

**Consistency of Standard Setting  
in an Augmented State Testing System**

Robert W. Lissitz

Hua Wei

University of Maryland

Paper presented at the annual meeting of the National Council on Measurement in  
Education, April 10, 2007, Chicago.

## Introduction

The No Child Left Behind (NCLB) Act (2001) requires that all schools and districts be reviewed annually on the basis of the percentage of students who perform at or above the state-defined proficient level and improvement from the previous year in terms of this percentage. Performance standards and cut scores are set for the statewide assessments and students are classified into various achievement categories based on their scores on the exams. Students identified as below proficient are provided with additional assistance or enrolled in remediation programs that are designed to help them reach the required level. Schools that fail to show adequate yearly progress (AYP) for consecutive years will receive serious sanctions. Therefore, the determination of cut scores that leads to student classifications has important implications for school accountability and instructional practice.

Traditionally, standard setting is a grade by grade activity. This practice often leads to the result that the performance standards are not set with a consistent level of rigor across grades (Ensigh et al., 2002). Besides, the variation among the resulting cut scores across grades may be too big to be considered reasonable. Inconsistency among the cut scores across grades is often reconciled by a single articulation committee, whose job is to evaluate the grade-by-grade committee recommendations (Buckendahl et al., 2005). However, the results are hard to justify due to a lack of guiding policy and accepted methodology.

The NCLB Act requires an effort to align the cut scores across grades and construct a coherent and consistent system of achievement standards. As defined by Lewis and Haug (2005), “Consistent proficient cut scores across grades within a content area should

represent the knowledge, skill, and abilities required in each grade that, within reasonable growth, would allow the student to meet the expectations represented by the cut score at the next grade assessed” (pp. 14-15). Consistency of standards or cut scores is conventionally examined by looking at the resulting proportions of examinees at the different performance levels. Kane (2001) stated that “To the extent that the proportions assigned to each performance level are the same for the two assessments, the standards on the two assessments are comparable in their stringency. To the extent they are different, the standards are different” (p. 76).

Lissitz and Huynh (2003) suggested an approach to consistency of standard setting that was based on a judgmental process by the standard setting committees. Huynh and Schneider (2005) suggested imposing “a consistent trend line” on the percentages of students assigned to the proficient category across grades, and by “consistent” they meant “no change, a moderate level of increase, and a moderate level of decrease” (p. 106). The four models of cut score consistency proposed by Lewis and Haug (2005) reflected the same concept and the four models result in an equal percentage, an approximately equal percentage, a smoothly decreasing percentage, or a smoothly increasing percentage of students categorized as proficient across grades.

The call for the use of multiple measures for making decisions about students, as specified in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), gave rise to augmented state testing programs, which include a state-developed criterion-referenced test that meets the assessment requirement of the NCLB Act, and an external norm-referenced assessment

that generates nationally normed scores. The norm-referenced test augments the criterion-referenced test in the sense that it provides a base line to confirm or evaluate information given by the state test. It serves as an external benchmark against which the state test and its standards can be compared in terms of technical vigor and credibility.

In this paper, we address the issue of consistency in standard setting in the context of an augmented state testing program. Information gained from the external norm-referenced test scores is used to help make an informed decision upon the determination of cut scores on the state test. The research questions to be answered in this study include:

- 1) Is there a pattern among the cut scores across grades? Is there any evidence that the cut scores for categorizing students as proficient are equally (or not equally) demanding across grades?
- 2) What are the indicators of consistency among the cut scores across grades?
- 3) How can the scores or associated indices from the norm-referenced test be used to help obtain consistent standards on the state test?

### **Data**

The data examined in the study are Maryland cross-sectional student performance assessment data for Grades 3 through 8 for three consecutive years from 2003 through 2005. Two tests, a criterion-referenced test (CRT) and a norm-referenced test (NRT), were administered to each grade in the content areas of mathematics and reading. In 2003, mathematics and reading assessments were administered in Grades 3, 5, and 8. In 2004 and 2005, all the grades from Grade 3 through 8 were assessed on both subject areas. CTB/McGraw-Hill and Harcourt Assessment were responsible for the

mathematics and reading assessments, respectively. The NRT scores were computed using TerraNova or Stanford 10 items, and the CRT scores were calculated using the custom items written to the Maryland content standards (Maryland Voluntary State Curriculum) plus a subset of TerraNova or Stanford 10 items that were aligned with the state content standards.

Standard setting was conducted for Grades 3, 5, and 8 in 2003, and for Grades 4, 6, and 7 in 2004. The Bookmark method was applied in both years. The cut scores obtained from the two sessions of standard setting were used to assign students to three performance levels: basic, proficient, and advanced. Information about the standard setting procedure and the resulting cut scores can be found in the Maryland Standard Setting Technical Report (Maryland State Department of Education, 2003).

### **Methodology**

In the state testing programs with multiple measures, accuracy and appropriateness of the high-stakes decisions on performance standards and cut scores can be enhanced by taking advantage of the relationship among “different tests that are intended to measure the same general construct but for different purposes” (Baker, 2003, p. 14). As indicated above, the content overlap between the CRT and NRT in the Maryland School Assessment (MSA) system is substantial, and, presumably, results from the two tests should lead to similar conclusions about student achievement. As a matter of fact, the correlations between the NRT and CRT scores ranged from .80 to .85 in mathematics, and from .89 to .90 in reading (Maryland State Department of Education, 2004). Therefore, the NRT scores and other measures associated with the scores can be used as yardsticks by which the consistency of the CRT standards is evaluated. In this study, the

degree of consistency in standards is detected by translating the CRT cut score for each grade into a NRT national percentile rank and then comparing the percentile ranks across grades. If the percentile ranks are smoothly increasing, decreasing, or remain constant across grades, we have some evidence to say that the cut scores are consistent across grades. Following the same line of thought, a consistent system of cut scores can be established for consecutive grades if we set the cut scores by forcing their corresponding NRT national percentile ranks to be consistent. We can have the same fixed percentile rank for each grade or we can have the percentile ranks increasing or decreasing continuously for the consecutive grades to reflect state policies regarding standards.

The Maryland standard-setting committees had an interest in gradually increasing the expectation for achieving the minimal proficient level of performance across grades when they set the standards in 2003 (Personal communication with staff at Maryland State Department of Education). The committees, as far as we have been able to determine, did not use any systematic approach to achieve this outcome. The method proposed in this study serves the function of providing a mechanism to steer the committee decisions in the direction that they have declared an interest in going.

## **Results**

### *Distributions of performance categories across years*

An analysis of the CRT results yields basic statistics about the distribution of performance categories for each grade across the three years. Two sets of graphs, from Figure 1 through 12 as shown below, display the percentage of students in each performance category for each grade across years in the two content areas. A pattern that is observed in every graph is that the percent of students in the basic category was

decreasing across years as they, presumably, moved to the proficient category, and the percent of students in the higher two categories (proficient and advanced) was increasing. This finding is encouraging. It is our understanding that the CRT scores for each grade have been equated across years (i.e., horizontal equating), and in the Technical Report 2004 for Maryland School Assessment (Maryland State Department of Education, 2004), there was an explicit indication of the equating procedure. Therefore, assuming that the CRT scores for the same grade are comparable across years, this finding implies that new students perform better than their predecessors. This is exactly what a school system and the state would like to see. (Please note that the data in each of the figures are cross-sectional, but we have connected the dots with lines to make them easier to understand. The lines do not imply longitudinal data.)

-----  
Insert Figures 1-12 about here  
-----

*Comparison of cut scores in terms of NRT national percentile ranks*

The following graphs, from Figure 13 through 18, display the relationship between the CRT cut scores and the NRT national percentile ranks in each year across grades in reading and mathematics. Students who were at the boundaries of two performance levels, i.e., students whose CRT scaled scores were equal to the cut score were selected and the median of their NRT national percentile ranks was identified. For some particular year and grade combinations, cut scores did not actually occur among students. When this was the case, students who scored closest to the cut score were selected.

-----  
Insert Figures 13-18 about here  
-----

A pattern that is common in all the graphs is that the cut score that distinguished the basic from the proficient category was associated with a general increasing trend in terms of the NRT national percentile rank with increasing grades for each year. This is what was desired by the standard setting committees, according to MSDE staff. It is noticeable that the cut scores for Grades 3, 5, and 8 were dissimilar from those for Grades 4, 6, and 7, probably because they were set at two time points (in 2003 and 2004, respectively), and by different committees. The cut scores set in the same year were increasingly demanding as the grade level increased, although the trend was less obvious over the entire span of grades.

In contrast, the national percentile ranks corresponding to the cut scores that distinguished the advanced from the proficient category bounced around with no clear trend. Therefore, a conclusion can be drawn from this finding that the cut score for the proficient category was generally more demanding for higher grades than for lower grades. In other words, students had to become better, at least compared to the national norming group, in order to remain in the proficient category as they progressed through school. If they were not, they would drop down into the basic category in subsequent years.

*Comparison of cut scores between content areas*

The following graphs, from Figure 19 through 24, display the results of comparison of the relationship between the CRT cut scores and the NRT national percentile ranks in mathematics with that in reading.

-----  
Insert Figures 19-24 about here  
-----

Figures 19 through 21 suggest that the cut score separating the proficient category from the basic category is more demanding in mathematics than in reading, because the median national percentile rank associated with the cut score in mathematics is greater than that in reading. In the years of 2004 and 2005, the difference between the two cut scores is the smallest in Grades 4 and 5, and the largest in Grade 6.

The cut score that distinguishes the advanced category from the proficient category exhibits a somewhat different pattern. In Grades 3 and 4, the cut score in reading is more demanding than that in mathematics, but the opposite is true for Grades 5 through 8. The difference between the two cut scores is smaller in the lowest and the highest grades, and larger in the middle grades.

Whether this result was intentional or not, we do not know, but we suspect it was not planned. Instead, we suspect that the standard setting committees in mathematics were composed of teachers who were more demanding than those in reading, at least as far as achieving proficiency is concerned.

#### *Adjustment of cut scores*

Tables 1 and 2 show the original cut scores that were determined for Grades 3, 5, and 8 in 2003, and the median NPR associated with the students at or near each cut score.

Looking at the NPRs across grades in each year, we see evidence that the cut scores are increasingly demanding as the grade level increases, as we discussed above.

-----  
Insert Tables 1 and 2 about here  
-----

Different approaches could be adopted to adjust the CRT cut scores by fixing the NRT national percentile ranks associated with students at cut scores. In the first approach, we used the NPR associated with the original cut score for the proficient category in Grade 3 in 2003 as a baseline and applied it to the other two grades assessed in the same year. The same NPR was translated into different cut scores for the three grades, and we used the newly obtained cut scores to re-categorize students in 2004 and 2005. In this way, we allow differences in student performance in years 2004 and 2005 for all three grades to show increases or decreases relative to the fixed NPR of Grade 3 in 2003 as the baseline.

Tables 3 and 4 summarize the results of the adjustments in the two content areas. It is noticeable that the resulting median NPRs associated with students who performed at or near the cut scores for the three grades in 2004 and 2005 were almost the same in reading, but they bounced around in mathematics, especially in the year of 2005.

-----  
Insert Tables 3 and 4 about here  
-----

Alternatively, in our second approach, we used the average empirical NPR of Grades 3, 5, and 8 in 2003 as the baseline and translated it into different scale scores for the

three grades. Clearly, we could have used the median of the NPRs for the three grades and the associated scale score at each grade. In our example, it would not have made any difference. The same set of scale scores was used to re-categorize students in 2004 and 2005. Again, this allows differences in performance to emerge in 2004 and 2005 relative to the average of the performance fixed in 2004 as the baseline. Results of the adjustments are summarized in Tables 5 and 6. In the content area of reading, the resulting median NPRs bounced around in 2004, but remained almost constant in 2005. In mathematics, the resulting median NPRs showed more differences across grades in 2005 than in 2004.

-----  
 Insert Tables 5 and 6 about here  
 -----

The results of the adjustments are satisfactory in the subject area of reading. The median NPRs associated with students at or near the adjusted cut scores are similar across grades and across years. The fluctuation in the NPRs is within reasonable limits and can be partially accounted for by the error of estimation associated with the test. Consider the fact that for some grade and year combinations, the cut score was not obtained by any of the students. In that case, we chose a score closest to the cut score and found the median NPR associated with students having that score. Therefore, the median NPR we obtained was not exactly corresponding to the cut score, but to a score one or two points above or below that score.

The fluctuation of the NPRs is larger in mathematics, especially in the year of 2005. The adjusted cut scores are more demanding for lower grades than for higher grades,

because the associated NPRs go down consistently as the grade level increases. It reflects the fact that students categorized as proficient must maintain a higher ranking compared with the national norming group as they progress through school. This is consistent with the results we have presented before. The difference in this analysis is that we have adjusted the cut scores in the later years to be consistent relative to the performance expectation obtained at the base year of 2003. We have tried to eliminate or at least lessen the differences in performance that are due to changes in the standard.

### **Discussion**

How to examine and set consistent cut scores across grades and subject areas is the theme of this study. We have identified two approaches to setting consistent cut scores. One is to influence the judgments that the standard setting committees render using empirical data related to the performance levels. The other is to statistically adjust the results of independently working standard setting committees after they have completed their tasks and data have been obtained from the testing. In the case of augmented testing, the first approach might be accomplished by presenting the NRT percentile ranks at the cut point, as we illustrated above. To our knowledge this has never been attempted, but appears to be feasible. In the typical standard setting approach, the committee is presented with information on the percentage of students in the school system (usually their specific state) that would be declared proficient or below proficient based on the selection of the cut point. As we said, we do not know of any case where the cut point is directly compared to a national level of performance by looking at the national percentile rank (NPR) for students at each cut point. How this use of a national perspective would influence the work of the state committee and how to control that

influence is unknown at this time. We are very interested in exploring this issue and believe it to be a worthwhile direction in the study of standard setting.

The second approach of statistically adjusting standard setting to insure various options for consistency is also an interesting problem and worth further exploration. We have been trying to create a system to accomplish this. Different approaches could be adopted to adjust the CRT cut scores, as we have shown above. The method we are currently favoring is to use the NRT percentile ranks in an augmented testing environment to provide a mechanism to statistically adjust the standard setting decision in the direction that the state chooses. Basically, the following sequence, as illustrated above, could be utilized to accomplish this task:

1. Decide on policy regarding the type of consistency we want to achieve in terms of a normative standard (NRT national percentile rank), i.e., do we want the standards to be constant, smoothly increasing, or smoothly decreasing across grades? If we want higher standards for higher grades, what is the level of growth we expect from one grade to the next? If we want lower standards for higher grades, what is the level of decrease that we expect from one grade to the next? Do we want mathematics to have the same standard that is held for reading?
2. Meet with the standard setting committee to explain the policy that was determined in Step 1. Before they set the standards, the committee needs to know that we may adjust their results. One option is to use the standard for a specific grade (perhaps Grade 3) as a base year standard and have consistency defined relative to that grade. If this were done, the Grade 3 (or other) standard setting committee could be the only

committee that sets standards. Standards for the other grades would be the result of statistical analysis operationalizing a policy decision regarding consistency.

3a. If the policy from Step 1 says that the standards should be constant across grades, convert the standard from Step 2 to a scale score for each grade. For example, if the standard we set in Step 2 has a NPR of 25 and we fix it for each grade, what we need to do is:

- 1) For each grade, identify all the students whose percentile ranks are equal to 25 on the NRT.
- 2) Compute the mean/median CRT score for those students. That is the cut score for that particular grade associated with the chosen NPR.

It is very likely that the same NRT normative standard is associated with a different CRT scale score in each grade, unless considerable care was given to equating and standardizing the scales across grades with national data. In most cases this will not have been done, and the relationship between the NRT and CRT will vary to some extent from grade to grade.

3b. If the policy from Step 1 says that the standards should be smoothly increasing or decreasing, adjust the standard set in Step 2 for each of the other grades to reflect that policy.

For example, if the standard we set in Step 2 is an NPR of 25 and we want the standards to be increasing by an increment of 3 percentiles, then:

- 1) For Grade 3, identify all the students whose NPR is equal to 25 on the NRT.
- 2) Compute the mean/median CRT score for those students. That is the cut score for Grade 3.

- 3) For Grade 4, identify all the students whose NPR is equal to 28 on the NRT.
  - 4) Compute the mean/median CRT score for those students. That is the cut score for Grade 4. The same thing can be done for the other grades.
4. Apply the cut scores determined in Step 3 to other subject areas to achieve consistency across subject areas.

This approach requires several assumptions, of course, including the idea that the norming groups for different grades and for different subject areas are equivalent or adjusted to be so. We also assume that the selection of items for the CRT is representative of the items in the NRT, and the norming was done well enough to serve as a source for a standard. The most important assumptions have to do with the horizontal equating of the tests across years and subjects and the specification of the scale for each grade, year and subject matter. We have assumed that the tests are measuring comparable qualities across grades and doing so using a scale score system that is directly comparable. The robustness of the adjustment procedure to these assumptions needs to be investigated.

### **Conclusions**

In an augmented testing program, measures provided by the national norm-referenced test can be used to inform the standard setting for the state or local criterion-referenced test, assuming that the two tests are highly related. In this paper, a standard-setting design is outlined that uses the NRT results as impact data to influence the judgments of the standard setting committee. Statistically adjusting the cut scores by using information from the NRT is another approach to fostering consistency across grades, and it is illustrated with real data.

In a coherent educational assessment system, it can be argued that not only should the standards across grades of a content area be consistent, but the standards across content areas should also be consistent. By using the approaches discussed in this paper, consistency of standards across subjects can be addressed. Of course, the assumption needs to be satisfied that the content standards are of equivalent difficulty across subjects and that the educational resources are equally allocated, so that we can be confident that the approaches are justified

#### **Note**

1. A portion of this paper was presented at the International Conference on Educational Evaluation, November 11-12, 2006, National Taiwan Normal University, Taipei.
2. We would like to thank the Maryland State Department of Education for their financial support of this project and Mark Moody and Marty Kehe for their advice.

#### **References**

- Baker, E. L. (2003). Multiple measures: Toward tiered systems. *Educational Measurement: Issues and Practice*, 22, 13-17.
- Buckendahl, C. W., Huynh, H., Siskind, T., & Saunders, J. (2005). A case study of vertically moderated standard setting for a state science assessment program. *Applied Measurement in Education*, 18, 83-98.
- Ensign, G., MacQuarrie, D., & Beck, M. (2002, June). *Validation of state achievement performance standards*. Paper presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Palm Desert, CA.

- Huynh, H., & Schneider, C. (2005). Vertically moderated standards: Background, assumptions, and practices. *Applied Measurement in Education, 18*, 99-113.
- Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives*. London: Lawrence Erlbaum Associates.
- Lewis, D. M., & Haug, C. A. (2005). Aligning policy and methodology to achieve consistent across-grade performance standards. *Applied Measurement in Education, 18*, 11-34.
- Lissitz, R. W., & Huynh, H. (2003). Vertical equating for state assessments: Issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment Research & Evaluation, 8*(10). Retrieved from <http://pareonline.net/getvn.asp?v=8&n=10>.
- Maryland State Department of Education. (2003). Maryland Standard Setting Technical Report. Retrieved from <http://www.marylandpublicschools.org/MSDE/divisions/planningresultstest/Maryland+Standard+Setting+Technical+Reports.htm>.
- Maryland State Department of Education. (2004). Technical Report 2004 for Maryland School Assessment. Retrieved from <http://www.marylandpublicschools.org/MSDE/divisions/planningresultstest/2004+MSA+Technical+Reports.htm>.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425.
- Schafer, W. (2003). A state perspective on multiple measures in school accountability. *Educational Measurement: Issues and Practice, 22*, 27-31.

**Figures**

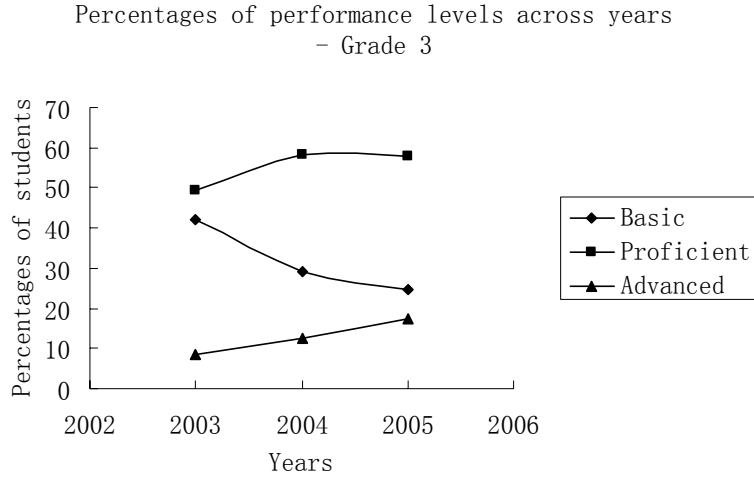


Figure 1: Distribution of performance categories across years for Grade 3 in Reading

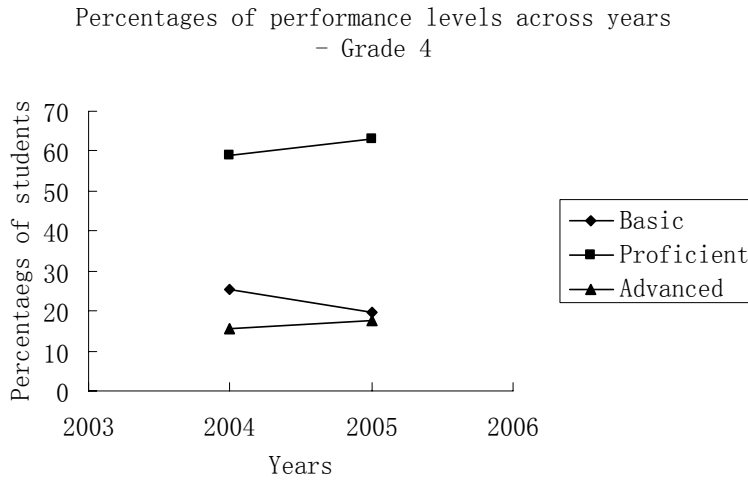


Figure 2: Distribution of performance categories across years for Grade 4 in Reading

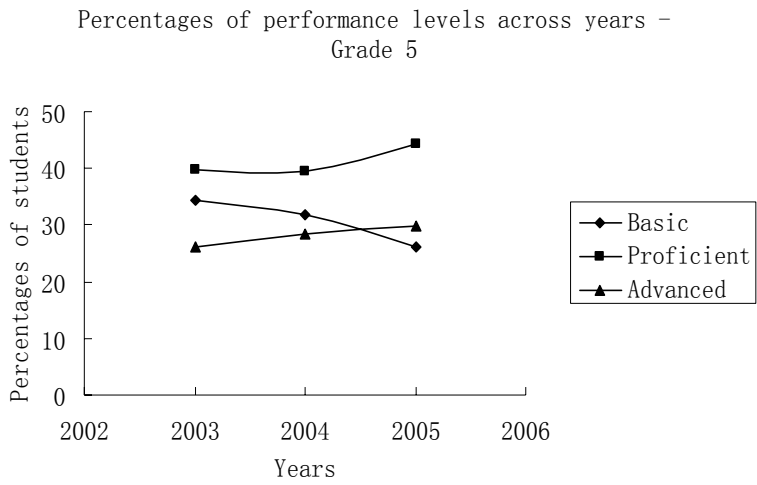


Figure 3: Distribution of performance categories across years for Grade 5 in Reading

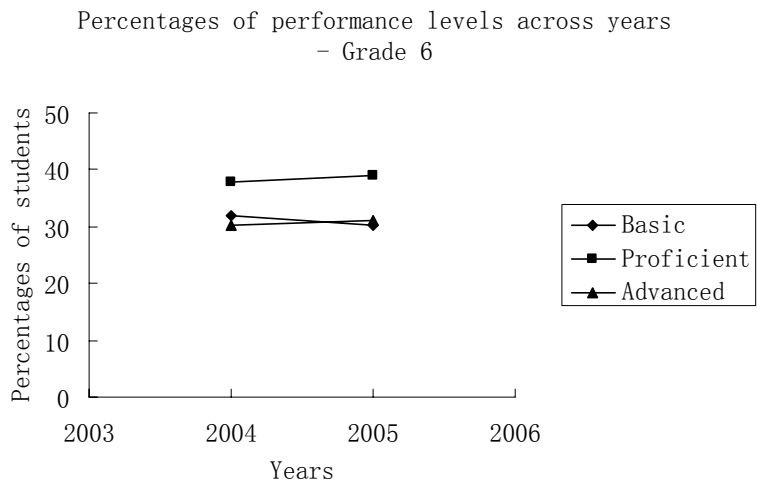


Figure 4: Distribution of performance categories across years for Grade 6 in Reading

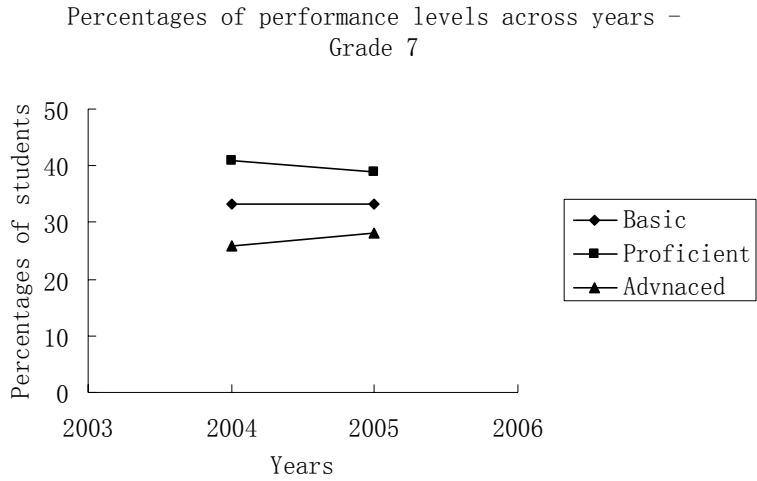


Figure 5: Distribution of performance categories across years for Grade 7 in Reading

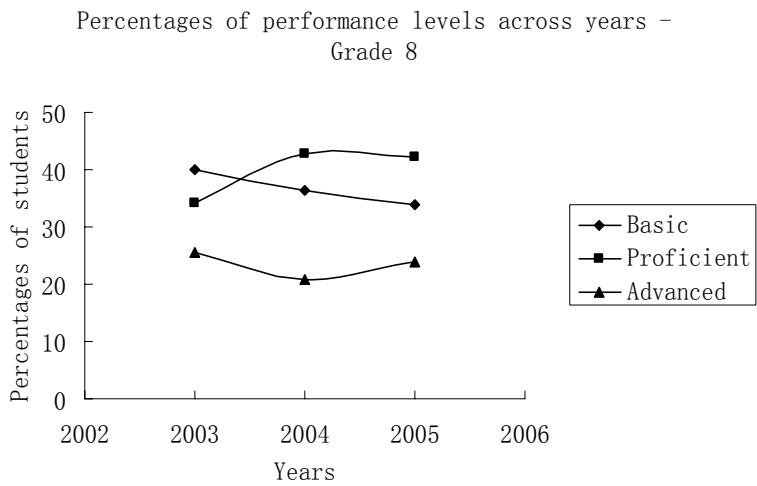


Figure 6: Distribution of performance categories across years for Grade 8 in Reading

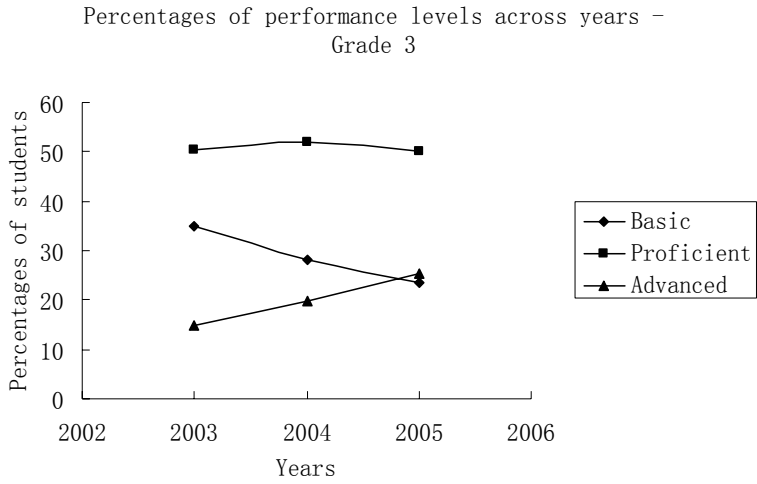


Figure 7: Distribution of performance categories across years for Grade 3 in Mathematics

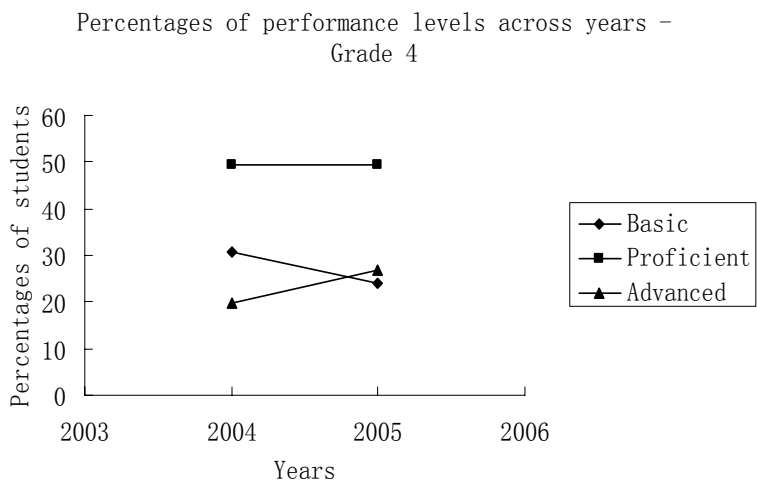


Figure 8: Distribution of performance categories across years for Grade 4 in Mathematics

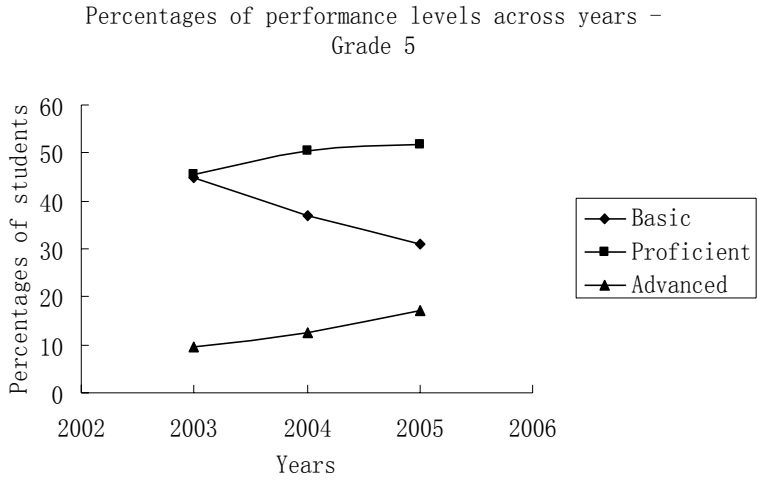


Figure 9: Distribution of performance categories across years for Grade 5 in Mathematics

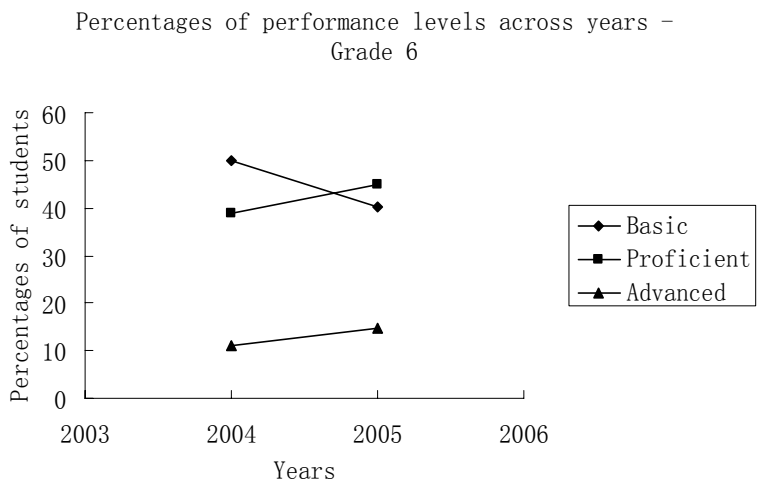


Figure 10: Distribution of performance categories across years for Grade 6 in  
Mathematics

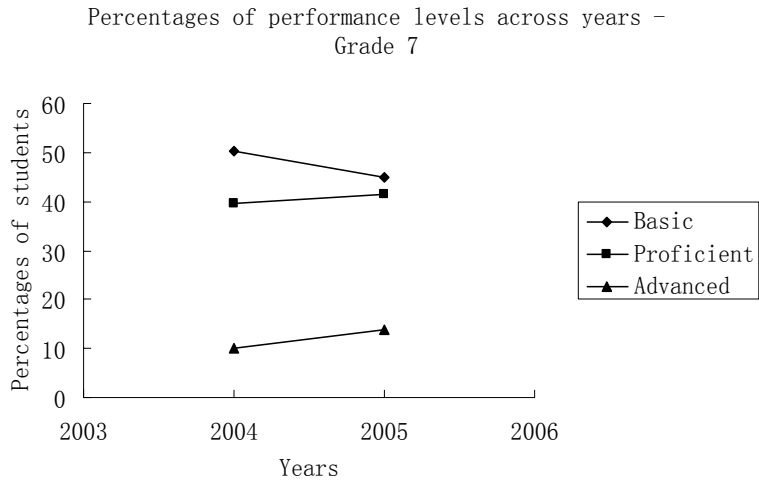


Figure 11: Distribution of performance categories across years for Grade 7 in Mathematics

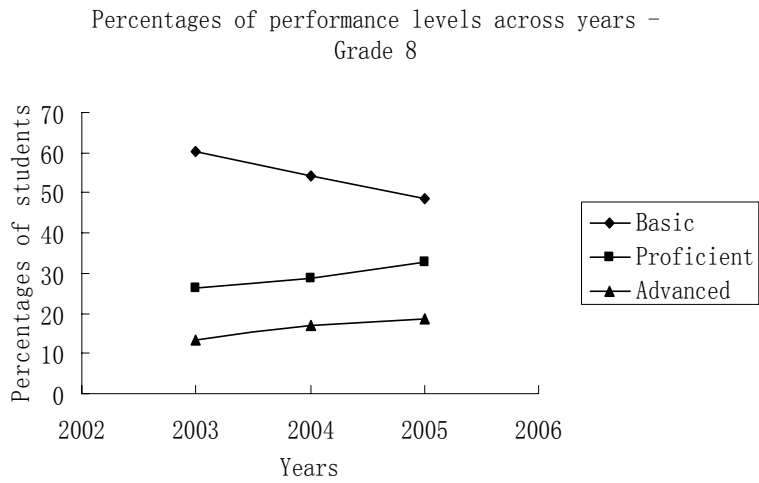


Figure 12: Distribution of performance categories across years for Grade 8 in Mathematics

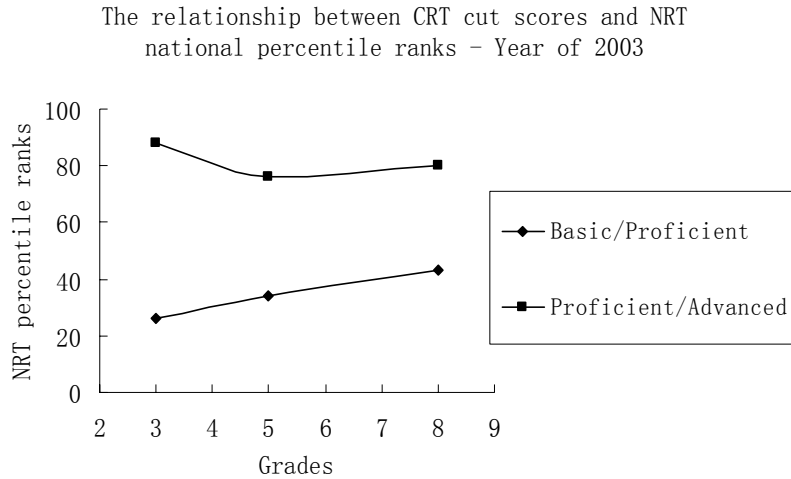


Figure 13: Relationship between CRT cut scores and NRT national percentile ranks across grades in Reading in 2003

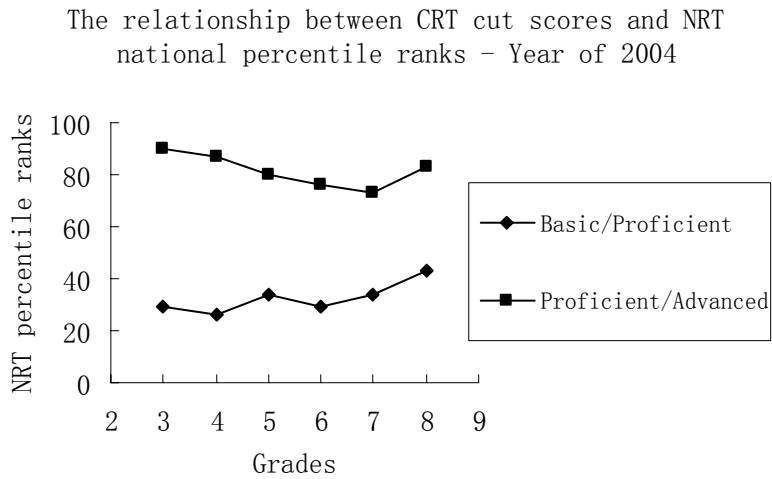


Figure 14: Relationship between CRT cut scores and NRT national percentile ranks across grades in Reading in 2004

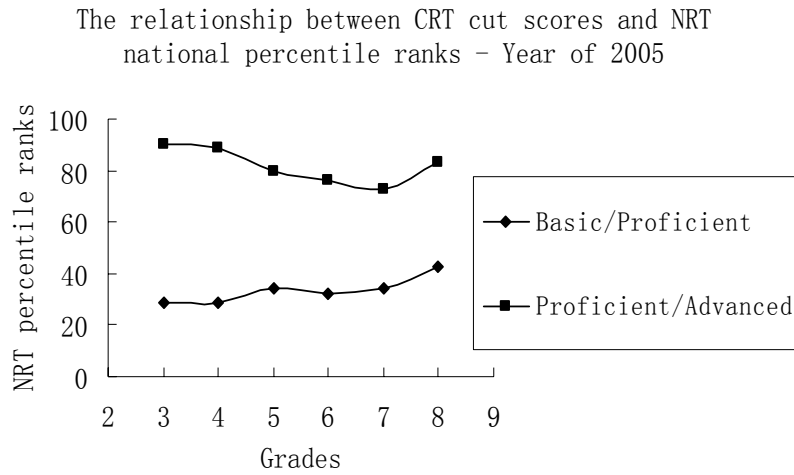


Figure 15: Relationship between CRT cut scores and NRT national percentile ranks across grades in Reading in 2005

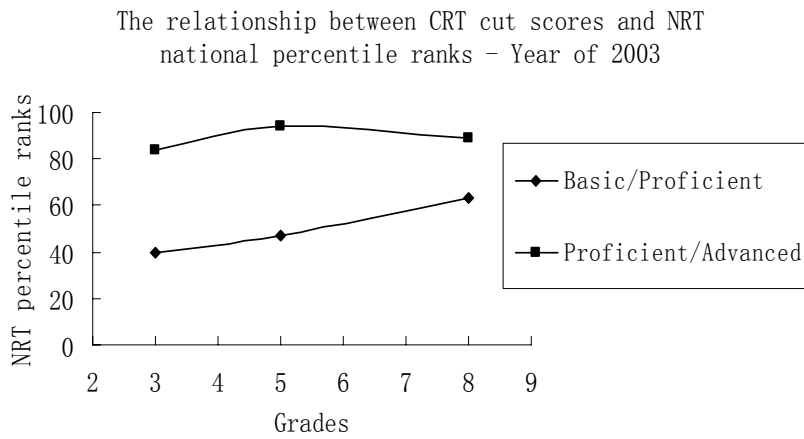


Figure 16: Relationship between CRT cut scores and NRT national percentile ranks across grades in Mathematics in 2003

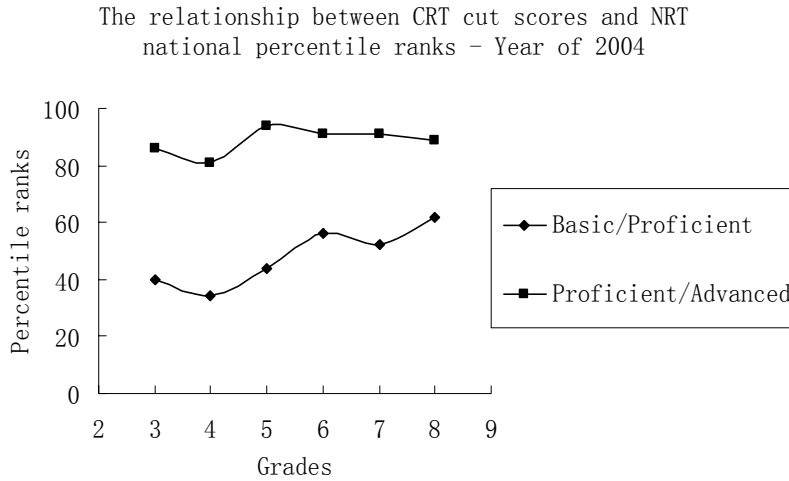


Figure 17: Relationship between CRT cut scores and NRT national percentile ranks across grades in Mathematics in 2004

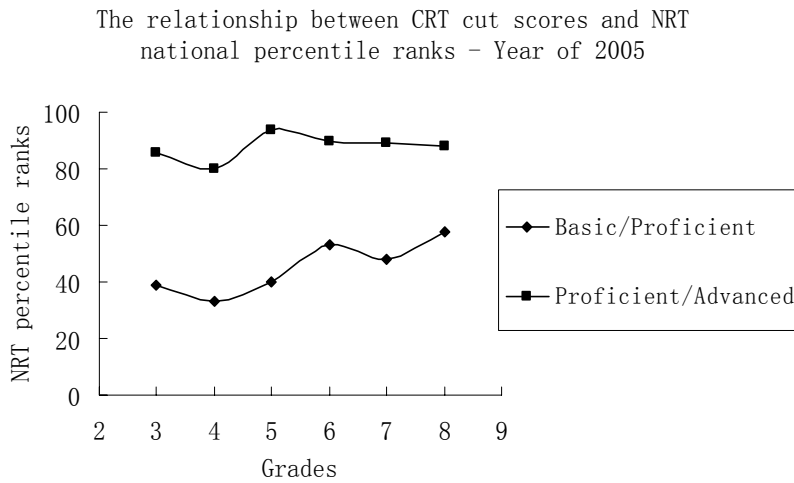


Figure 18: Relationship between CRT cut scores and NRT national percentile ranks across grades in Mathematics in 2005

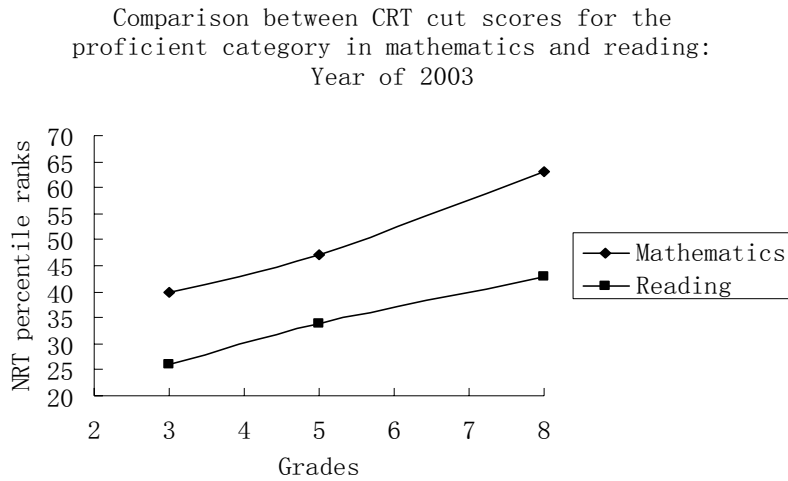


Figure 19: Comparison between CRT cut scores for the proficient category in Mathematics and Reading in 2003

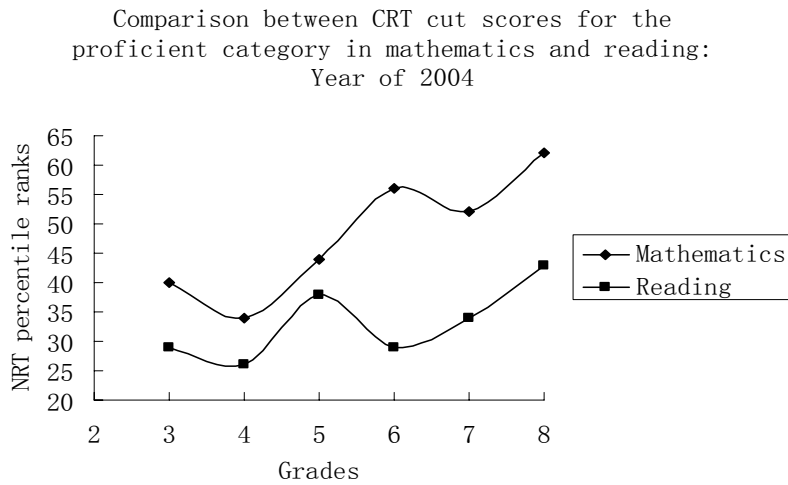


Figure 20: Comparison between CRT cut scores for the proficient category in Mathematics and Reading in 2004

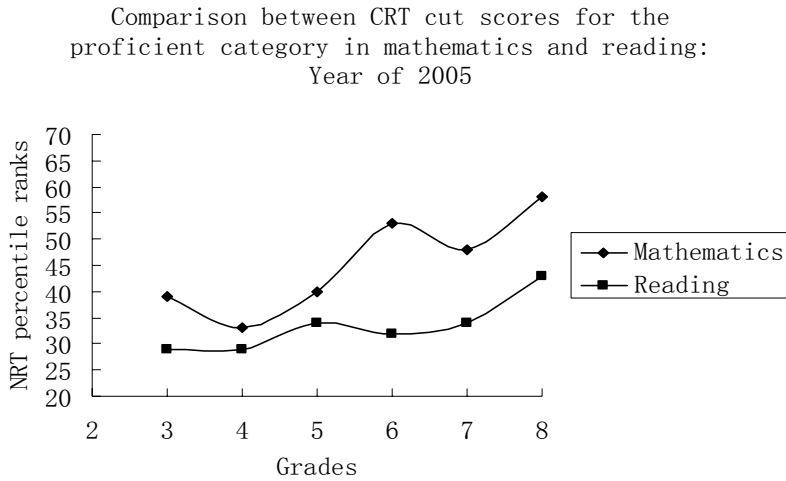


Figure 21: Comparison between CRT cut scores for the proficient category in Mathematics and Reading in 2005

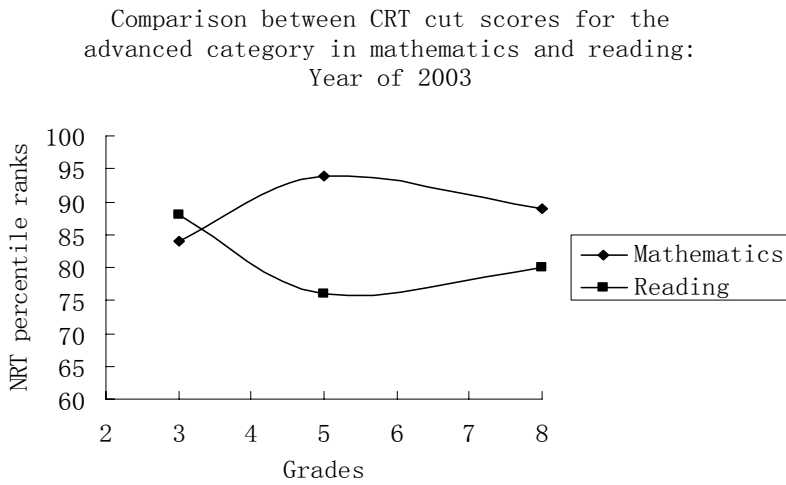


Figure 22: Comparison between CRT cut scores for the advanced category in Mathematics and Reading in 2003

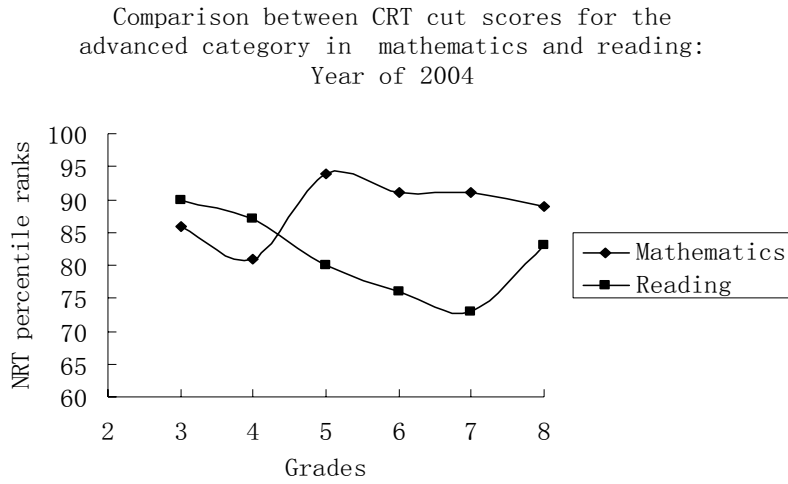


Figure 23: Comparison between CRT cut scores for the advanced category in Mathematics and Reading in 2004

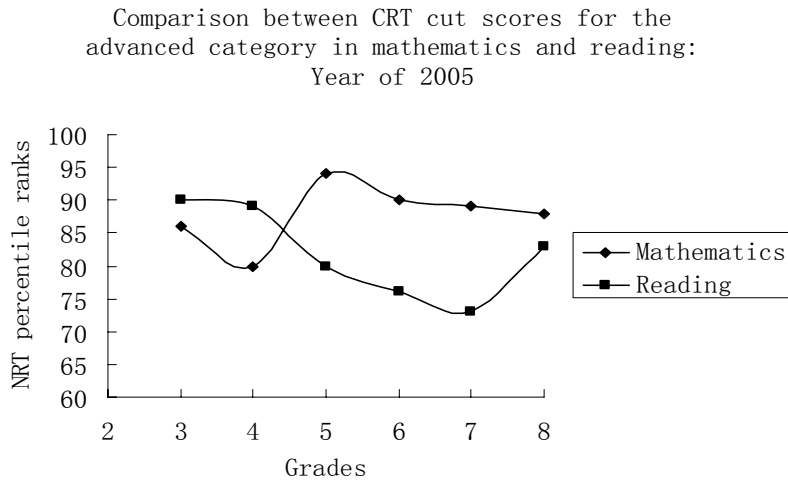


Figure 24: Comparison between CRT cut scores for the advanced category in Mathematics and Reading in 2005

**Tables**

Grade	Cut Score	Median NPR		
		2003	2004	2005
3	388	26 (n = 675)	29 (n = 2103)	29 (n = 1739)
5	384	34 (n = 644)	34 (n = 2238)	34 (n = 1161)
8	391	43 (n = 375)	43 (n = 1511)	43 (n = 1405)

Table 1: Unadjusted cut scores for Grades 3, 5, and 8 in Reading

Grade	Cut Score	Median NPR		
		2003	2004	2005
3	379	40 (n = 540)	40 (n = 511)	39 (n = 424)
5	392	47 (n = 611)	44 (n = 536)	40 (n = 531)
8	407	63 (n = 622)	62 (n = 621)	58 (n = 654)

Table 2: Unadjusted cut scores for Grades 3, 5, and 8 in Mathematics

Grade	Cut Score	Median NPR		
		2003	2004	2005
3	388	26 (n = 1454)	29 (n = 2103)	29 (n = 1739)
5	378	26 (n = 1877)	31 (n = 2188)	29 (n = 1029)
8	374	26 (n = 1712)	28 (n = 931)	28 (n = 1770)

Table 3: Results of adjusting cut scores by using Grade 3 as the baseline in Reading

Grade	Cut Score	Median NPR		
		2003	2004	2005
3	384	40 (n = 632)	45 (n = 489)	45 (n = 441)
5	386	40 (n = 520)	41 (n = 505)	35 (n = 485)
8	378	40 (n = 586)	43 (n = 499)	35 (n = 471)

Table 4: Results of adjusting cut scores by using Grade 3 as the baseline in Mathematics

Grade	Cut Score	Median NPR		
		2003	2004	2005
3	398	34 (n = 1477)	34 (n = 2435)	37 (n = 1997)
5	386	34 (n = 1997)	38 (n = 1233)	34 (n = 1187)
8	381	34 (n = 278)	33 (n = 1125)	33 (n = 2127)

Table 5: Results of adjusting cut scores by empirical average as the baseline in Reading

Grade	Cut Score	Median NPR		
		2003	2004	2005
3	392	50 (n = 690)	51 (n = 558)	49 (n = 482)
5	396	50 (n = 564)	51 (n = 593)	44 (n = 529)
8	389	50 (n = 685)	48 (n = 527)	44 (n = 606)

Table 6: Results of adjusting cut scores by empirical average as the baseline in Mathematics