, dass SINUS nicht in allen Schularten die gleiche Wirksamkeit erzielt hat. In erster Linie und vornehmlich scheinen die Hauptschulen, die Schulen mit mehreren Bildungsgängen und die Integrierten Gesamtschulen vom SINUS-Programm profitiert zu haben.

Ein wesentliches Ziel des SINUS-Programms war es, die Kooperation zwischen Lehrkräften und die Qualitätssicherung an den Schulen aufzubauen und zu fördern. Für die Evaluation wurde angenommen, dass sich in diesem Bereich die deutlichsten Wirkungen von SINUS zeigen sollten. Die Ergebnisse der Evaluationsstudie belegen für die Bereiche der Kooperation und der Qualitätssicherung die größten Effektstärken. Diese Befunde sprechen dafür, dass die Programminhalte an den teilnehmenden Schulen gut umgesetzt wurden. Der Effekt zeigt sich allerdings nicht bei den Integrierten Gesamtschulen, was zum Teil auf einen „Deckeneffekt“ zurückgeführt werden kann. Dennoch finden sich auch hier an den Integrierten Gesamtschulen Vorteile zugunsten der SINUS-Schulen in Bezug auf die Nutzung lernunterstützender Innovationen. Einzig bei den Gymnasien scheint die Umsetzung des SINUS-Programms hinsichtlich der Kooperation weniger gut gelungen zu sein. Unter Umständen wirken hier bestimmte „gymnasiale Traditionen“, die noch stärker mit einem Bild der Lehrkraft als individuell agierendem Fachexperten verbunden sind (vgl. TERHART 2001). Auffällig ist zudem, dass sich bedeutsame Effektstärken in Hauptschulen, Schulen mit mehreren Bildungsgängen und Realschulen nicht nur punktuell, sondern weitgehend über alle Merkmale nachweisen lassen. Dieser Befund unterstreicht, dass SINUS an den Schulen in der Breite gewirkt und nicht nur Einzeltechniken (z.B. spezifische Gruppentechniken) vermittelt hat. Die Vielzahl der bedeutsamen Effektstärken lässt weiterhin vermuten, dass sich die Lehrkräfte an SINUS-Schulen intensiv mit dem Programm, der Programmidee und den Modulen auseinander gesetzt haben.

Ein weiteres Ziel von SINUS war, auf der Basis der Module zu Veränderungen des Unterrichts zu gelangen. Insgesamt geben die Ergebnisse der Evaluation Hinweise darauf, dass sich der Unterricht an SINUS-Schulen in relevanten Merkmalen des Unterrichts geändert hat: Schülerinnen und Schüler aus SINUS-Schulen berichten im Vergleich zu PISA-Schulen in verstärktem Maße davon, dass in ihrem Unterricht Alltagsbezüge hergestellt werden. Die Aufgaben wirken kognitiv aktivierend, die Übungsformen werden wahrgenommen als anspruchsvoll und abwechslungsreich, und die Schülerinnen und Schüler setzen sich selbstständig mit Lerninhalten auseinander. Damit erreicht der Unterricht an SINUS-Schulen aus der Sicht der Schülerinnen und Schüler eher die Qualitätsmerkmale, die von der Lehr-Lern-Forschung als bedeutsam erachtet werden (vgl. PRENZEL 2000b). Trotz dieser positiven Befunde führt der Vergleich mit PISA-Schulen bei den Unterrichtsmerkmalen zu geringeren Effektstärken und zeigt auch geringere Unterschiede zwischen den Schularten. Bei diesen Befunden sollte auch berücksichtigt werden, dass Einschätzungen von Schülerinnen und Schülern bei der Beschreibung des Unterrichtsgeschehens oder bei der Beurteilung didaktischer Kompetenzen zum Teil kritisch diskutiert werden und ihnen eine bedingte Aussagekraft zugeschrieben wird (vgl. BAUMERT u.a. 2004). So finden BAUMERT u.a. (vgl. 2004) Hinweise darauf, dass die Unterrichtseinschätzungen von Schülerinnen und Schülern an Hauptschulen mit Vorbehalt zu sehen sind. Vor diesem Hintergrund kann man vermuten, dass sich bei den Hauptschulen trotz der großen Effektstärken auf Seiten der Kooperation zwischen Lehrkräften keine positiven Effekte für SINUS in der

Unterrichtswahrnehmung zeigen. Anders verhält es sich bei den Schulen mit mehreren Bildungsgängen und Integrierten Gesamtschulen: Hier treten deutliche Effekte bei einer Vielzahl von Unterrichtsmerkmalen auf, und gleichzeitig wird die von den Lehrkräften berichtete Kooperation im Kollegium auch von der Schülerschaft wahrgenommen.

Letztlich hat SINUS vor dem Hintergrund der TIMSS- und PISA-Befunde zum Ziel, die Kompetenzen sowie die Interessen und Haltungen der Schülerinnen und Schüler in Mathematik und den Naturwissenschaften zu fördern. Die Ergebnisse der Evaluation zum Interesse und Selbstkonzept verhalten sich konform mit den bisherigen Befunden: Es lassen sich klare Vorteile zugunsten von SINUS bei Schulen mit mehreren Bildungsgängen und Integrierten Gesamtschulen beobachten. Dieser Befund ist bedeutungsvoll, da man auf der Basis der Befunde aus der Forschung davon ausgehen kann, dass das fachbezogene Interesse und das Selbstkonzept relativ stabile Persönlichkeitsmerkmale darstellen und die mit SINUS verbundenen Lernaktivitäten und -prozesse (z.B. mehr nachdenken, selbständiger arbeiten, kognitiv fordernde Aufgaben) bei der Schülerschaft auch zu einer Verunsicherung oder Reaktanz führen könnten. Insofern kann es durchaus als Erfolg gelten, wenn das Interesse und das Selbstkonzept nicht sinken und wenn das Interesse bei den Schulen mit mehreren Bildungsgängen und den Integrierten Gesamtschulen zu steigen scheint.

Bei den erreichten Kompetenzen fallen die Effektstärken in den Naturwissenschaften deutlich zugunsten von SINUS aus. In diesem Bereich liegen die Hauptschulen, die Schulen mit mehreren Bildungsgängen und die Integrierten Gesamtschulen des SINUS-Programms in der Kompetenzentwicklung im Durchschnitt zwischen einem halben und einem ganzen Schuljahr vor den Schulen der repräsentativen Vergleichsstichprobe aus PISA. In der Mathematik beträgt der Vorsprung bei drei der fünf Schularten ein halbes Schuljahr, jedoch nicht bei den Realschulen und Gymnasien.

Zusammenfassend zeigen sich also in einigen Schularten (Hauptschulen, Schulen mit mehreren Bildungsgängen, Integrierte Gesamtschulen) konsistente Effekte des Programms in einer bedeutsamen Größenordnung. Führt man sich bei der Bewertung dieser Ergebnisse noch einmal die „Strenge“ des Evaluationsdesigns vor Augen, das ja nicht nur die Klassen von engagierten SINUS-Lehrkräften untersucht hat, muss von flächendeckenden Effekten auf Schulebene ausgegangen werden. Insbesondere haben die Schularten vom SINUS-Programm profitiert, für die ein stärkerer Entwicklungsbedarf festzustellen ist (vgl. PRENZEL u.a. 2004, 2005). Die aktuellen Befunde weisen darauf hin, dass die Leistungszuwächse zwischen PISA 2000 und PISA 2003 in den Naturwissenschaften und in der Mathematik eher bei den leistungsstärkeren Schülerinnen und Schülern erzielt wurden (vgl. ROST u.a. 2005). Tatsächlich sind insbesondere die leistungsschwächeren Schülerinnen und Schüler auf die didaktische Qualität des Unterrichts angewiesen. Kompetenzstarken Schülerinnen und Schülern gelingt es auch, schlecht aufbereitete Unterrichtsinhalte zu erschließen. Insofern zeigen die Befunde der SINUS-Evaluation, dass sich Programme zur Professionalisierung und Unterrichtsentwicklung lohnen.

Literatur

- BAUMERT u.a. 1997 = BAUMERT, J./LEHMANN, R./LEHRKE, M./SCHMITZ, B./CLAUSEN, M./HOSENFELD, I./KÖLLER, O./NEUBRAND, J. (1997): TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. – Opladen.
- BAUMERT u.a. 2002 = BAUMERT, J./ARTELT, C./KLIEME, E./NEUBRAND, M./PRENZEL, M./SCHIEFELE, U./SCHNEIDER, W./STANAT, P./TILLMANN, K.-J./WEIB, M. (Hrsg.) (2002): PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. – Opladen.
- BAUMERT u.a. 2004 = BAUMERT, J./KUNTER, M./BRUNNER, M./KRAUSS, S./BLUM, W./NEUBRAND, M. (2004): Mathematikunterricht aus Sicht der PISA-Schülerinnen und -Schüler und ihrer Lehrkräfte. In: PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROLFF, H.-G./ROST, J./SCHIEFELE, U. (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. – Münster, S. 314-354.
- BLUM u.a. 2004 = BLUM, W./NEUBRAND, M./EHMKE, T./SENKBEIL, M./JORDAN, A./ULFIG, F./CARSTENSEN, C. H. (2004): Mathematische Kompetenz. In: PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROLFF, H.-G./ROST, J./SCHIEFELE, U. (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. – Münster, S. 47-92.
- BORTZ, J./DÖRING, N. (1995): Forschungsmethoden und Evaluation für Sozialwissenschaftler. – Berlin.
- Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (1997): Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“. – Bonn: Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie.
- CARSTENSEN u.a. 2004 = CARSTENSEN, C. H./KNOLL, S./ROST, J./PRENZEL, M. (2004): Technische Grundlagen. In: PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROLFF, H.-G./ROST, J./SCHIEFELE, U. (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. – Münster, S. 371-387.
- COHEN, J. (1988): Statistical power analysis for the behavioral sciences – New York.
- HELMKE, A. (1992): Selbstvertrauen und schulische Leistungen. – Göttingen.
- HELMKE, A./VAN AKEN, M. (1995): The causal ordering of academic achievement and self-concept of achievement during elementary school: A longitudinal study. In: Journal of Educational Psychology, Vol. 87, pp. 624-637.
- HOFSTEIN, A./LUNETTA, V. N. (2004): The laboratory in science education: Foundations for the twenty-first century. In: Science Education, Vol. 88,1, pp. 28-54.
- KLIEME, E./RAKOCZY, K. (2003): Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In: BAUMERT, J./ARTELT, C./KLIEME, E./NEUBRAND, M./PRENZEL, M./SCHIEFELE, U./SCHNEIDER, W./TILLMANN, K.-J./WEIB, M. (Hrsg.): PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. – Opladen, S. 333-60.
- LAMPERT, M. (1990): When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. In: American Educational Research Journal, Vol. 27,1, pp. 29-63.
- LEHMANN u.a. 2000 = LEHMANN, R. H./PEEK, R./GANSFUB, R./LUTKAT, S./MÜCKE, S./BARTH, I. (2000): QuaSUM: Qualitätsuntersuchungen an Schulen zum Unterricht in Mathematik. Ergebnisse einer repräsentativen Untersuchung im Land Brandenburg. – Potsdam: Ministerium für Bildung, Jugend und Sport im Land Brandenburg.
- LUYTEN, H. (1994): Stability of school effects in Dutch secondary education: The impact of variance across subjects and years. In: International Journal of Educational Research, Vol. 21,2, pp. 197-216.
- OSER, F./HASCHER, T./SPYCHIGER, M. (1999): Lernen aus Fehlern. Zur Psychologie des „negativen“ Wissens. In: ALTHOF, W. (Hrsg.): Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern. – Opladen, S. 11-41.
- OSTERMEIER, C. (2004): Kooperative Qualitätsentwicklung in Schulnetzwerken. Eine empirische Studie am Beispiel des BLK-Modellversuchsprogramms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“ (SINUS). – Münster.
- OSTERMEIER, C./PRENZEL, M. (2005): What can we learn from different forms of evaluation: Experiences from a quality development program in science and mathematics instruction. In: BENNETT, J./

- HOLMAN, J./MILLAR, R./WADDINGTON, D. (Hrsg.): Making the difference: Evaluation as a tool for improving science education. – Münster, pp. 145-158.
- OSTERMEIER u.a. 2004 = OSTERMEIER, C./CARSTENSEN, C. H./PRENZEL, M./GEISER, H. (2004): Kooperative unterrichtsbezogene Qualitätsentwicklung in Netzwerken: Ausgangsbedingungen für die Implementation im BLK-Modellversuchsprogramm SINUS. In: Unterrichtswissenschaft, 32. Jg., H. 3, S. 215-237.
- PERELS, F./GÜRTLER, T./SCHMITZ, B. (2005): Training of self-regulatory and problem-solving competence. In: Learning and Instruction, Vol. 15,2, pp. 123-140.
- PRENZEL, M. (2000a): Qualitätsentwicklungsansätze im Rahmen des BLK-Modellversuchsprogramms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts“. In: Qualitätssicherung im internationalen Wettbewerb. – Bonn: Forum Bildung, S. 719-728.
- PRENZEL, M. (2000b): Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts: Ein Modellversuchsprogramm von Bund und Ländern. In: Unterrichtswissenschaft, 28. Jg., H. 2, S. 103-126.
- PRENZEL, M./OSTERMEIER, C. (2003): Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts – Ein unterrichtsbezogenes Qualitätsentwicklungsprogramm. In: Beiträge zur Lehrerbildung. Fachzeitschrift zu Theorie und Praxis der Aus- und Weiterbildung von Lehrerinnen und Lehrern, 21. Jg., H. 2, S. 265-276.
- PRENZEL u.a. 2004a = PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROLFF, H. G./ROST, J./SCHIEFELE, U. (2004a): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. – Münster.
- PRENZEL u.a. 2004b = PRENZEL, M./DRECHSEL, B./CARSTENSEN, C. H./RAMM, G. (2004b): PISA 2003 – eine Einführung. In: PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROLFF, H. G./ROST, J./SCHIEFELE, U.: PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. – Münster, S. 13-46.
- PRENZEL u.a. 2005 = PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M., PEKRUN, R./ROST, J./SCHIEFELE, U. (2005): PISA 2003 – Der zweite Vergleich der Länder in Deutschland. Was wissen und können Jugendliche? – Münster.
- RAMM u.a. 2005 = RAMM, G./RÖNNEBECK, S./KIENZL, A./CARSTENSEN, C. H./PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROST, J./SCHIEFELE, U. (2005): Handbuch zu den Skalen aus PISA 2003. – Münster [im Druck].
- RENKL, A. (1997): Lernen durch Lehren: Zentrale Wirkmechanismen beim kooperativen Lernen. – Wiesbaden.
- REUSSER, K./STEBLER, R. (1997): Every word problem has a solution – The social rationality of mathematical modelling in schools. In: Learning and Instruction, Vol. 7,4, pp. 309-327.
- ROST u.a. 2004 = ROST, J./WALTER, O./CARSTENSEN, C. H./SENKBEIL, M./PRENZEL, M. (2004): Naturwissenschaftliche Kompetenz. In: PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROLFF, H.-G./ROST, J./SCHIEFELE, U. (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. – Münster, S. 111-146.
- ROST u.a. 2005 = ROST, J./SENKBEIL, M./WALTER, O./CARSTENSEN, C. H./PRENZEL, M. (2005): Naturwissenschaftliche Grundbildung im Ländervergleich. In: PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROST, J./SCHIEFELE, U. (2005): PISA 2003 – Der zweite Vergleich der Länder in Deutschland. Was wissen und können Jugendliche? – Münster, S. 113-134.
- SAMMONS, P. (1999): School effectiveness: Coming of age in the twenty-first century. – Lisse.
- SCHEERENS, J./BOSKER, R. J. (1997): The foundations of educational effectiveness. – Oxford.
- SEIDEL, T. (2003): Lehr-Lernskripts im Unterricht. – Münster.
- SENKBEIL u.a. 2004 = SENKBEIL, M./DRECHSEL, B./ROLFF, H.-G./BONSEN, M./ZIMMER, K./LEHMANN, R. H./NEUMANN, A. (2004): Merkmale und Wahrnehmungen von Schule und Unterricht. In: PRENZEL, M./BAUMERT, J./BLUM, W./LEHMANN, R./LEUTNER, D./NEUBRAND, M./PEKRUN, R./ROLFF, H.-G./ROST, J./SCHIEFELE, U. (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. – Münster, S. 296-314.
- SENKBEIL u.a. 2005 = SENKBEIL, M./ROST, J./CARSTENSEN, C. H./WALTER, O. (2005): Der nationale Naturwissenschaftstest PISA 2003. Entwicklung und empirische Überprüfung eines zweidimensionalen Facettendesigns. In: Empirische Pädagogik, 19. Jg., S. 1-24.

TERHART, E. (2001): *Lehrerberuf und Lehrerbildung: Forschungsbefunde, Problemanalysen, Reformkonzepte.* – Weinheim.

WHITE, B. Y./FREDERIKSEN, J. R. (1998): Inquiry, modeling, and metacognition: Making science accessible to all students. In: *Cognition and Instruction*, Vol. 16,1, pp. 3-118.

Anschriften der Verfasser und der Verfasserin: Prof. Dr. Manfred Prenzel, Prof. Dr. Claus H. Carstensen, Dr. Martin Senkbeil, Dr. Christian Ostermeier und Prof. Dr. Tina Seidel, Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN) an der Universität Kiel, Olshausenstraße 62, D-24098 Kiel; e-mail: carstensen@ipn.uni-kiel.de; prenz@ipn.uni-kiel.de; senkbeil@ipn.uni-kiel.de; ostermeier@ipn.uni-kiel.de; cmseidel@standford.edu