



EDUCATION

- THE ARTS
- CHILD POLICY
- CIVIL JUSTICE
- EDUCATION
- ENERGY AND ENVIRONMENT
- HEALTH AND HEALTH CARE
- INTERNATIONAL AFFAIRS
- NATIONAL SECURITY
- POPULATION AND AGING
- PUBLIC SAFETY
- SCIENCE AND TECHNOLOGY
- SUBSTANCE ABUSE
- TERRORISM AND HOMELAND SECURITY
- TRANSPORTATION AND INFRASTRUCTURE
- WORKFORCE AND WORKPLACE

This PDF document was made available from www.rand.org as a public service of the RAND Corporation.

[Jump down to document](#) ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

Support RAND

[Browse Books & Publications](#)

[Make a charitable contribution](#)

For More Information

Visit RAND at www.rand.org

Explore [RAND Education](#)

View [document details](#)

Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use.

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL REPORT

Breaking Ground

Analysis of the Assessment System and Impact of Mexico's Teacher Incentive Program "Carrera Magisterial"

Lucrecia Santibáñez, José Felipe Martínez,
Ashlesha Datar, Patrick J. McEwan,
Claude Messan Setodji, Ricardo Basurto-Dávila

Prepared for the Mexican Ministry of Education



RAND EDUCATION

The research in this report was conducted by RAND Education, a unit of the RAND Corporation. This research was sponsored by the Mexican Ministry of Education (*Secretaría de Educación Pública*).

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

RAND® is a registered trademark.

© Copyright 2007 *Secretaría de Educación Pública*

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the *Secretaría de Educación Pública*.

Published 2007 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: <http://www.rand.org/>
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

SUMMARY

During the early 1990s, Mexico implemented a national education reform known as the National Agreement for the Modernization of Basic Education.¹ The federal government, state governments, and the national teachers' union (*Sindicato Nacional de Trabajadores de la Educación*, or SNTE) signed the agreement in May 1992. As part of this agreement, Mexico implemented Carrera Magisterial (CM), a program intended to give recognition to teachers and to provide economic incentives for superior teaching performance.

By using salary bonuses as an incentive tool, CM seeks to "help improve educational quality by rewarding and stimulating the work of the best teachers ... and reinforcing teacher interest in professional development and continuous improvement" (*Comisión Nacional SEP-SNTE*, 1998).² CM is actually more akin to a salary adjustment or horizontal promotion system; i.e., it adjusts teachers' salary levels without changing their job descriptions. CM offers salary incentives to teachers who participate in professional development courses and who consent to be evaluated through teacher and student tests, as well as peer reviews. Most teachers in public primary and secondary schools are eligible to participate. Indeed, the majority of Mexican teachers participate in CM.

To allocate the salary bonuses, CM conducts annual evaluations in which teachers voluntarily participate. The CM evaluation system focuses on six dimensions, or "factors," as they are called in the program: highest degree earned, years of seniority, a professional performance rating by a committee composed of the teacher's peers, the teacher's score on a test following federal and state professional development courses, the teacher's score on a teacher knowledge test, and a score reflecting the teacher's classroom average on a standardized student achievement test.

After the evaluation concludes, each teacher is awarded a total point score. If this score is above a specified cutoff, the teacher is included in CM at one of five levels (A-E), receiving a salary bonus associated with that

¹ *Acuerdo Nacional para la Modernización de la Educación Básica* (ANMEB). Note that in Mexico, basic education includes three grades of preschool (declared mandatory in 2002; will be implemented gradually between 2004 and 2008), primary school (grades 1-6), and secondary school (grades 7-9).

² Authors' translation.

level. Teachers who enter the first level (A) receive a salary bonus of about 20 percent of their base salary (determined by the traditional salary schedule using seniority and education). Most of the teachers in CM are at this level. Teachers in the program's highest level (E) receive more than 200 percent over their salary base.

By implementing Carrera Magisterial, Mexico became one of the first countries in the world to link teacher salaries to performance in public schools. However, the political nature of the program—which is managed centrally by a joint commission of the Ministry of Education and the teachers' union (*Comisión Nacional SEP-SNTE*) – has resulted in ambiguities with respect to program objectives. The two main actors have not always agreed upon a unified vision. CM can be considered both a new salary schedule (*SNTE's* vision) and a pay-for-performance program (*Secretaría de Educación Pública's* vision).

THE PURPOSE OF THIS STUDY

Despite more than a decade of implementation at considerable cost to and effort of the Ministry of Education, CM has never been formally and independently evaluated.³ Therefore, it is unclear whether the program functions adequately. Our study is not a comprehensive evaluation of the program as public policy. Rather it is an evaluation of its internal functioning and its potential for affecting educational quality. More specifically, our study evaluates the adequacy of the instruments used by CM to measure teacher performance. We also evaluate the effects of the salary incentives offered by CM on some indicators of educational quality. To do this, we focus on the following broad research questions:

1. Are the instruments used by CM to measure teacher and student performance technically sound? Are the test development and administration procedures adequate?

³ This is not to say that CM has never been studied. For example, Schmelkes (2001) conducted an evaluation of the program using a sample of teachers working in marginal areas of the country. Ornelas (2002) summarized the program's main features and some of its effects. Other studies on the topic include García Manzano (2004), Tyler (1997), and Santizo (2002). None of these studies, however, have utilized the full dataset of the program to determine its impact at the national level nor have they specifically studied its evaluation system.

2. Are the instruments and procedures used to measure the peer-review factor adequate?
3. What is the relationship among the program's factors? Are they positively related to educational quality?
4. Is the program meeting its goal of helping to improve educational quality?

DATA AND METHODS

Data for this study come from four sources: the Carrera Magisterial program, the Ministry of Education's Evaluation Directorate, the National Program for Teacher Professional Development (PRONAP), and the National Statistical Institute (INEGI). We constructed a database with information on every teacher who participated in CM from 1998-2003, including scores on the six program factors and additional information such as teachers' gender, age, grade, and subject taught, and socioeconomic indicators of the school and region.

We employed various methodological approaches, mainly quantitative in nature. Whenever relevant, we complemented our analysis with targeted literature reviews to build on existing research and draw lessons from other countries.

To examine the reliability and technical soundness of CM's tests, we used standard psychometric indicators of internal consistency and item reliability. To examine the alignment of test content with the curriculum, we convened subject-matter expert consultants with knowledge of the education system in Mexico to evaluate subject-by-subject content frameworks. Last, to explore the levels of cognitive demand elicited by the teacher and student tests, we adapted frameworks from the literature that specified several cognitive categories, and the experts to make judgments on the cognitive demand of the items. To evaluate the impact of CM incentives on educational quality, we first performed an analysis of the relationships among the program's factors and explored the extent to which each factor was related to student test scores and to other indicators of educational quality, such as teacher test scores and peer review ratings. We then estimated two regression discontinuity models to understand (1) whether, for teachers attempting to enter CM Level A, the program's incentives have a positive effect on student test scores and (2) whether the salary incentives offered by CM have

subsequent positive effects on student test scores after successful admission into CM or promotion to one of CM's higher levels.

SCOPE OF THE ANALYSIS

Most of the empirical analyses in this study used data from 1998 through 2003. This time period covers five evaluation cycles, or *Etapas* as they are called in the program (*Etapas* 8 to 12). Before *Etapas* 8 (corresponding to the 1998-1999 school year), the program underwent considerable reforms.

We focused on primary and secondary classroom teachers, who in CM are referred to collectively as *primera vertiente*. The analysis included primary school teachers in grades one to six, as well as secondary school teachers teaching mathematics, Spanish, geography, history, and civics and ethics. We did not evaluate school administrators or academic support staff members, who are also eligible to participate in CM.

As described above, the CM evaluation system focuses on six factors: highest degree earned, years of seniority, a professional performance rating by a committee composed of the teacher's peers, the teacher's score on a test following federal and state professional development courses, the teacher's score on a teacher knowledge test, and a score reflecting the teacher's classroom average on a standardized student achievement test.

Our analysis of CM teacher assessment instruments was restricted to those used in grades 1 to 6 in primary school and the academic subjects listed above in secondary school. We did not evaluate the instruments used to test knowledge of the content provided in the various professional development courses, nor did we evaluate the content or procedures of the courses themselves. Our analysis of professional development was limited to studying the relationship between the scores obtained in this factor and other measures of teacher performance. In addition, our analyses of student test scores were restricted to teachers whose students are tested (grades 3 to 6 in primary schools and "academic" subjects—math, Spanish, geography, history, and civics and ethics—in secondary schools).

Our empirical strategy attempts to isolate the impact of CM on specific indicators of educational quality. The quantitative nature of the analysis does not allow us to identify the mechanisms or processes that might have caused any observed effects.

SUMMARY OF FINDINGS

Our analysis of the teacher tests found that, although some of the testing procedures are adequate, other features of the process are problematic. In general, the CM tests used to measure teacher knowledge have adequate or near adequate levels of internal consistency and reliability, represent broad coverage of the curriculum, and are well aligned with test guides and specifications given to teachers. However, the majority of the items in the subject-matter section of the tests, which measure teacher knowledge of his subject, demands only low-level cognitive responses. In addition, other test sections evaluating knowledge of pedagogy and Mexico's education system have low levels of internal consistency reliability.

Although the student tests also represent broad coverage of the curriculum, other characteristics are faulty. These tests have lower levels of internal consistency reliability than do the teacher tests. In some cases, these levels are lower than those that would be desirable under international standards (particularly in the case of the student tests at the secondary level). As was the case with the teacher tests, the subject-matter-specific tests administered to secondary students demand mostly low-level cognitive responses.

Our evaluation of CM test development and administration procedures also produced mixed results. We found that testing procedures were generally adequate. However, there are no procedures in place to check for bias or ensure confidentiality of results on either the teacher preparation or the student achievement test. Although the teachers' professional preparation/knowledge tests appear to be carefully developed and monitored, the student achievement tests appear to suffer from less attention in their development, which in turn results in tests of comparatively lower quality. Moreover, we found a general lack of adequate documentation about item development and revision, and about technical quality, including reliability, particularly for the student tests. The documentation that is available reveals technical shortcomings in the statistical procedures used to develop and analyze tests.

Our analysis of the peer-review instruments and procedures found limitations as well. The instruments and evaluation criteria are not tied to explicit teaching standards on which to base subjective judgments about teaching practice. Furthermore, CM has not conducted empirical evaluations of

the reliability of the peer review instrument or its ranking levels to ensure that they are based on meaningful definitions of quality and sound psychometric properties. In terms of the peer review procedures, we highlight concerns regarding purpose and process. Examples from the literature lead us to conclude that peer review is usually done with more formative purposes to help teachers improve practice. These examples also suggest that peer review is typically conducted by people outside the school or district to prevent conflicts of interest – protocol not followed within the CM program.

The results from our impact analysis of CM incentives on student test scores suggest that these incentives do not have any discernible effects on student test scores for primary school teachers and very modest positive effects on student test scores for secondary school teachers who are vying for admission into CM. In secondary schools, these effects are more evident after *Etapa 10*, when states adhered more closely to the minimum cutoff required for admission. Magnitudes of the effects for secondary school teachers range from 3 to 15 percent of a standard deviation, and they are only for teachers we classified as being within a group most likely to benefit from the incentive program (which was comprised of fewer than 20 percent of the total sample). After teachers were admitted into the program or received a promotion, we observed modest negative effects on their student test scores.

Because the composition of the sample varies from year to year, these results can be explained in part by attrition. They could also be the result of the CM teacher incentive structure. Perhaps, since salary bonuses are guaranteed for one's entire career, teachers might exert lower levels of effort after receiving a promotion.

These findings are consistent with the literature on teacher incentives, which offers mixed evidence on the effects of salary incentives offered to teachers to improve student achievement. Most of the programs we reviewed in this study showed only modest improvements in student test scores, and these improvements (when they were observed) were often short-lived. Moreover, there was evidence that teachers devoted extra time to test preparation and this extra attention might have been partly responsible for the positive effects observed in some programs.

POLICY RECOMMENDATIONS

The mere fact that the *Secretaría de Educación Pública* (SEP) manages to annually administer more than 45 different tests to millions of students and hundreds of thousands of teachers with a limited budget is laudable. Perhaps the lack of sufficient resources is at the root of some of the deficiencies we identified in CM's assessment system.

The main recommendations of this study target the shortcomings observed in the teacher and students tests used to measure teacher performance. Even the best incentive systems might not show improvements in educational quality if the instruments used to measure them are flawed. It should also be noted that this research uses the program's own measures of quality as outcomes of interest (student test scores, teacher test scores, and peer review ratings). To the extent that these measures do not accurately reflect educational quality, our results will suffer from the same shortcoming.

Given our findings that the effect on student test scores of the incentives offered by the program is insignificant or very weak and that these very weak effects become insignificant after the bonus has been allocated, policymakers might consider reviewing the main features of the CM evaluation system, the factors it includes, and how these are assessed. To ensure that instruments used by test-based accountability systems are adequate, it is important to design and implement them according to internationally accepted standards of quality. To assure quality, technical manuals for test development and application and reporting of test results should be developed and continuously updated. These manuals should also include guidelines on reviewing items for bias. Moreover, they should incorporate detailed guidelines for confidentiality of the results and for communication and reporting of information to teachers, schools, state authorities, researchers, and the general public.

The CM teacher tests could benefit from several improvements. Currently, these tests do not increase in difficulty as teachers seek promotion into higher levels even though these levels are associated with higher salary bonuses. These tests are also much narrower than those used in other programs around the world, primarily measuring knowledge of one's subject matter. Other material on the test is of questionable value; the testing of knowledge of the Mexican education system is not supported by any theory of determinants of teacher performance. Policymakers should consider developing tests that

measure teaching practice more precisely (using concrete measures of teaching competencies) as well as subject matter knowledge.

In addition to improving the teacher test, the peer review process should be further evaluated. International experience indicates that peer evaluation could be designed to give reliable and valid information about teacher performance while at the same time serving as an important formative feedback and improvement tool. The peer review factor could also be adjusted to avoid subjecting teachers to potential conflicts of interest by evaluating other teachers who work in the same building. A possible solution to the current conflict of interest situation is to invite teachers in the highest levels of CM to become mentors and peer reviewers. This last option, however, would depend on CM revising its assessment system so that teachers in the highest levels are those who have genuinely demonstrated greater teaching competencies.

A common element across the student and teacher tests and the peer evaluation instrument is the lack of standards that link the evaluation mechanisms to models of teaching or learning. We recommend developing performance standards to undergird all assessment measures used by CM. Standards should detail subject-matter content, abilities, and/or knowledge to be evaluated, accompanied by detailed performance criteria and indicators. They are invaluable tools in the development of testing instruments. For policymakers, setting standards represents an opportunity to clearly establish desired teacher characteristics and actions; for teachers, standards provide a guide or framework for improvement efforts; for evaluators, standards constitute basic specifications or terms of reference that serve to guide test development and testing procedures.

CM has placed a strong emphasis on professional development. But it appears that undergoing professional development at the national level is only very weakly related to improvements on teacher tests, student tests, and peer-review ratings. Because an evaluation of the professional development factor, its instruments, and its course content fell outside the scope of our work, such an evaluation warrants further consideration.

Other reforms of a more structural nature should also be considered. In general, incentive programs should avoid "double counting"; for example, if seniority and education are already determining the base salary, they should not also be used to determine the size of the bonus. Double counting is

particularly problematic when these factors have been found to not be strongly linked to the outcomes of interest, as is the case in CM.

The program could also be modified so that it uses more than one year of achievement data when measuring teacher performance – using value-added methods that provide accurate and valid measures of teacher contributions to achievement gains over time. Such longitudinal assessments allow for providing continuous incentives for improvement as well as the possibility of implementing penalties or providing extra support if performance falls below acceptable levels. Evidence from the economics literature suggests that these kinds of actions could reduce “noise” in the performance measure as well as reduce risk to the worker, improving the efficiency of the incentives. It is also important that incentive programs minimize potential gaming behaviors, such as teaching to the test.

Teacher evaluation programs of the size and scope of CM benefit from technical and content advisory boards to supervise key psychometric and statistical aspects of its evaluations, curriculum coverage, and compliance with the necessary documentation on security, fraud detection, and other essential aspects of the assessment process. Such boards would work to ensure the quality of the tests and testing procedures.

All public policy programs strive to use resources efficiently. This is particularly important when, as is the case in Mexico, resources are scarce. One way to improve efficiency in the CM program would be to target bonuses only to those teachers who have strong incentives to perform.

Or, given that a national curriculum is already in place in Mexico, educational authorities could focus resources on developing a unified national testing system that could be used by CM as well as by other programs needing to measure student achievement. A unified national testing program would require tremendous technical, logistical, and administrative efforts. However, unified national tests would allow for external validation of the results of CM in relationship to other school reform programs at the federal and state levels. A unified system would also foster better use of available resources and technical capabilities. It would be important that policymakers analyze, with the assistance from technical experts, the option of consolidating the resources spent on CM’s testing program as well as other testing programs currently administered by the Directorate of Evaluation of the Ministry of Education into one single testing program. Recent efforts by

the Mexican Ministry of Education and other relevant organizations (such as the National Education Evaluation Institute) to develop a national education evaluation policy may represent a key opportunity.