Determining the Efficacy of the California Reading Instruction Competence Assessment (RICA)

By Sheryl O'Sullivan & Ying Hong Jiang

Introduction

On December 31, 2000, on the eve of the new millennium, the *Los Angeles Times* ran an article by Harvard professor Howard Gardner entitled "The Testing Obsession." In this article, Gardner accuses the nation in general, and California in particular, of taking part in "a frenzy of testing." Further, he states that few involved in this testing frenzy have ever asked the important question: "What is the relation between test scores and a quality education?" (Gardner, 2000, p. M1).

Sheryl O'Sullivan is an associate professor of teacher education and Ying Hong Jiang is an assistant professor of educational leadership, both with the School of Education at Azusa Pacific University, Azusa, California.

Those of us who are involved in the enterprise of education, especially those of us laboring in California, know first hand what Gardner means when he refers to current practice as a testing obsession. Elementary and secondary students are tested over an ever-widening array of discrete skills, and their harried teachers are themselves being increasingly tested using exams with acronym names like MSAT, PRAXIS, SSAT, BCLAD, and RICA. As with the

testing designed for elementary and secondary students, the increased testing for prospective teachers appears to have been mandated before asking whether the testing is related to learning or teaching.

With such policies increasing, educators may well wonder just what this new millennium will come to mean for us and for our students. Currently all of us from preschoolers on up are marching to the accountability tune. But without suitable reflection and research on the accountability of our testing programs, we are likely to continue to enforce policies that may or may not have any relationship to quality education. This article will first review some of the research surrounding current standardized testing practices in general, and will then turn to an examination of the efficacy of the RICA examination in particular. Finally, the results of a study that correlates RICA test scores with existing test data will be reported. The authors of this article seriously question our current testing trajectory, especially as it relates to RICA, and it is our intent to encourage others to question this path at least enough to desire solid research before moving blindly forward.

Our Current Testing Environment

The bandwagon against our current testing policies is a very crowded bandwagon indeed. And it is crowded with educators from very diverse backgrounds. Organizations as different as the International Reading Association, the American Educational Research Association, and the California Council on Teacher Education (*CCNews*, 2001) among others have issued resolutions which support authentic assessment but firmly oppose all high-stakes standardized examinations like the currently used California SAT-9. Highly respected educators as diverse as Alfie Kohn, Elliot Eisner, and the previously mentioned Howard Gardner have published articles and books decrying the increasing use of high-stakes tests at all levels of schooling. Perhaps with this many well-respected and experienced voices united against our current policy we ought to ask ourselves if their concerns have merit. What are the specific concerns that these and other educators cite involving standardized testing?

One recurrent concern is that standardized testing has the effect of narrowing the curriculum. This happens in a variety of ways. One noticeable way is that the content of the standardized test dictates the content of the curriculum. Kohn (2001) calls this cannibalizing the curriculum. If reading skills are heavily featured on the test, reading skills will be heavily featured in the classroom. Conversely, anything not on the test, such as science, history, music, expository writing, or literature analysis, will be given short shrift in the classroom. This might be defensible if we were certain that discrete, easily-measured skills were the important ones for students to know. However, as Berliner and Biddle (1997) point out in their book, *The Manufactured Crisis*, we now have many years of psychological research that indicates that learning is more complex, personally constructed, and integrated than we previously thought. Discrete facts are of limited value and do not transfer

well to other areas without guidance from a teacher. Standardized tests are unlikely to measure this type of learning adequately, and therefore the use of these tests as sole measures of achievement will rather quickly result in the narrowing of what teachers teach as well as what students learn.

Another concern about over-reliance on standardized testing focuses less on what teachers teach and more on how they teach. Kohn (2001) cites studies that find a positive correlation between high scores on standardized exams and shallow thinking in students. This is predictable when we realize the enormous range of subjects covered on most standardized tests. For example, Ohanian (2001) relates that questions on the California SAT-9 for sixth graders might include items on "the requirements for a police search of one's home, the climate in Cairo, and why the Republican party was formed in 1854" (p. 364). As long as tests place a priority on discrete facts of unintegrated knowledge, teachers feel compelled, as Berliner and Biddle (1997) say, "to cover course content at a gallop" (p. 318). And since there is a very limited amount of time that teachers have with their students we must ask ourselves what teachers are failing to teach while they are relaying the information about police search that students need to achieve a higher score on the test. Eisner (2001) puts this concern succinctly when he charges that an over-emphasis on extrinsically motivated, quantitative exams will undermine the curiosity, risk-taking, and exploration necessary to encourage intellectual life.

Perhaps the most worrisome concern, though, about our present-day use of standardized testing is the misuse of these tests away from their intended purpose. This happens in several ways. First, these tests are being used as the *sole* indicators of the worth of a school, a teacher, or a student. It would be very difficult to find anyone in the field of education or psychology who would advocate using the results of one test, especially a norm-referenced test, as the basis for making life-altering decisions about anyone. Yet this is how many standardized test programs are now functioning. Test scores alone are being used in various places to decide promotion and retention issues for students, merit raises for teachers, and job prospects for principals. When the results of one test carry so much weight, these tests are termed high-stakes, and as Gardner (2000) notes, "once a high-stakes test-measuring instrument has been revealed, the minds of everyone—students, teachers, parents, and the media—are wonderfully concentrated" (p. M1). Though the use of high-stakes testing is routinely condemned by researchers and practitioners alike, our minds are certainly concentrated on them while they wield so much power.

In addition to being misused as a sole indicator of merit, standardized tests often are also being used to judge attributes they were never designed to measure. Thompson (2001), for instance, notes that standardized tests designed to compare schools nationally are being misused to evaluate the curriculum or teachers of a specific school. And because the base of a standardized test is a standard, we are being ruled by a one-size-fits-nobody curriculum. When we pretend that all children are alike and that their needs are identical, we actually meet the needs of fewer and

fewer students. Some researchers believe that poorer school districts and poorer children are actually bearing a disproportionate amount of the burden of the current emphasis on standardized testing. Berliner and Biddle (1997) theorize that minority children are often subjected to the most drill and practice type teaching because their teachers are under the most pressure to increase test scores. Jonathan Kozol, in a recent address to the American Association of School Administrators, is even more strident in his criticism when he calls 'shameful' the practice of denying poor children equal preschool and school opportunities, but then testing them profusely in order to label them inadequate (Kozol, 2001). Clearly to these authors, and others, standardized testing goes far beyond theoretical issues to matters of equity.

While there are many other concerns about our current testing practices, these may suffice to encourage us to at least question the wisdom of high-stakes standardized tests. Most of the authors previously mentioned have been addressing the issue of such testing for children. It is the intent of this article to extend the concern about this type of testing to its use with adults, especially adults preparing to be teachers.

As stated at the beginning of this paper, elementary and secondary California students are not the only ones being subjected to standardized, high-stakes testing. Teacher candidates face an ever-increasing tide of alphabet soup tests in their quest to become credentialed. First, candidates must prove they thoroughly know any subject they may be called upon to teach by passing the MSAT (for elementary teachers) or the SSAT (for secondary teachers). For elementary teachers who teach multiple subjects, this means they must prove in-depth, specific knowledge of just about everything. Since the first-time passage rates for these tests are ridiculously low, many adults who would be strong teachers and who know how to research the many topics they may teach are kept out of credentialing programs, although not out of elementary and secondary classrooms. Some of these candidates spend hundreds of dollars on test preparation courses and additional testing before finally being able to enter a credentialing program. Many of these candidates teach on emergency permits with little or no training during this entire time.

Second, California is becoming increasingly interested in testing not only subject matter knowledge of teacher candidates, but also their pedagogical knowledge and skills. While most experts would agree that pedagogical skills are difficult to assess using objective, standardized tests, this is exactly what is taking place. The Reading Instruction Competence Assessment (RICA), required of all multiple subject and special education credential candidates, has been in place since 1998. Rumors of similar tests for mathematics (MICA) and science (SICA) have failed to materialize, but with the recent passage of SB 2042 vast new teaching performance expectations (TPE) will be tested using a teaching performance assessment (TPA) that has yet to be developed.

A recent report from the National Research Council (Mitchell, et al.2001) was critical of using standardized tests to license teachers because the Council did not

find that these types of tests adequately revealed what teachers understand, nor were they able to predict success in the classroom. Despite persistent and growing criticism of these types of testing programs for perspective teachers, however, legislatures continue to mandate these actions without adequate research about their effectiveness. Let us now turn our attention to one small piece of California's testing program, the RICA examination for prospective elementary and special education teachers.

Background of the RICA

The Reading Instruction Competence Assessment, or RICA™, was developed by National Evaluations Systems, Inc. (NES), at the request of the California Commission for Teacher Credentialing (CCTC). This was undertaken in response to legislation passed in 1996 in California as a part of the California Reading Initiative. This reading initiative is a broad political plan to improve the reading performance of California students, and one provision of this plan included instructions to the CCTC to "develop, adopt, and administer a reading instruction competence assessment" (NES, 1997, p.1). Passage of this exam is required of all candidates for multiple subject or special education credentials, but is not required for people teaching with internship, emergency, or single subject credentials.

The RICA can be taken in either a written or a video format. The vast majority of candidates take the written assessment, and this study examines only this format. The RICA Written Examination is organized around four domains that are considered important to the teaching of reading. These domains are:

Domain 1: Planning and Organizing Instruction Based on Ongoing Assessment
Domain II: Developing Phonological and Other Linguistic Processes Related to Reading
Domain III: Developing Reading Comprehension and Promoting Independent Reading
Domain IV: Supporting Reading Through Oral and Written Language Development

There are two sections to the written test, a multiple choice section of 60 scoreable items and a constructed response section which includes four short answer instructional tasks and one longer case study. Each section of the exam addresses all four domains. A score of 81 is needed to pass the exam (Carlson, 1998).

The RICA was first administered in June of 1998 at a cost of $178 per examinee. The statewide pass rate for the first year was 91.3 percent. In recent years the cost of the test has been reduced to $122 per applicant and the statewide pass rate has leveled off to 84.8 percent.

The stated goal of the RICA is ". . . to measure an individual's knowledge, skill and ability relative to effective reading instruction" (NES, 1997, p.1). This is a worthy goal, and with a well-respected test-developer like NES we can expect that there was due rigor in the development and validation of this test. This turns out to be true. In

fact the lengths that NES went to to produce a strong test are commendable. The test manual is over 100 pages long, explaining in exhaustive detail the test development process (NES, 1997). Prior to even beginning this process, another booklet, this one of 56 pages, explains efforts to find another measure of reading instruction competency that would alleviate the need for designing a new test (Zack, 1997). None existed. Finally, a 165-page document details the many steps taken by CCTC to establish a passing score on the exam (Carlson, 1998). Obviously due diligence was taken in the development of the RICA, and it is not the intent of this article to question whether the design of the test is well-researched. Rather, it is the intent of this article to question the entire endeavor. In other words, is the RICA a good idea?

Advisability of the RICA

Unfortunately the foundational question of whether the RICA is a good idea appears never to have been asked, or at least has never been the subject of research reported in the literature. Efforts to review the research pertaining to the RICA yield nothing about the topic beyond the previously reported documents on test development and validation. In our haste in California to comply with legislation, it appears we may have confused movement with progress. A great deal of activity has taken place to produce a test that may or may not have any impact on the quality of reading instruction in California classrooms.

Since there is so little in the literature relating directly to the RICA, one way to consider the question of advisability is to include the RICA in a larger category in which it fits, and apply what we know of the general category to a particular test. This is the main reason so much effort was taken at the beginning of this article to consider the question of our current program of standardized testing in general. As a high-stakes, standardized, legislated, written, objective test, could the RICA have the same effects as other examinations that fit into this category? Let's examine the three concerns listed at the beginning of this paper with the RICA in mind.

A major concern of tests like the RICA is that they narrow the curriculum, and encourage shallow learning. Without real research on the subject it is impossible to know for sure, but it appears the RICA could be having this effect. Half of the RICA is made up of multiple-choice questions. Fully 60 of the 81 points needed to pass could be attained through correct answers on multiple-choice questions. Since multiple-choice questions have only a single correct answer, they are testing low-level, literal knowledge. Moreover, some of the essay questions require students to cover the answer in as few as 50 words. Very little of any depth can be covered in such a constricted way. Since college reading professors are like teachers everywhere and want their students to do well on exams, it is reasonable to expect that more and more time in already compacted reading courses will be devoted to literal level, buzzword learning with time-consuming application and reflective activities being abbreviated to make time.

Another concern of our present testing program is that the tests themselves are very high-stakes. This is also true of the RICA. The RICA must be passed to gain a teaching credential in California. It does not matter if candidates have already followed a state-approved college major or passed the MSAT to demonstrate competence in their subject. It does not matter if they have completed a state-approved certification program that included a state-approved course in the teaching of reading. It also does not matter if candidates are already employed as teachers and everyone supervising them is pleased with their performance with children. Numerous highly-trained educators can work closely with candidates and attest to their fitness to teach reading, but if candidates do not pass this one written test they cannot be credentialed. Stakes this high, of course, invite abuses. Additionally, now the State of California, in response to federal directives, is using this test designed for testing the teaching of reading to rank entire university education departments, a purpose for which the RICA was never intended. As Gardner (2000) stated, a high-stakes test brings out incredible focus in all constituents involved. As it is in the case of other tests used as sole indicators of merit, it would appear the RICA has had the same ill effects.

Finally, in general standardized testing there is always a concern about equity issues. Especially in high-stakes testing when a person's chosen future depends on the outcome, we must always question whether some people are being hurt more than others. For the RICA it is not possible to determine from the score data reported to universities whether gender or ethnicity makes any difference in pass rates. However, according to the National Research Council (Mitchell, et al. 2001) minority students tend to have lower scores on standardized tests in general. This brings into question whether the diversity of classroom teachers is being narrowed through the use of the RICA.

One other area of concern about equity is virtually built into the regulations concerning the RICA. Teachers who are on internship or emergency credentials do not need to take the examination until they apply for a regular credential. This may seem like a small worry until we notice the huge number of people teaching on emergency permits in California, and the high concentration of these people in our most needy schools. The current policies which encourage students to begin teaching before they engage in any formal training are especially hard on returning adult students. These older students already have families to support, and the temptation to delay training as long as possible, especially if high-stakes testing is involved, means that many untrained people teach for years before seeking further education. If non-passage of the RICA had been shown to be useful in screening out poor teachers of reading, this delay in putting fully credentialed teachers in our classrooms would be defendable, in fact welcome. However, without any sort of proof of this, the RICA becomes just one more barrier to credentialing, a barrier felt more keenly by mature returning students who wish to teach.

As previously stated, there is no research recorded in the literature that

considers any of these concerns about the RICA. While it is logical to consider that a specific test may have the same effects as tests in a general category to which it belongs, this has not been addressed through research. As a beginning in building a research base concerning the wisdom of the RICA it is reasonable to ask the question: Does the RICA provide us with any new information about prospective teachers? In other words, what is the discriminant validity of the RICA? The research study here reported sought to answer that question.

Method

Design of the Study

During the 1998-99 and 1999-2000 academic years data were collected on students who were enrolled in one California university's teacher preparation program. Data collected included scores on RICA, California Basic Educational Skills Test (CBEST), and Multiple Subjects Assessment for Teachers (MSAT), cumulative grade point average (GPA) at time of admission, and grades attained in the reading methods course required of all students. All of these categories of data were compared individually and in groups with RICA scores using logistic regression to test for relationships among the variables.

Sample and Targeted Population

All students in the study were enrolled in the state-approved teacher education program of a small independent liberal arts university in California. All students were pursuing a multiple subject credential. Some students had followed a liberal studies major and were undertaking teacher education as part of an undergraduate program. Others had completed undergraduate degrees in a variety of majors and were enrolled in a fifth-year program leading to the preliminary credential. The studied sample was comparable in terms of gender and ethnicity to the statewide population taking the RICA. The initial sample numbered 171, but since not all data were available for every student only 106 cases were included in the final analysis.

Data Analysis and Results

Since all students had passed CBEST before entering the program, this variable was not useful and was, therefore, discarded. Each of the other predictor variables (MSAT scores, incoming G.P.A. and grades in reading) was correlated individually with RICA pass/fail scores. Table 1 shows the correlations among the variables.

The same predictor variables were then correlated in different combinations with RICA scores. Logistic regression was used to test the strength of each correlation. This test was chosen because the outcome variable RICA scores were only available in dichotomous form, pass or fail. A request to NES to supply continuous data for the subjects in this study was denied, which was consistent with the experience of the National Research Council (Mitchell, et al., 2001) that found

it was unable to obtain enough information about this company's tests to study them well. Logistic regression is useful for situations in which you want to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous. Logistic regression was therefore chosen as the appropriate statistical test, and some of the continuous data of the other variables were changed to dichotomies to serve the purpose of the study. For example, the MSAT scores were transformed to pass/fail. In the logistic regression analysis, we first build a model with three predictors. These predictors are MSAT pass or fail, raw GPA, and reading grades. These variables are used to predict the probability of passing RICA. Table 2 shows the results of our initial logistic regression model.

From the results of our initial logistic regression model, we found that MSAT and GPA are two significant predictors of probability of passing RICA, with p value less than .05. Reading grade is not a significant predictor in this model. A forward

Table 1
Zero Order Correlations among MSAT Raw Score,
GPA, Reading Grade, and Rica Pass or Fail

Variables	MSAT Raw Score	GPA	Reading Grade	RICA Pass or Fail
MSAT Raw Score	-	.121	.605	.344**
GPA		-	.243*	.235*
Reading Grade			-	.080
RICA Pass or Fail				-

Notes. $\underline{n}$=106. *$\underline{p}$ < .05. **$\underline{p}$ <.01.

Table 2
Summary Statistics for the Logistics Regression Models

Models	Variables	B	S.E.	Wald Statistic	p	R
Initial Model Using Three Predictors	MSAT Pass or Fail	1.6277	.6620	6.0453	.0139	.2082
	GPA	1.3560	.6892	3.8716	.0491	.1416
	Reading Grades	.2990	.7567	.1562	.6927	.0000
Final Model Using Two Predictors	MSAT Pass or Fail	1.6080	.6580	5.9729	.0145	.2063
	GPA	1.4117	.6721	4.4115	.0357	.1607

Wald test then was chosen to determine the final model including only the important predictors through the Wald test. Table 2 compares the results from our initial model with three predictors and the final model with two predictors.

The final logistic regression model is a more parsimonious model in that it only includes two significant predictors of probability of passing RICA. This model suggests that RICA could be predicted with 95 percent of confidence when a participant passed MSAT and holds a fairly high GPA.

Our next analysis involves determining the GPA cut-off value to accurately predict probability of passing RICA. Table 3 shows the percentage of passing RICA for each combined category of MSAT and GPA.

From the table, we can see that 30 out of 34 participants who passed MSAT and had a GPA between 3.0 and 3.49 also passed RICA, in other words, the passing rate for these participants are 88.2 percent. We can also see that 24 out of 24 students who passed MSAT and had a GPA of 3.5 or above also passed RICA, in other words, the passing rate for these participants is 100 percent.

Discussion

This study contained several limitations. The sample size was relatively small and taken from only one university. Continuous data were not available for the RICA, and the analysis of the MSAT data was based upon a transformed rather than the original variable. However, within these limitations, these findings lend

Table 3: Percentage of Passing RICA for Each Combined Category of MSAT and GPA

MSAT				RICA Fail	RICA Pass	Total
Fail	GPA	2.0-2.49	Count	0	1	1
			% within GPA Category	0.00%	100.0%	100.0%
		2.5-2.99	Count	4	3	7
			% within GPA Category	57.1%	42.9%	100.0%
		3.0-3.49	Count	1	1	2
			% within GPA Category	50.0%	50.0%	100.0%
		3.5-4.0	Count	1	3	4
			% within GPA Category	25.0%	75.0%	100.0%
	Total		Count	6	8	14
			% within GPA Category	42.9%	57.1%	100.0%
Pass	GPA	2.0-2.49	Count	1	4	5
			% within GPA Category	20.0%	80.0%	100.0%
		2.5-2.99	Count	6	23	29
			% within GPA Category	20.7%	79.3%	100.0%
		3.0-3.49	Count	4	30	34
			% within GPA Category	11.8%	88.2%	100.0%
		3.5-4.0	Count	0	24	24
			% within GPA Category	0.00%	100.0%	100.0%
	Total		Count	11	81	92
			% within GPA Category	12.0%	88.0%	100.0%

credence to the hypothesis that the RICA may be redundant. It is unlikely that a person who has passed MSAT will fail the RICA on the first attempt. And it is virtually assured that such a person will pass the RICA on the second attempt. We can therefore predict that the population who would have entered the teaching force after passing the MSAT will enter that force whether or not a RICA exam exists. It may take longer and cost several hundred dollars more for each candidate to become a teacher, but virtually the same people will become teachers with or without the RICA.

We can also notice this redundancy when we look at the combined predictive value of a passing score on the MSAT and an incoming cumulative GPA of 3.0 or above. These two variables together can predict that the candidate is likely to pass the RICA, and if the GPA is above 3.5, passage of RICA is virtually assured. Even just having information that a student's GPA is above 3.0 will correlate to a passing RICA score at a .05 level of significance. Again, according to this study, students who have gone through the fairly routine university screen for GPA and have also passed the MSAT will be virtually the same students who become teachers whether or not the RICA is part of their requirements.

Conclusion

California's, and indeed the nation's, over-hasty leap into the game of high-stakes testing is like diving head first into a pool without first checking the depth. It may be perfectly safe, but then again it may not be. We are risking the futures of a great many teachers and students with such a precipitous dive made with so little study. It seems unwise to continue in this direction at least in the case of the RICA without more research into whether there is any correlation with the quality of education being provided to our children.

With that in mind, here are some suggestions for research questions that should be addressed:

1. Does the RICA have any effect on the coursework being offered to pre-service teachers? In other words, how different are the reading methods courses now from the ones offered before the RICA was required, and have those changes been positive or negative when measured against what we know of good teaching?

2. Does the RICA have any effect on the behaviors of teachers in elementary classrooms? In other words, are teachers doing a better job of reading instruction since the advent of the RICA? Do the students of RICA teachers read better?

3. Can we discern which teachers are RICA tested and which ones are not? In other words do the teachers who are teaching without taking or preparing for the RICA teach reading differently in comparison with those teachers who have taken the RICA?

4. What measures do students take if they do not pass the RICA the first time which leads them to pass on subsequent attempts? Do these measures lead to improved reading instruction in the classroom?

These suggestions are very basic questions which should have been asked long before we undertook such an expensive and time-consuming testing program. It may well be that preparing for and passing the RICA can be shown to have measurable positive effects on the quality of reading instruction in our elementary classrooms. Improved instructional quality would presumably lead to better readers. This, of course, is the goal of all reading teachers everywhere, and if the RICA plays a part in this, by all means we should embrace it. But until at least some evidence can be found that links the RICA with improved instruction, we should stand solidly against any head first dives into pools of unknown depth.

References

Berliner, D.C. & Biddle, B.J. (1997). *The manufactured crisis.* White Plains, NY: Longman.

Carlson, R. Jr. (1998). *Establishing passing standards on the Reading Instruction Competence Assessment.* Sacramento, CA: California Commission on Teacher Credentialing.

CCET, (2001). CCET joins with other educational groups to oppose high stakes standardized tests. *CCNews, 9*(2), 3.

Eisner, E.W. (2001). What does it mean to say a school is doing well? *Phi Delta Kappan, 82*(5), 367-372.

Gardner, H. (2000, December 31). The testing obsession. *The Los Angeles Times,* pp. M1, M6.

Kohn, A. (2001). Fighting the tests: A practical guide to rescuing our schools. *Phi Delta Kappan, 82*(5), 348-357.

Kozol, J. (Speaker). (2001). National conference on education keynote address (Cassette Recording No. AASA01-01). Orlando, FL: American Association of School Administrators.

Mitchell, K.J., D.Z. Robinson, B.S. Plake, & K.T. Knowles (Eds.) (2001). *Testing teaching candidates: The role of licensure tests in improving teacher quality.* Washington, D.C.: National Research Council, National Academy Press.

National Evaluation Systems, Inc. (1997). *Development and validation of the content specifications for the Reading Instruction Competence Assessment.* Amherst, MA: Author.

Ohanian, S. (2001). News from the test resistance trail. *Phi Delta Kappan, 82*(5), 363-366.

Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan, 82*(5), 358-362.

Zack, J. (1997). *Search for and analysis of extant measures of a teacher's reading instruction competence.* Sacramento, CA: California Commission on Teacher Credentialing.