**The Assessment of Students
With Disabilities in Kentucky**

CSE Technical Report 431

Daniel Koretz
CRESST/RAND Institute on Education and Training

July 1997

Center for the Study of Evaluation
National Center for Research on Evaluation, Standards,
and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

PREFACE

The research reported here could not have been carried out without the assistance of many people. I want to express my gratitude to Ed Reidy, Deputy Commissioner of the Kentucky Department of Education (KDE), for providing us access to the KIRIS data and for many other forms of assistance. I am particularly grateful to Jonathan Dings of the KDE, who generously contributed his time and expertise on innumerable occasions. I also want to thank Brian Gong, Scott Trimble, and Lou Spencer for their assistance. While this work could not have been carried out without the help these individuals provided, they bear no responsibility for errors of fact or for the interpretations put forward in this report. I also want to thank three RAND staff members: Laurie McDonald for her programming, Gina Schuyler for many kinds of research assistance, and Suzanne Welt for preparation of the document.

THE ASSESSMENT OF STUDENTS WITH DISABILITIES IN KENTUCKY

Daniel Koretz

CRESST/RAND Institute on Education and Training

Summary

Until very recently, many students with disabilities have been excluded from most large-scale assessment programs. Over the past few years, however, the National Assessment and many state assessment programs have undertaken to increase the inclusion of these students in regular assessments. This change raises important questions of measurement quality and impact. How feasible is it to assess most of these students? How reasonable a basis do the scores of students with disabilities provide for inferences about their knowledge and skills? What are the effects of providing accommodations—that is, modifications in the administration of the assessment?

Kentucky has led the nation in this regard, instituting very strict rules to ensure the inclusion of most students with disabilities in the regular Kentucky Instructional Results Information System (KIRIS) assessment. In the spring of 1995, Kentucky collected data on the primary disability of each assessed student with disabilities and on the accommodations each of these students were provided. Thus the KIRIS data provide an invaluable opportunity to examine the assessment of students with disabilities in a highly inclusive system.

This report describes analyses of the KIRIS assessment results for students with disabilities. It examines the rate of inclusion of students with disabilities, the frequency with which accommodations were used, the relationships between accommodation and scores, and a variety of indicators of assessment quality.

Kentucky succeeded in including the large majority of students with disabilities in the regular KIRIS assessment. Precise estimates are not possible, but it appears that roughly 80% or more of students with disabilities were assessed. The majority of those assessed—roughly 80% in the 4th grade and more than 60% in Grades 8 and 11—were provided with at least one accommodation, and most of these students were provided with two or more.

The results of analysis of the quality of the assessment results for these students were mixed. It is important to bear in mind that the majority of students with disabilities assessed with KIRIS, like the majority of those served in schools nationwide, have cognitive disabilities—primarily learning disabilities. Many other disabilities, including the physical disabilities that have provided much of the focus for debate about the assessment of individuals with disabilities, have very low prevalence rates and therefore are little reflected in the results of this analysis. Students with visual or hearing impairments, for example, are very few in

number. Analysis focused specifically on these smaller groups might yield very different answers.

Unlike some other studies, this study found no evidence that KIRIS test items discriminate less well for students with disabilities. Regardless of the provision of accommodations, the correlations between individual items and total scores were essentially the same for students with and without disabilities. When item means were plotted against total scores, the curves for students with disabilities were all monotonically increasing and generally showed slopes very similar to those for students without disabilities. Both statistical tests and visual inspection revealed little evidence of differential item functioning (DIF) for students assessed without accommodations, although these constituted only a minority of students with disabilities. On the other hand, some of the items used in mathematics appear to have been too difficult for a substantial number of students with disabilities.

A variety of findings raise doubts about the quality of scores obtained with assessment accommodations. The high frequency of accommodations, particularly in the fourth grade, itself raises questions about possibly inappropriate use. Some scores obtained with specific combinations of accommodations seemed implausibly high. For example, the average scores of fourth-grade learning-disabled students with certain combinations of accommodations were well above those of students without disabilities. Similarly, the average scores of mildly retarded fourth-grade students given certain accommodations were only 0.1 standard deviation below the average for nondisabled students in reading and 0.1 standard deviation above the average in science—unreasonable results, given that retarded students by definition have generalized cognitive deficits. Of the accommodations recorded in Kentucky, providing students with the opportunity to dictate responses (offered to more than half of the learning-disabled or retarded fourth graders assessed) had by far the strongest positive association with scores. In addition, DIF was quite common for students assessed with accommodations, and in mathematics, numerous instances of DIF were substantively large.

The need for additional research on these topics is clear. Research in other contexts is needed to explore the generality of these findings. Additional research is needed on the use and effects of accommodations, and new methods need to be explored for evaluating the validity of scores for students with disabilities. In addition, research is needed to explore the effects of the inclusion of students with disabilities in large-scale assessments on the educational opportunities they are provided and on their eventual performance. While additional research evidence is accumulating, policymakers can more closely monitor the assessment of students with disabilities, for example, by tracking the use of accommodations, the difficulty of assessments for students with disabilities, and mean scores of students with and without accommodations. In addition, policymakers can take steps to further inform educators about the appropriate uses of accommodations and other issues that arise in the assessment of students with disabilities.

THE ASSESSMENT OF STUDENTS WITH DISABILITIES IN KENTUCKY

Daniel Koretz
CRESST/RAND Institute on Education and Training

Abstract

Recent education reforms have spurred efforts to increase the inclusion of students with disabilities in large-scale assessments. These efforts are hindered, however, by a lack of experience and relevant research. This study explored the inclusion of students with disabilities in Kentucky's KIRIS assessment, which is arguably the most inclusive statewide assessment in the nation. The study found that Kentucky has managed to include the large majority of identified students with disabilities in its main assessment. In Kentucky as nationwide, the majority of these students have cognitive impairments—in particular, learning disabilities. Students with discrete physical impairments, such as visual disabilities, have been the focus of much of the discussion, are relatively few, and had little effect on the results reported in this study. Unlike some other research, this study found no evidence that test items discriminated less well for students with disabilities. It also found very little evidence of differential item functioning (DIF) for students assessed without accommodations. Only a minority of students with disabilities, however—about 20% in the fourth grade—were assessed without accommodations. The high frequency of some accommodations raises questions about the appropriateness of their use, In addition, the scores of students with some accommodations were implausibly high, and DIF was common and often large for students assessed with accommodations. Additional research is needed—for example, research on the uses and effects of accommodations and on the validity of scores obtained by students with disabilities.

Until very recently, many students with disabilities were routinely excluded from large-scale assessments. For example, guidelines pertaining to the exclusion of students with disabilities from statewide assessments differ from one state to another, and the estimated rate of participation of students with disabilities varies markedly across states and is often low (Erickson, Thurlow, & Thor, 1995; McGrew, Thurlow, Shriner, & Spiegel, 1992; Shriner & Thurlow, 1992). In addition, decisions about whether to include students with disabilities are often made by local school personnel, such as the team responsible for students' Individualized Education Plans, or IEPs (Erickson & Thurlow, 1996), and this introduces additional variation in patterns of inclusion. In many cases, educators face incentives to exclude from assessments students with disabilities who may score poorly.

Recent reforms aimed at raising standards for all students have focused attention on the exclusion of students with disabilities from assessments, and numerous efforts are underway to increase their participation. These efforts are motivated by several goals. It is hoped that the inclusion of students with disabilities in large-scale assessments will provide better information not only about their own performance, but also about the aggregate performance of the schools in which they study. In addition, it is hoped that including these students in assessments—especially, the large-scale assessments tied to accountability in standards-based reforms—will make schools more accountable for, and thus more attentive to, the academic performance of students with disabilities.

These efforts to increase the participation of students with disabilities in large-scale assessments, however, are hindered by a lack of experience and information (National Research Council, 1997). For example, there is little systematic information on the use or effects of special testing accommodations for elementary and secondary students with disabilities.

Thus, as new policies attempt to increase the participation of students with disabilities, researchers will confront three broad questions. First, to what extent is it feasible to include students with disabilities in large-scale assessments? The answer to this question will presumably vary with the nature of both disabilities and assessments. For example, an academically able but physically disabled student may function very well, perhaps with accommodations, in an assessment that is too difficult for mentally retarded or autistic children.

Second, what is the quality of results from the assessment of students with disabilities? How valid a basis do the results for students with disabilities provide for inferences about their own academic achievement, and how does their inclusion in assessments affect the validity of inferences about the performance of schools or other groups that include them? The answers to this question may also hinge on the nature of the assessments as well as the nature of any testing accommodations provided to these students. For example, the continued exclusion of students with disabilities who are able to participate in assessments may bias estimates of the performance of schools or districts. If a school that reports that 60% of its students have reached a given performance standard but tests only 85% of its students, it has in fact demonstrated only that 51% of its students have reached the standard. Similarly, providing too few accommodations may lead to underestimates of the performance of students with disabilities and of the schools they attend, whereas providing too many or inappropriate accommodations may produce overestimates.

Finally, what are the diverse effects of inclusion? In what ways does it change how educators respond to students with disabilities and the achievement of those students? Does it have effects on the assessments employed, on the allocation of resources, or on the education provided to students without disabilities?

In an effort to begin answering some of the questions posed by the assessment of students with disabilities, this paper investigates the results of Kentucky's efforts to include these students in its statewide assessment. Kentucky is currently the leader among states in the inclusion of students with disabilities in statewide assessments; it is currently the only state that mandates that most students with disabilities be assessed using the regular state assessments, known as KIRIS (the Kentucky Instructional Results Information System). Maryland is implementing similar policies, and other states are also moving toward greater and more systematic inclusion of students with disabilities in their assessments. In addition, in 1995, Kentucky collected data for every assessed student indicating the student's primary handicapping condition, if any, and the special accommodations the student was offered in taking the assessment. Kentucky's KIRIS data thus provide a unique opportunity to explore the assessment of students with disabilities. This report addresses the first two questions noted above: the feasibility of assessing students with disabilities and the quality of their assessment results. It describes the assessment of students with disabilities in three grades (4, 8, and 11) and the accommodations they were offered. It then describes the performance of students with disabilities on KIRIS, overall and as a function of primary disability and testing accommodations. A subsequent section examines the performance of students with disabilities on individual test items to evaluate potential bias and other aspects of the quality of assessment results for these students. A final section discusses the findings and explores implications for policy and research.

Kentucky's Policies for the Assessment of Students With Disabilities

Kentucky's policies for the assessment of students with disabilities are guided by the premise that only a small number of students with disabilities—1% to 2%

of the total student population, comprising primarily students with moderate to severe cognitive disabilities—should be excluded from the regular KIRIS assessment. Most of those excluded are to be tested with a different assessment, called the KIRIS Alternate Portfolio. (Data from the Alternate Portfolio program are not considered in this report.) Students who are in ungraded programs are tested on the basis of age. The decision rules for determining inclusion are as follows:

- Students without an IEP or Section 504 plan participate in KIRIS without accommodation or modification.

- Students who meet several criteria indicating severe cognitive limitations, including being unable to complete a regular diploma program by reason of disability, even with extended services, accommodations, and modifications, are eligible for the KIRIS Alternative Portfolio.

- All students with IEPs or Section 504 plans who do not meet the preceding criterion are to be assessed using the regular KIRIS assessment.

Students with disabilities with IEPs or Section 504 plans may be administered KIRIS with either accommodations or modifications, subject to explicit limitations. (Kentucky defines an accommodation as "an alteration in the testing environment or process" and a modification as "an alteration in the assessment instrument" (Kentucky Department of Education, 1996, *Procedures for Considering Student Inclusion*, footnote 2). State policy allows the use of "adaptations and modifications including the use of assistive technology devices that are consistent with the instructional strategies specified on the student's . . . IEP or 504 plan and available to the student in the course of his/her instructional process" (Program Advisory No. OCAA-93-94, February 9, 1993, cited in Kentucky Department of Education, 1996, Attachment G). These accommodations and modifications:

1. must be part of the student's ongoing instructional program;

2. may not be introduced for the first time during the KIRIS assessment;

3. must be "based on the individual needs of the students and not on a disability category"; and

4. shall not "inappropriately impact the content being measured." (Kentucky Department of Education, 1996, Attachment G, A1).

Accommodations Offered

Several of the accommodations commonly offered in some other assessment programs are not specifically offered in the KIRIS assessment. Provision of additional time is one of the most commonly offered accommodations in some assessment programs. Most parts of KIRIS, however, are not intended to be speeded, and additional time can be offered to both disabled and other students without any notation on the testing record. Students with disabilities might in fact be offered additional time more frequently than other students, or might be offered on average more additional time, but there are presently no data pertaining to this question. Kentucky Department of Education (KDE) guidelines, however, do indicate that it is permissible to provide students with disabilities with breaks during testing time, if doing so is consistent with their IEPs or 504 plans. KDE does not provide specific guidance about the use of separate assessment settings—another frequently offered accommodation. KDE makes KIRIS available in large-type and Braille formats. Oral presentation by tape is not available.

KDE collects information about six specific accommodations:

- paraphrasing;
- oral presentation of the assessment (providing a reader);
- allowing dictation of responses (providing a scribe);
- cueing;
- use of an interpreter; and
- technological aids.

In addition, proctors could indicate the use of other, unspecified accommodations.

Restrictions on the Use of Accommodations

KDE provides detailed guidelines about the use of accommodations in the KIRIS Assessment, including numerous specific questions and answers about the uses of specific accommodations. Given the frequency with which various accommodations were used in KIRIS in 1995, the guidelines pertaining to paraphrasing, oral presentation, and dictation are particularly important.

Guidelines about the use of paraphrasing are specific and restrictive. The guidelines note that paraphrasing is allowed only for directions, not for reading and content passages. Paraphrasing is labeled an intrusive technique, and educators are told that they should use the least intrusive method possible. Paraphrasing can include repeating, rephrasing, or breaking down directions. However, it should not entail changes in "critical words" and should not be used "simply because vocabulary or content has not been taught/learned" (Kentucky Department of Education, 1996, Attachment G, A18). No concrete examples of appropriate or inappropriate paraphrasing are provided.

Guidelines for oral administration differentiate between the reading assessment and other content areas and include the following:

- On-demand tasks in general may be read to a student if the student has a verified disability in the area of reading, the student's IEP documents the use of a reader in instruction, and use of a reader "is not a replacement for reading instruction or technology" (Kentucky Department of Education, 1996, Attachment G, A20).
- Reading assessments "may be read to a student on the premise that the intent of reading is to measure comprehension, only if this is the normal mode through which the student is presented regular print materials and is documented on the student's IEP or 504 . . . plan" (Kentucky Department of Education, 1996, n.p.).

KDE's guidelines for providing a scribe in on-demand parts of the KIRIS Assessment include the following:

> A scribe may only be used for the KIRIS assessment when:
>
> - a student has a verified disability in the area of written expression or a physical disability which impedes the motor process of writing;
> - a student is motorically able to print or use cursive techniques . . . ; however, the student's written language deficit is so severe that the student cannot translate thoughts into written language even though the student can express those thoughts orally. This is a very rare situation in which such students cannot recognize written words or make sound-symbol associations;
> - a student can write, but writes very slowly and the time constraint of the . . . task will inhibit the student's ability to produce the required product. (Kentucky Department of Education, 1996, Attachment G, A6)
>
> A scribe may not be used for the KIRIS Assessment:
>
> - to enhance student products, i.e., the student is able to produce the product, but the product would be better if it were scribed. (Kentucky Department of Education, 1996, Attachment G, A6)

Although careful reading suggests that KDE intended that these criteria be interpreted as very restrictive, the two last criteria might introduce uncertainties into the decision about offering a scribe. At what point does a student's slowness in writing change from merely degrading the quality of the product (indicated to be insufficient grounds for accommodation by the last criterion) to "inhibiting" production of the product (indicated to be sufficient ground for accommodation by the previous criterion)? The fact that additional time is allowed for most parts of KIRIS might further cloud the decision.

Who Is Tested, and How?

For several years, KDE has marked the records of students with disabilities in the KIRIS database (although scores are reported without regard to disability), and in the spring of 1995 proctors were asked to indicate both students' primary handicapping conditions and which of the six specified accommodations were provided. Data on accommodations were limited in important ways, however. Although proctors could indicate that other accommodations were offered, specific information was not collected on the provision of extra time, additional breaks, the use of separate settings, or the use of large-type or Braille editions. The information obtained also does not indicate whether a given accommodation was provided only for a few tasks, only for one subject area, or for the entire assessment.

These data suggest that Kentucky has been quite successful in meeting the goal of including most students with disabilities in the regular KIRIS assessment. However, the use of accommodations was widespread and raises questions about their appropriateness and impact on validity.

Number and Percent of Students Tested

In order to evaluate the success of KIRIS in including students with disabilities in the main assessment, it is necessary to compare the counts of students tested to other evidence about the numbers of students with disabilities in the state. Unfortunately, there is no count of students with disabilities served in Kentucky schools that is directly comparable to the counts of those tested in KIRIS, so this comparison can only be approximate.

The best data on students with disabilities served in the schools are those the states provide yearly to the U.S. Department of Education for the Department's

Annual Report to Congress on the Implementation of The Individuals With Disabilities Education Act (e.g., U.S. Department of Education, 1996). These data include counts of students served under the Individuals With Disabilities Education Act (IDEA), along with counts of the much smaller number of severely disabled students served under Title I State Operated Programs. These counts include the relatively few severely disabled students who are expected to be excluded from the regular KIRIS assessment under KDE's guidelines. However, these data exclude a modest number of students with disabilities who do not qualify for services under either of those two programs but do qualify under Section 504 of the Rehabilitation Act of 1973. Moreover, the *Annual Report* data are tabulated by age, not grade. Data showing the distribution of students served by age and grade (which would allow one to crosswalk between the age tabulation in the *Annual Report* and the grade-based KIRIS data) are not available.[1]

The data in the *Annual Report* place Kentucky's data in the context of national prevalence rates. Reported prevalence rates vary dramatically from one state to another, so findings from one state may not apply to another. For example, in the 1993-94 school year, the total percentage of children ages 6-17 identified as disabled and served by IDEA or Chapter 1 State Operated Programs ranged from a low of 5.5% (in Puerto Rico) to a high of 15% (in Massachusetts). Prevalence rates for some specific disabilities varied even more; for example, the percentage identified as having specific learning disabilities ranged from 2.1% to 9.3% (U.S. Department of Education, 1995, Table AA15). Nonetheless, a comparison between Kentucky's prevalence rates and the national average of prevalence rates provides a rough gauge of the potential relevance of Kentucky's experience to the nation as a whole.

In most respects, Kentucky's reported rates of disability mirror national rates quite closely. Roughly 10% of children ages 6-17 are served under Part B of IDEA in both Kentucky and the nation as a whole (Table 1). In both Kentucky and the nation, more than half of the children with disabilities served by these programs are identified as having cognitive disabilities, either learning disabilities or mental retardation. In Kentucky, however, the reported prevalence of learning disabilities is somewhat lower than average, while that of mental retardation is

[1] The KIRIS data provide age as well as grade for those students with disabilities who are tested but of course include no data for those not tested. Moreover, they include data only for a few grades. Thus, for example, they indicate the proportion of tested fourth graders who are age 10, but they do not allow one to calculate the proportion of 10-year-olds who are in Grade 4.

Table 1

Percentage (Based on Estimated Resident Population) of Children Ages 6-17 Served Under IDEA, Part B, by Disability During the 1994-95 School Year

Classification	Kentucky	Nation
All disabilities	9.68	10.43
Specific learning disabilities	3.18	5.34
Speech or language impairments	2.72	2.28
Mental retardation (MR)	2.53	1.12
Serious emotional disturbance	.66	.91
Multiple disabilities	.17	.17
Hearing impairments	.11	.14
Orthopedic impairments	.06	.13
Other health impairments	.15	.23
Visual impairments	.06	.05
Autism	.02	.05
Deaf-blindness	.00	.00
Traumatic brain injury	.01	.01

Source: U.S. Department of Education, 1996, Table AA12.

somewhat higher. About half of all students served under IDEA nationwide, as compared to about one third in Kentucky, are learning disabled. Conversely, roughly 10% of the students served nationwide, but one fourth of those served in Kentucky, are identified as mentally retarded. Other differences between the reported prevalence rates in Kentucky and in the nation as a whole are relatively minor.

Note that in Kentucky as nationwide, the prevalence of most disabilities other than the cognitive disabilities is very low. Some of the discrete, physical impairments that have influenced much of the thinking about assessment accommodations, such as visual and hearing impairments, affect very few students: 15 or fewer per 10,000 nationwide. These very low prevalence rates preclude separate analysis of many disability groups in this report; there are too few such students tested to provide reliable estimates. On the other hand, these students are also so few that the impact of their performance on the validity of the aggregate, school-level scores used in the KIRIS accountability system will generally be minor.

The percentage of students assessed in the regular KIRIS assessment who are identified as disabled drops substantially with increasing grade level. In the spring of 1995, 10% of the students tested in the 4th grade, 8% of those tested in the 8th grade, and 5% of those tested in the 11th grade were identified as disabled (Table 2). This does not necessarily imply, however, that KIRIS is less successful in including students with disabilities in the higher grades. A sizable decline across the grades in the number of identified students with disabilities tested is expected. First, the count of identified and served students with disabilities declines progressively with age after age 9 or 10, and particularly after age 15 (U.S. Department of Education, 1996, Table AA6). Second, disabled students as a group have an unusually high dropout rate (U.S. Department of Education, 1996).[2] Finally, some students with disabilities move to the general education program and are thereafter not counted. However, with a few exceptions (for example, students identified as having speech and communication disorders in the early grades), these students constitute only a modest share of the total population served under IDEA (e.g., U.S. Department of Education, 1996, Table 1.6).

It appears that KDE succeeded in its efforts to include the large majority of students with disabilities in KIRIS in 1995, but in the absence of data showing the total numbers of students with disabilities enrolled by grade, one can estimate the percentage included only very roughly. A modest percentage of students with disabilities who had records in the KIRIS database, ranging from 2.5% in Grade 4 to 3.4% in Grade 11, lacked scores in the four subject areas for reasons that are not documented (Table 3). (By contrast, the percentage of students without disabilities lacking scores was roughly half as large.) The counts of students with

Table 2
Students With Disabilities Assessed With Regular KIRIS, 1995

	Grade 4	Grade 8	Grade 11
All students tested in KIRIS	49,469	51,513	41,862
Students with disabilities tested in KIRIS	4,917	4,168	2,074
Tested students with disabilities as a percent of all tested students	10.0%	8.1%	5.0%

[2] Some calculations suggest that the percentage of students dropping out each year is not in the aggregate higher for students with disabilities but that the cumulative (cohort) dropout rate is higher because the percentage of disabled dropouts returning to school is much lower than the percentage of nondisabled dropouts (U.S. Department of Education, 1996, p. 16).

Table 3

Students With Disabilities Assessed With Regular KIRIS and Estimated Numbers Enrolled, 1995

	Grade 4	Grade 8	Grade 11
All students with disabilities recorded in KIRIS	4,917	4,168	2,074
Students with disabilities with KIRIS scores	4,792	4,047	2,004
Percent of recorded students with disabilities lacking scores	2.5%	2.9%	3.4%
Estimated number of students with disabilities in grade served under IDEA or Chapter 1 SOP	5,900	4,800	2,300
Estimated percentage of students with disabilities excluded from KIRIS	~15%	~15%	~10%
Estimated percentage of students with disabilities without KIRIS scores	~20%	~15%	~15%

Note. Estimated counts are rounded to two figures; estimated percentages are rounded to the nearest 5%.

scores can be compared to a rough estimate of the number served under IDEA.[3] These rough estimates suggest that all but 10% to 15% of students with disabilities were included in KIRIS, and all but 15% or 20% had scores on the assessment (Table 3). If that is so, fewer than 2% of all students were excluded from the regular KIRIS assessment because of disabilities.

The group of students with disabilities with scores on the regular KIRIS—who are the subject of all the analyses reported here—differ somewhat from the total group served because of the atypical characteristics of students excluded from KIRIS. Nonetheless, because relatively few students are excluded, the group with scores mirrors the total group in many respects. The characteristics of the group with scores in the fourth grade are shown in Table 4. Although students with severe cognitive impairments are excluded from the regular KIRIS assessment, students with cognitive disabilities constituted nearly three fourths of all of the fourth-grade students with disabilities for whom regular KIRIS scores were obtained. Students with learning disabilities constituted almost half of the students with scores, while mentally retarded students made up more than a

[3] The total number of students with disabilities enrolled in each grade was estimated by taking a weighted average of the published counts of identified students at each age, where each weight was the percentage of tested students with disabilities who were of that age. The age distribution of tested students with disabilities was used instead of the overall age distribution because students with disabilities tend to be considerably older than other students in the same grade.

Table 4

Percentages of All Students With Disabilities With Most Common Primary Disabilities, 1994-95 School Year, Kentucky (Scores in Grade 4) and the Nation (Served at Age 10)

Kentucky classification	Federal classification	Kentucky percentage, grade 4	Federal percentage, age 10
Specific learning disability (SLD)	Specific learning disability (SLD)	44	54
Mild mental retardation	None	27	
Functional mental retardation	None	1	
Mental retardation (sum of functional and mild)	Mental retardation (MR)	28	10
Communication/speech	Speech or language impairment	12	22
Emotional-behavioral disability	Serious emotional disturbance	6	7
Other health impairments	Other health	3	2
Sum of SLD and MR	Sum of SLD and MR	71	64
Sum of all listed disabilities	Sum of all listed disabilities	93	95

Sources: RAND tabulations of 1995 KIRIS files; U.S. Department of Education, 1996, Table AA6. Components may not sum to totals because of rounding.

fourth. Students with communication or speech disorders, who are relatively rare in the higher grades, constituted 12% of those with scores. All other disability groups were smaller. Except where otherwise noted, all of the students with scores were included in the analyses that follow, but where specific disabilities were the focus of analysis, only the largest groups could be included.

Two differences are apparent when the composition of the fourth-grade group with scores is compared to the national distribution of 10-year-olds served under IDEA. The first is the relatively large percentage of students classified as retarded in Kentucky, a finding that also appeared above in the counts of all students with disabilities served under IDEA in Kentucky (Table 1).[4] The second difference is the relatively low percentage of students with communication and speech disorders in Kentucky, which did not appear in the counts of all Kentucky children served. This may reflect a lack of accommodations in KIRIS suitable for some students with communication disorders or the occurrence of communication disorders severe enough to preclude participation in the assessment.

[4] Counts of students with disabilities served provided to the U.S. Department of Education reflect data collected each December. Counts of students tested were obtained directly from assessment records from the spring administration of KIRIS.

Frequency of All Accommodations

Despite KDE's seemingly restrictive guidelines on the use of accommodations, the large majority of students with disabilities tested with the regular KIRIS assessment were given accommodations, and many were given more than one. This pattern was most striking in Grade 4, in which 81% of students with disabilities were given at least one accommodation, and 66% were given more than one (Table 5). Accommodations were used somewhat less in secondary schools, although a majority of students with disabilities still received at least one. About one third of 8th- and 11th-grade students with disabilities were assessed with no accommodations at all, and somewhat under half were given multiple accommodations (Table 5).

The widespread use of multiple accommodations complicates description of accommodations for two reasons. First, Kentucky educators provided accommodations in a large number of combinations. For example, in Grade 4, a total of 66 different combinations of accommodations were recorded.[5] Most combinations of accommodations were provided to very few students, but nonetheless, many students were given one or another of the infrequent combinations. Second, analysis of the relationships between accommodations and performance must disentangle the effects of accommodations given jointly. For example, 1,167 (54%) of the 2,158 learning-disabled students tested in Grade 4 were permitted dictation, but only 43 of them were provided dictation without any other accommodations. The 43 students who received only dictation clearly do not provide a reasonable basis for assessing the relationship between that accommodation and performance on KIRIS; they are too few and too likely to be atypical of the 1,167.

Table 5

Percentage of Students With Disabilities Receiving Assessment Accommodations, by Grade

Accommodation	Grade 4	Grade 8	Grade 11
No accommodations	19	33	39
One accommodation	15	22	20
Multiple accommodations	66	45	41

[5] This includes the 6 specific accommodations about which proctors were questioned, the residual "other" category, and 59 combinations of two or more accommodations.

Accordingly, the following discussion distinguishes between *simple counts* of accommodations and *mutually exclusive combinations* of accommodations. Simple counts are obtained by tabulating all accommodations provided, regardless of the multiple accommodations given to many students. Thus, a single student can be counted in the simple counts for several accommodations. In contrast, to obtain mutually exclusive combinations, students are divided into categories based on the single or multiple accommodations each received, and each student is counted only once. To illustrate, in the case of the previous example, the simple dictation count would show 1167 fourth-grade learning-disabled students as receiving dictation, almost all of whom also received other accommodations. In contrast, the exclusive dictation variable would show that only 43 fourth-grade learning-disabled students received dictation with no other accommodations. All other students receiving dictation also received other accommodations and therefore were placed in other mutually exclusive categories.

Simple Counts of Accommodations

Oral presentation and paraphrasing are the most frequently provided accommodations (Table 6). Despite KDE's restrictive guidelines, roughly half of all students with disabilities in each of the three tested grades received paraphrasing.

Table 6

Percentage of Students With Disabilities Receiving Assessment Accommodations, by Grade (Based on Simple Counts)

Accommodation	Grade 4	Grade 8	Grade 11
None	19	33	39
Oral presentation	72	56	45
Paraphrasing	49	48	47
Dictation	50	14	5
Cueing	10	12	10
Technological aid	3	5	5
Interpreter	2	3	4
Other	8	5	6

Note. Individual students may receive multiple accommodations.

In contrast, the use of oral administration varied appreciably by grade. Nearly three fourths of 4th-grade students with disabilities were administered at least part of KIRIS orally; this dropped to 56% in Grade 8 and 45% in Grade 11. Dictation was provided to half of the tested students with disabilities in the 4th grade. None of the other recorded accommodations was provided to more than 10% of students with disabilities in any grade.

A very important pattern is the dramatic decline in the use of dictation across the grades. Although half of 4th-grade students with disabilities were allowed to dictate at least part of the assessment, this option was provided to relatively few secondary-school students with disabilities (14% in Grade 8 and only 5% in Grade 11). It will be shown later that dictation had an unusually large association with test scores, and differences in its use may help to explain differences in observed scores for students with disabilities across the three grades.

Students with learning disabilities or mild mental retardation, who together make up almost three fourths of tested fourth graders with disabilities, were provided with nearly identical accommodations. (See Table 7, in which only the most common disabilities and most frequent accommodations are noted.) The overwhelming majority of these students were provided oral administration; over half were allowed to dictate responses, and over half were provided with paraphrasing. Students with emotional-behavioral disabilities were less frequently provided with any of the three most common accommodations, and students with communication disorders were less likely yet to receive any of the three.

In Grades 8 and 11, the accommodations provided to learning-disabled students remained similar to those provided to mildly mentally retarded students, albeit less so than in Grade 4 (Table 7). Roughly half of both groups in both grades were provided with paraphrasing. Oral presentation was provided to approximately half of the learning-disabled students but to somewhat more of the mentally retarded students. Dictation was provided to about a fifth of the mentally retarded students in Grade 8 but to few in Grade 11 and to few learning-disabled students in either grade. In both of the secondary grades, about half of the students with behavioral disabilities received no accommodation, but a third or more received either oral presentation or paraphrasing.

Table 7

Percentage of Students With Disabilities Receiving Assessment Accommodations, by Grade and Disability (Based on Simple Counts)

Accommodation	Learning disability	Mental retardation	Behavioral disability	Communication disorder
Grade 4				
None	8	9	28	62
Paraphrasing	57	56	44	24
Oral presentation	83	88	58	31
Dictation	54	60	41	23
Grade 8				
None	33	20	53	a
Paraphrasing	48	57	34	a
Oral presentation	55	72	35	a
Dictation	11	21	8	a
Grade 11				
None	39	29	54	a
Paraphrasing	48	55	42	a
Oral presentation	43	62	34	a
Dictation	4	7	6	a

Note. Individual students may receive multiple accommodations.

[a] Communication disorders are not tabulated for Grades 8 and 11 because of small numbers.

Frequency of Mutually Exclusive Categories of Accommodations

Despite the large number of combinations in which accommodations were provided to students with disabilities, only six mutually exclusive categories were provided to more than 5% of the students with disabilities in at least one of the three tested grades. The use of several of these five combinations of accommodations varied substantially across the three tested grades.

Only two of the individual accommodations were provided without others to at least 5% of the students with disabilities in at least one of the tested grades: oral presentation and paraphrasing. From 9% to 14% of students with disabilities received oral presentation only, depending on grade, and from 2% to 8% received paraphrasing only (Table 8). These small percentages contrast with the large percentages of students with disabilities—ranging from 45% to 72%—who received these accommodations in total (either singly or in combination with others; see Table 6).

Table 8

Percentage of Students With Disabilities Receiving Assessment Accommodations, by Grade and Disability (Based on Mutually Exclusive Categories)

	Grade 4	Grade 8	Grade 11
None	19	33	39
Oral presentation only	10	14	9
Paraphrasing only	2	6	8
Oral presentation and dictation	16	2	0
Oral presentation and paraphrasing	11	18	20
Oral presentation, paraphrasing, and dictation	20	6	2
Oral presentation, paraphrasing, and cueing	1	6	5
Other multiple accommodations	17	13	14

Only four specific combinations of accommodations, all involving oral presentation, were used in at least 5% of the cases in at least one grade, and the use of these varied markedly across grades. The combination of oral presentation and dictation (without other accommodations) was provided to 16% of students with disabilities in the 4th grade but to virtually none in the secondary grades (Table 8). Oral presentation in combination with paraphrasing was provided to a sizable share of students with disabilities in each grade, but in different forms. In the 4th grade, these two accommodations were more often provided jointly with the opportunity to dictate responses; about one student in five was offered this combination of three accommodations. In contrast, oral presentation and paraphrasing were more often offered without the opportunity to dictate in the secondary grades; the combination of these two accommodations without others was offered to about one in five students with disabilities in Grades 8 and 11. The combination of oral presentation, paraphrasing, and cueing was provided to 5% or 6% of students with disabilities in the secondary grades but to virtually none in the 4th grade. The many additional combinations of accommodations taken together were provided to 13% to 17% of students with disabilities.

How Do Students With Disabilities Perform on KIRIS?

To analyze the performance of students with disabilities on KIRIS, it is necessary to take into account both primary disabilities and accommodations. Some advocates for the disabled argue against classifying students based on specific disabilities, arguing that classifications may be potentially detrimental and are often irrelevant to deciding on appropriate educational interventions. As Shepard (1989) noted, however, the taxonomy useful for making decisions about services may be fundamentally different from that needed for research. The wide range of conditions subsumed under the rubric of "disability" have greatly different implications for the interpretation of assessment results. For example, a bright student who scores poorly in the absence of accommodations because of a visual disability is not comparable to a student with good eyesight who scores poorly because of mental retardation. Moreover, the scarcity of research on accommodations and the sometimes negative findings about the validity of scores obtained with accommodations (Anderson, Jenkins, & Miller, n.d.; Wightman, 1993; Willlingham et al., 1988) indicates the importance of examining performance separately for different accommodations.

Accordingly, the following sections present several different views of the performance of students with disabilities on KIRIS. The first section provides a simple description of performance for all students with disabilities, without regard to accommodations. The second section describes the performance of students with specific disabilities, again without regard to accommodations. The third section describes the performance of all students with disabilities, separately for those who were assessed with or without accommodations. This is followed by a description of the performance of students who received the most common mutually exclusive categories of accommodations. The final section explores the relationships between individual accommodations and performance on KIRIS.

KIRIS Results for All Students With Disabilities

To provide a consistent standard of comparison, all KIRIS scores reported here were restandardized to a mean of 0 and a standard deviation of 1 in the nondisabled population separately for each subject and grade.

The gap in performance between students with and without disabilities widens as students progress through the grades. In the 4th grade, the average scores of students with disabilities was 0.3 or 0.4 standard deviation below the average for students without disabilities in every subject but science (Table 9). In science, the scores of students with disabilities were only 0.1 standard deviation below the mean for students without disabilities. In contrast, in the secondary grades, the mean difference between students with and without disabilities was in

Table 9

Means and Standard Deviations for All Students With Disabilities, by Subject and Grade

	Reading	Math	Science	Social studies
Grade 4				
Mean	-0.3	-0.4	-0.1	-0.3
Standard deviation	1.1	1.2	1.2	1.1
Grade 8				
Mean	-1.1	-1.0	-1.0	-1.0
Standard deviation	1.1	1.0	1.0	1.0
Grade 11				
Mean	-1.4	-1.1	-1.3	-1.2
Standard deviation	1.0	0.9	1.0	0.9

Note. Scaled to a mean of 0 and a standard deviation of 1 in the population of students without disabilities. Thus a mean of -1.0 is one standard deviation below the mean in the population without disabilities.

every instance at least a full standard deviation, and the largest difference—in the 11th grade—was nearly a standard deviation and a half. Thus in reading, the average student with disabilities in the 8th grade scored roughly at the 14th percentile on the distribution of students without disabilities. The average reading score for students with disabilities in the 11th grade was approximately at the 8th percentile on the distribution of students without disabilities.

The striking differences between the performance of fourth-grade and secondary-school students with disabilities cannot be fully explained by differences in reported prevalence rates, which differ only modestly between the 4th and 8th grades. One factor that may have contributed to these disparities in performance is the differences among grades in the use and apparent effects of assessment accommodations, discussed below.

Differences in prevalence rates do suggest, however, that the modest increase in the gap in scores in from Grade 8 to Grade 11 may in some respects be understated. The number of students with disabilities tested is much smaller in Grade 11 than in Grade 8, perhaps in part because the cohort dropout rate among students with disabilities is generally high (U.S. Department of Education, 1996). If the students dropping out of school are disproportionately low achievers, one would expect the high dropout rate among students with disabilities to shrink the gap between students with and without disabilities who remain in school. Therefore, if only students who were to remain in school through 11th grade were considered, the gap between students with and without disabilities might widen even more between the 8th and 11th grades. Longitudinal data would be needed, however, to test this speculation.

Despite the heterogeneity of the population of students with disabilities in other respects, they are not more variable than their nondisabled peers in terms of performance on KIRIS in the secondary grades. In both Grade 8 and Grade 11, the standard deviations of the scores of students with disabilities in every subject are very close to the value of 1.0 found in the nondisabled population, while students with disabilities show modestly more variability than their nondisabled peers in Grade 4 (Table 9).

KIRIS Results for Students With Specific Disabilities

Scores on KIRIS varied somewhat among disability groups, although the variation among groups within a grade was often much smaller than the differences within a single disability group across grades.

In Grade 4, learning-disabled students obtained average scores very near those of nondisabled students (Table 10). The average for students with learning disabilities essentially matched the average for students without disabilities in reading and social studies, fell 0.1 standard deviation below in mathematics, and *exceeded* the mean for nondisabled students by 0.2 standard deviation in science. Thus, in science, the mean student with a learning disability scored at about the 58th percentile on the distribution of nondisabled students.

Table 10

Mean Scores by Disability and Subject, Grade 4

	Number tested	Reading	Math	Science	Social studies
Specific learning disability	2150	0.0	-0.1	0.2	0.0
Mild mental retardation	1327	-0.5	-0.7	-0.3	-0.6
Emotional/behavioral	319	-0.5	-0.5	-0.5	-0.5
Communication/speech	597	-0.4	-0.5	-0.3	-0.4

Students with the other three most common disabilities—mild mental retardation, emotional-behavioral disabilities, and communication or speech disorders—had average scores that ranged from 0.3 to 0.7 standard deviation below the mean for nondisabled students. The performance of all three groups was similar.

As noted above, relative to their peers without disabilities, students with disabilities scored much more poorly in the secondary grades than in Grade 4. This difference was apparent in all of the three largest disability categories (Table 11).[6] Although there was some variation among subject areas, students with learning disabilities scored on average far below average: about 0.9 standard deviation below the nondisabled student mean in Grade 8 and about 1.1 standard deviations below the mean in Grade 11 (Table 11). As in Grade 4, students with mild mental retardation and emotional-behavioral disabilities scored on average somewhat lower yet. In Grade 11, students with mild mental retardation averaged about 1.5 standard deviations below the mean for nondisabled students, placing the average mildly retarded student at about the 6th percentile on the distribution of nondisabled students.

Table 11

Mean Scores by Disability and Subject, Grades 8 and 11

	Number tested	Reading	Math	Science	Social studies
Grade 8					
Specific learning disability	2107	-1.0	-0.8	-0.8	-0.9
Mild mental retardation	1215	-1.3	-1.3	-1.3	-1.3
Emotional/behavioral	401	-1.4	-1.2	-1.2	-1.3
Grade 11					
Specific learning disability	1055	-1.3	-1.0	-1.1	-1.1
Mild mental retardation	583	-1.7	-1.3	-1.6	-1.5
Emotional/behavioral	115	-1.5	-1.1	-1.3	-1.2

Note. Communication disorders are not tabulated for Grades 8 and 11 because of small numbers.

[6] Students with communication or speech disorders are not included in Table 11 because of their small numbers in Grades 8 and 11.

KIRIS Results for Students With and Without Accommodations

On average, across all disabilities and accommodations, students who received testing accommodations scored higher than the smaller number who received none. These differences between accommodated and unaccommodated students were larger in the 4th grade than in the other grades. In the 4th grade, students who received accommodations averaged from 0.4 standard deviation (in mathematics) to 0.7 standard deviation (in science) above students with disabilities who received no accommodations (Table 12). The mean difference between students with and without accommodations ranged from 0.1 to 0.3 standard deviation in both the 8th and the 11th grades.

These patterns do not appear to represent a change in the impact of accommodations across the grades. As shown below, the use of accommodations generally has a similar association with scores in all grades, although there are some exceptions to this generalization. Rather, the smaller gap between accommodated and unaccommodated students in the secondary grades appears to reflect a difference in the use of accommodations. In particular, secondary school students are rarely offered the opportunity to dictate responses to the assessment, and dictation is shown below to have an unusually strong association with scores. The much larger percentage of secondary school students assessed

Table 12

Number and Mean Scores for All Students With Disabilities, by Subject and Grade, With and Without Accommodations

	Number tested	Reading	Math	Science	Social studies
Grade 4					
No accommodations	821	-0.7	-0.7	-0.6	-0.7
Any accommodations	3971	-0.2	-0.3	+0.1	-0.2
Grade 8					
No accommodations	1225	-1.2	-1.2	-1.1	-1.1
Any accommodations	2792	-1.1	-0.9	-0.9	-1.0
Grade 11					
No accommodations	734	-1.5	-1.3	-1.3	-1.3
Any accommodations	1270	-1.4	-1.0	-1.2	-1.2

Note. Scaled to a mean of 0 and a standard deviation of 1 in the population of students without disabilities. Thus a mean of -1.0 is one standard deviation below the mean in the population without disabilities.

without accommodations may also contribute to these patterns if the students offered accommodations in the elementary grades but not the secondary grades are relatively high-scoring compared to other students with disabilities.

KIRIS Performance for Mutually Exclusive Categories of Accommodations

The performance of students with disabilities varied markedly depending on the assessment accommodations they received. This is apparent when students are classified in terms of both their primary disabilities and mutually exclusive categories of accommodations they received.

These comparisons require two caveats. First, even though almost all students with disabilities are included in the KIRIS data, most combinations of disabilities and accommodations include very few students. Accordingly, results are reported here only for the two most common disabilities—learning disabilities and mild mental retardation—and the most frequent mutually exclusive categories of accommodations. Other combinations of disabilities and accommodations included too few students to be used in this analysis.

Second, the extent to which differences in scores can be attributed to the accommodations themselves is uncertain. Groups that received different accommodations may have differed in other respects as well and therefore might not have obtained similar scores even if they had been given the same accommodations. In the case of the cognitive and learning disabilities analyzed in this section, it might be reasonable to expect that the students offered the most substantial accommodations would generally be those with the severest disabilities and therefore with lower performance. One might therefore expect accommodations to narrow the gap in scores between students with and without disabilities by offsetting the greater severity. However, if accommodations are used appropriately and serve to offset disabilities *unrelated to the constructs being measured*, there would be little reason to expect students with accommodations to score appreciably better than those without. Substantially higher scores among students with accommodations might suggest either that some students are being denied appropriate accommodations or that others are being given inappropriate or excessive accommodations.

The variation in performance among types of accommodations was dramatic across grades. It was largest among 4th-grade students with mild mental retardation. Among those students, the difference in average scores between the lowest-scoring condition (no accommodations) and the highest-scoring (the combination of oral presentation, paraphrasing, and dictation) averaged 1.3 standard deviations across all four subject areas (Table 13). That is, the average student receiving all three of these accommodations scored roughly at the 90th percentile on the distribution of mildly mentally retarded students receiving no accommodations. The largest difference among accommodations was smaller among mildly retarded students in the 8th grade and among learning-disabled students in both the 4th and 8th grades—about three fourths of a standard deviation. The differences among the highest- and lowest-scoring accommodations were much smaller yet in the 11th grade—about one third and one half a standard deviation. Thus, in the 11th grade, the average student in the highest-scoring group of mildly retarded students scored at roughly the 54th percentile of the lowest scoring group.

The variation in performance among categories of accommodation is even more striking and more revealing when compared with the performance of students without disabilities. For example, overall (across all categories of accommodations), fourth-grade students with mild mental retardation scored about half a standard deviation below the mean of nondisabled students, although they scored markedly lower in mathematics and somewhat better in science (Figure 1). The differences in performance among categories of accommodations, however, were very large. In fact, the difference between the highest- and lowest-scoring accommodations group was larger than the overall difference between mildly retarded and nondisabled students. At one extreme, mildly retarded students receiving no accommodations scored roughly 1.4 standard deviations

Table 13

Difference Between Highest-Scoring and Lowest-Scoring Combinations of Accommodations (Including No Accommodation) by Grade and Disability (in Standard Deviations)

	Mild mental retardation	Specific learning disability
Grade 4	1.3	0.8
Grade 8	0.7	0.8
Grade 11	0.2	0.5

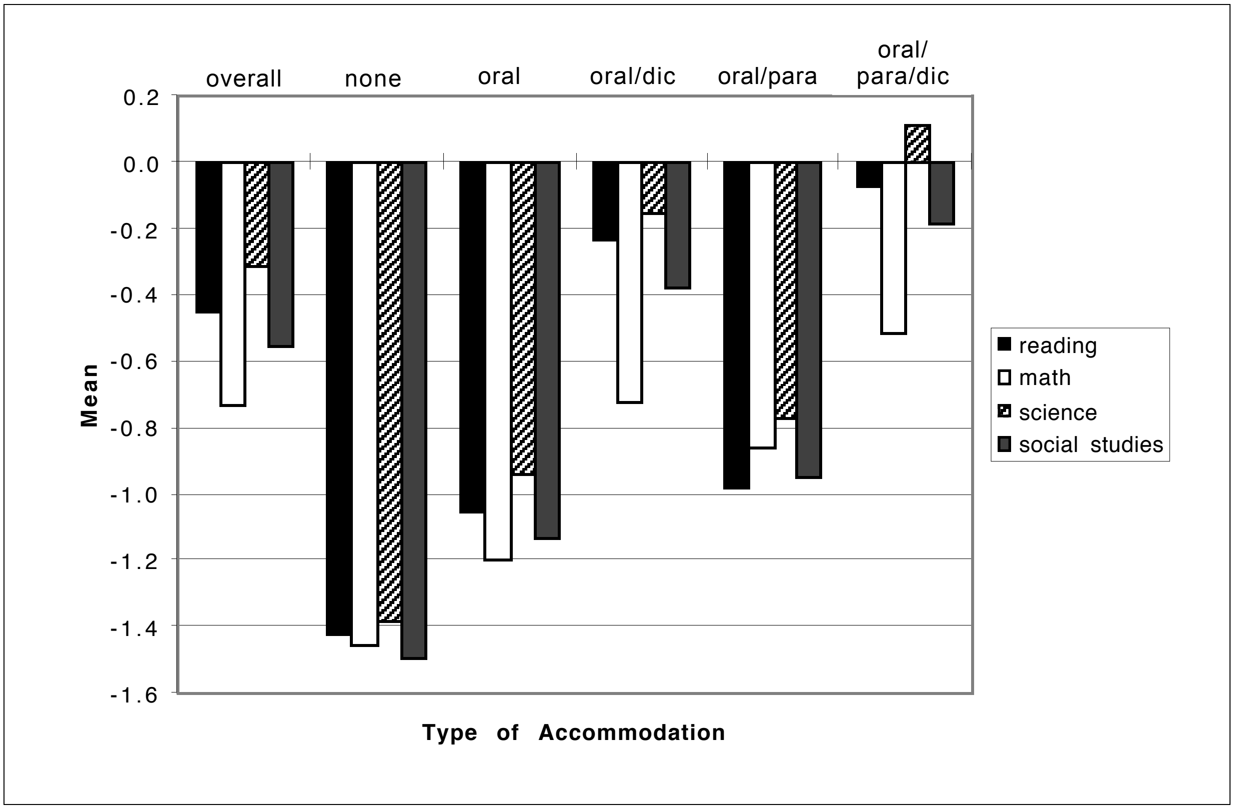


Figure 1. Grade 4 mildly mentally retarded students by accommodation and subject: Mutually exclusive combinations of accommodations.

below the mean for students without disabilities in all subject areas (Figure 1).[7] This places their average score at about the 8th percentile on the distribution of students without disabilities. Mildly retarded students given oral presentation without other accommodations scored somewhat higher, and those provided with both oral presentation and paraphrasing scored a bit higher yet.

The mildly retarded students offered dictation in combination with other accommodations, however, scored markedly higher than other groups, particularly in subjects other than mathematics. The average scores of mildly retarded students provided with the combination of oral presentation and dictation ranged from 0.4 to 0.7 standard deviation below the mean, and the averages for retarded students provided with both of these as well as paraphrasing were 0.5 standard deviation below the mean in mathematics, 0.1 standard deviation below the

[7] For simplicity, the phrase “below the mean” is used to denote “below the mean for nondisabled students.”

average in reading, and 0.1 standard deviation *above* the mean in science (Figure 1). These scores cannot be attributed only to reflect sampling error from small numbers. A total of 328 mildly retarded fourth graders were provided with the combination of oral presentation, paraphrasing, and dictation; a confidence band of two standard errors around their mean score in science extends from the mean of nondisabled students to 0.2 standard deviation above the mean.

Somewhat similar differences among categories of accommodations appeared among fourth-grade learning-disabled students, but with some important differences (Figure 2). Again, the overall means for these students—which are above average in science and near the average for nondisabled students in the other three subjects—mask striking differences among categories of accommodations. In this case, the lowest-scoring group were the learning-disabled students provided only with oral presentation; their averages ranged from 0.4 to 0.6 standard deviation below the mean for nondisabled students. Students receiving no accommodations scored modestly higher but still well below the average for nondisabled students. Students who received paraphrasing as well as oral presentation scored higher yet, particularly in science.

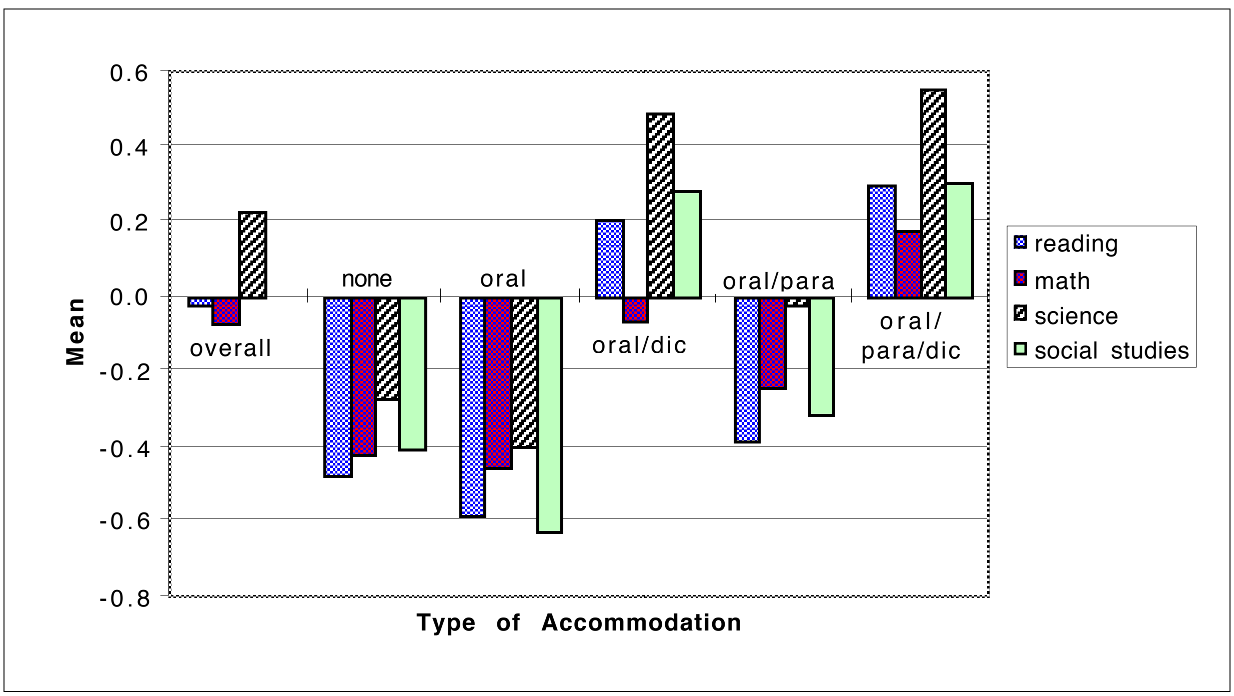


Figure 2. Grade 4 learning-disabled students by accommodation and subject: Mutually exclusive combinations of accommodations.

Here again, it was students receiving combinations of accommodations that included dictation who scored much higher. Learning-disabled fourth graders receiving dictation in combination with oral presentation scored above the average for nondisabled students in every subject but mathematics, and they scored almost half a standard deviation above the average in science. Learning-disabled fourth graders provided with paraphrasing in addition to these two accommodations scored even higher; their average scores ranged from about 0.2 standard deviation above the mean for nondisabled students in mathematics to 0.5 standard deviation above the mean in science. In other words, the average scores of learning-disabled students receiving these three accommodations together ranged roughly from the 57th percentile on the distribution of nondisabled students in mathematics to the 71st percentile in science.

Similar variations among accommodations appeared in Grade 8, although the scores of all eighth-grade groups with disabilities were lower than those of the corresponding groups in the fourth grade. Among eighth-grade students with mild mental retardation, those receiving no accommodation again scored lowest, and those receiving dictation (in combination with oral presentation and paraphrasing) again scored highest (Figure 3). Even this highest-scoring group, however, still scored far below the mean of students without disabilities (roughly 0.9 standard deviation below, when averaged across subjects). Results for eighth-grade learning-disabled students showed patterns similar to those of mentally retarded students, except that all groups scored somewhat higher (Figure 4). In addition, among learning-disabled students, the score of the one group receiving dictation differed much more from those of the other accommodations groups, averaging only about 0.3 standard deviation below the mean for students without disabilities.

In the 11th grade, scores showed less variations among the types of accommodations with enough cases to analyze, particularly in the case of students with mild mental retardation. However, an important caveat is that in the case of mentally retarded students, fewer categories included enough students to be included in the analysis. In particular, the categories that stood out from the others in the younger grades—those including dictation—included too few mentally retarded or learning-disabled students to be analyzed.

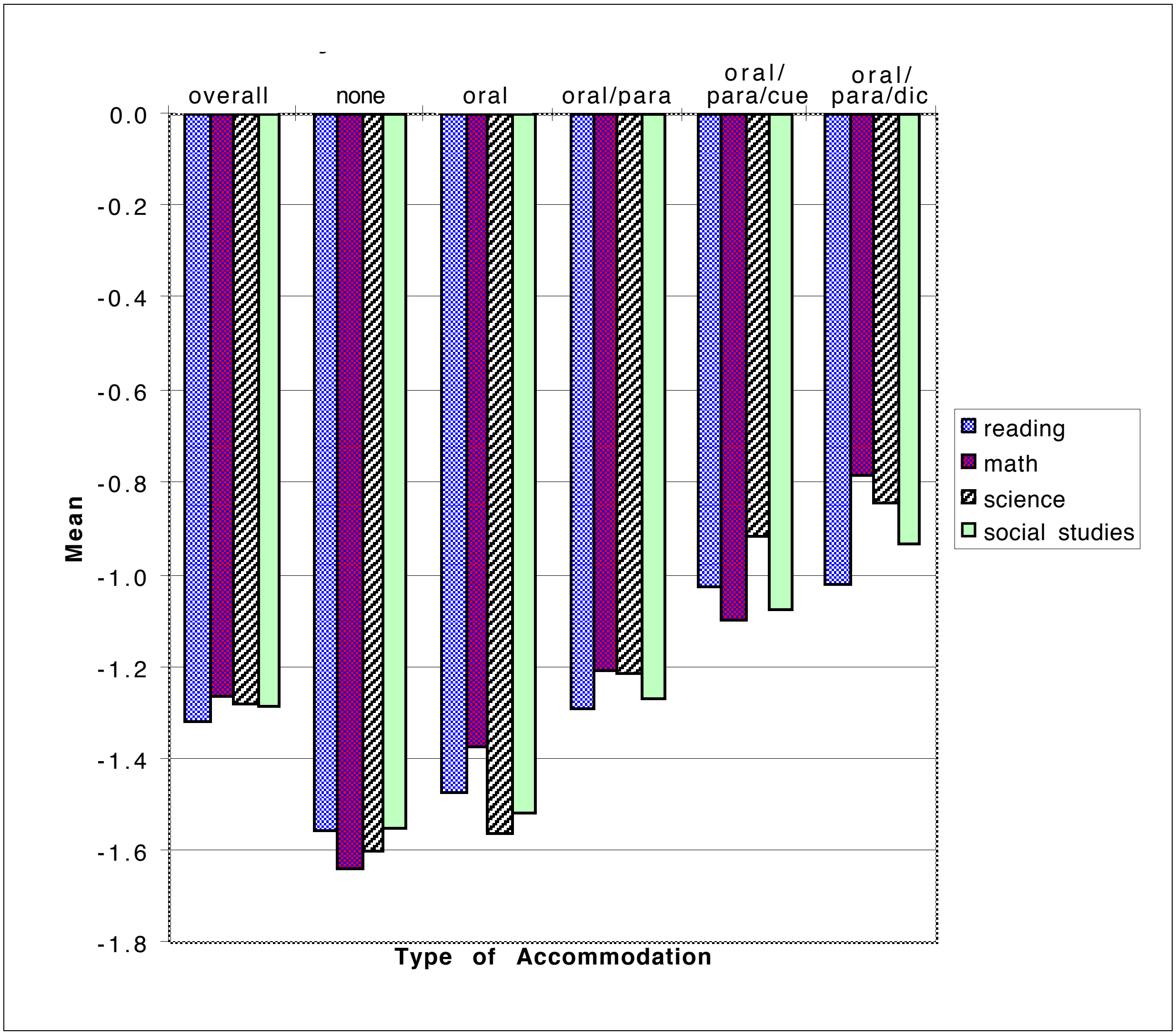


Figure 3. Grade 8 mildly mentally retarded students by accommodation and subject: Mutually exclusive combinations of accommodations.

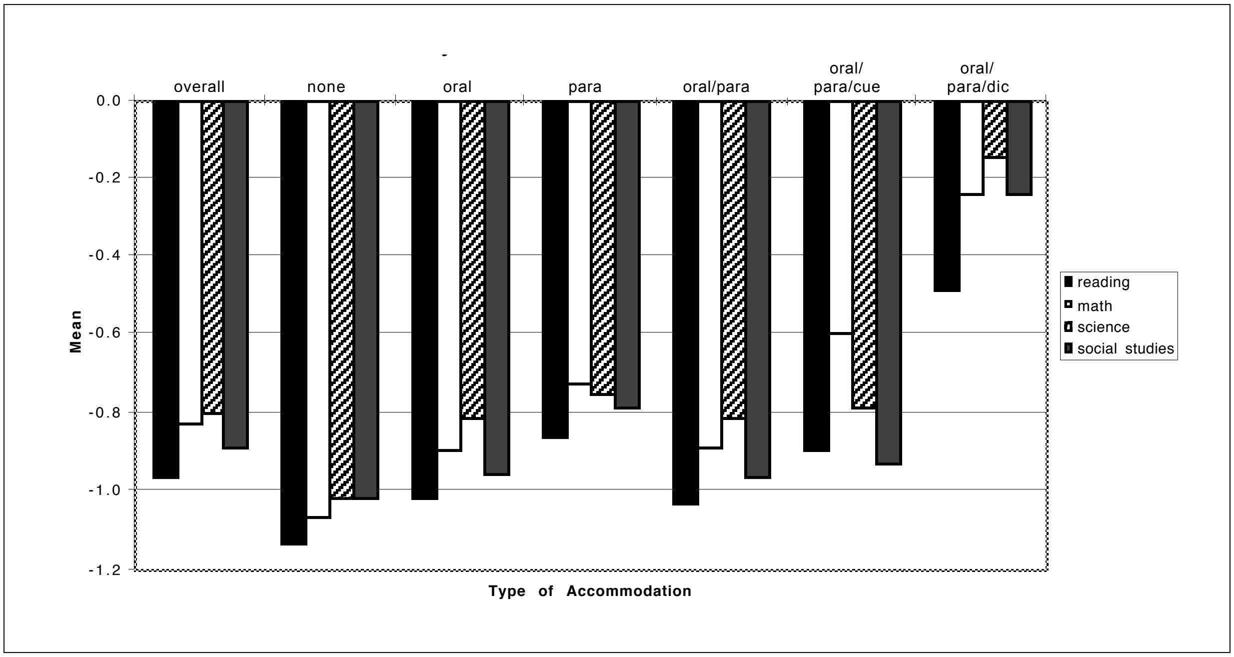


Figure 4. Grade 8 learning-disabled students by accommodation and subject: Mutually exclusive combinations of accommodations.

The Performance Correlates of Individual Accommodations

As noted earlier, analyzing the performance correlates of individual accommodations is complicated by the fact that most students with disabilities received two or more accommodations. In most instances, the groups of students receiving only a single accommodation were small and presumably atypical, so their performance on KIRIS could be misleading. Accordingly, we used a simple multivariate model to disentangle the independent relationships between individual accommodations and performance on KIRIS for the two largest disability groups, students with learning disabilities and students with mild mental retardation. More limited analyses were conducted for students with emotional or behavioral disturbances.

For each grade and subject area and for each of the two disability groups, we estimated two simple regression models in which scores were predicted by a constant and dummy variables indicating the presence or absence of the most common accommodations. The first model included only the three most frequent accommodations: oral presentation, paraphrasing, and dictation (see Table 6). The second model added a variable for cueing, which was much less common than the other three accommodations but was still provided to 10% or more of the tested students with disabilities in each grade. In these models, the constant estimates the score that students would have received in the absence of accommodations, while the coefficient for each dummy variable estimates the change in scores that would be obtained by adding that particular accommodation. To test the extent to which these simple models fit the actual KIRIS results, we predicted scores using the regression models for each of the most common mutually exclusive categories of accommodations and compared these predictions to the actual scores obtained by students in those groups.

The results presented here are limited to the more inclusive model that includes cueing. In some but not all cases, the coefficients for cueing were sizable and statistically significant. In addition, the models including cueing generally fit the observed data better than those that excluded it. The results of all of the models including cueing for these three disability groups are presented in Appendix A.

In general, the results of these models fit observed performance on KIRIS quite well, although there was one clear exception. The results in mathematics for eighth-grade students with learning disabilities illustrate the typically good fit provided by the models (Table 14). For all six of the mutually exclusive combinations of accommodations with substantial numbers of students, the predicted mean score was within 0.08 of the observed mean. (Recall that the scale for these scores has a standard deviation of 1 in the nondisabled population and generally similar variability among students with disabilities, so a difference of 0.08 is very small.)

Table 14

Observed and Predicted Scores in Mathematics, Grade 8 Learning-Disabled Students, by Accommodations

Mutually exclusive accommodations	Observed mean score	Predicted mean score	Number of students
No accommodations	-1.06	-1.00	697
Oral presentation	-0.90	-0.98	290
Paraphrasing	-0.73	-0.88	147
Oral presentation and paraphrasing	-0.89	-0.86	414
Oral presentation, paraphrasing, and cueing	-0.60	-0.59	138
Oral presentation, paraphrasing, and dictation	-0.24	-0.19	89

The primary exception to the close fits obtained by the models arose among learning-disabled students in the fourth grade. For one group of these students—those receiving only paraphrasing—the models substantially overestimated performance. In mathematics, for example, those students had a mean score of -0.70 (0.7 standard deviation below the mean for nondisabled students), but the model predicted an average of only about 0.3 standard deviation below the mean (Table 15). The models substantially overestimated the performance of this particular group of students in the other subject areas as well. Although the number of students receiving only paraphrasing was small, these differences appear to reflect more than sampling error.[8]

When the correlates of individual accommodations were disentangled, dictation had the strongest relationship to KIRIS performance for learning-disabled students. This was most striking in Grade 11. In that grade, learning-disabled students with no accommodations were predicted to score, on average, 1.23 standard deviations below the mean for students without disabilities (the value of 1.23 in the “constant” column for Grade 11 in Table 16). Providing dictation was associated with an increase of 0.8 standard deviation, leading to a predicted score for students with dictation alone of -0.43 (-1.23 + 0.80). In contrast, providing either paraphrasing or cueing was associated with an increase in scores of about one third of a standard deviation, and providing oral presentation was associated with an increase of only 0.12 standard deviation. The

Table 15

Observed and Predicted Scores in Mathematics, Grade 4 Learning-Disabled Students, by Accommodations

Mutually exclusive accommodations	Observed mean score	Predicted mean score	Number of students
No accommodations	-0.42	-0.52	164
Oral presentation	-0.47	-0.46	256
Paraphrasing	-0.70	-0.29	69
Oral presentation and paraphrasing	-0.25	-0.24	281
Oral presentation and dictation	-0.07	-0.03	398
Oral presentation, paraphrasing, and dictation	0.18	0.19	512

[8] The standard deviation of mathematics scores for learning-disabled students receiving only paraphrasing was 1.072. Thus, a confidence band of two standard errors around the observed mean for this group would extend from -0.44 to -0.96.

Table 16

Estimated Independent Effects of Individual Accommodations, for Learning-Disabled Students in Mathematics, by Grade

	Constant (no accommodation)	Paraphrasing	Oral presentation	Dictation	Cueing
Grade 4	-0.52	0.23	0.05	0.43	0.35
Grade 8	-1.00	0.12	0.02	0.67	0.27
Grade 11	-1.23	0.31	0.12	0.80	0.32

pattern in Grade 8 was largely similar. Dictation was associated with a smaller increase in scores in Grade 4 than in the secondary grades, but even there it had a stronger association with scores than the other three accommodations.

The strong positive association of dictation with the scores of learning-disabled students was consistent across subject areas as well as grade levels (Figure 5). In only one of the twelve combinations of grades and subjects was dictation associated with an increase of less than half a standard deviation in scores—the case of mathematics in Grade 4 shown in Table 16. In all other instances, dictation was associated with increases in scores ranging from 0.5 to 0.8 standard deviation. Across grades and subjects, all of the other three accommodations have a much weaker association with scores. Cueing had a moderately strong association with scores, but primarily in the fourth grade. Even though a majority of learning-disabled students have reading disabilities, oral presentation of the assessment did not have a consistent, appreciable association with scores.

The associations of accommodations with performance on KIRIS were somewhat similar among mildly mentally retarded students, although the patterns are more complex and vary more among grades and subjects. For example, in reading in Grades 4 and 11, dictation was associated with very large increases in scores—far larger than the increases associated with the other accommodations (Table 17). In Grade 4, dictation was associated with an increase in scores of fully 0.9 standard deviation, offsetting over two thirds of the difference between nondisabled students and mildly retarded students receiving no accommodation. In Grade 8, however, dictation had only a weak association with scores, and in Grade 11, cueing had quite a large association.

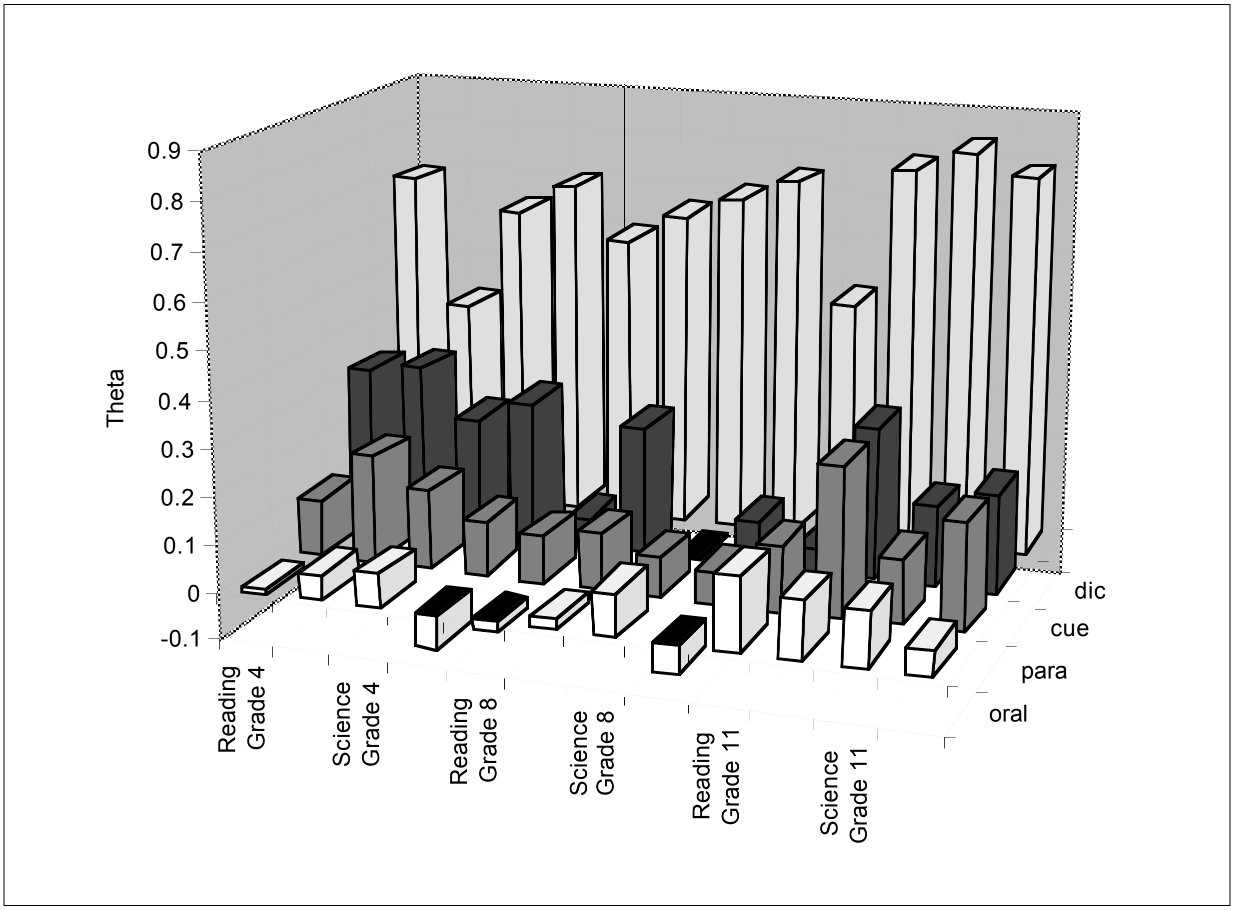


Figure 5. Learning-disabled student test scores, by grade, subject, and accommodation: Estimated effects of individual accommodation.

Table 17

Independent Effects of Individual Accommodations, for Mildly Mentally Retarded Students in Reading, by Grade

	Constant (no accommodation)	Paraphrasing	Oral presentation	Dictation	Cueing
Grade 4	-1.31	0.21	0.18	0.90	0.25
Grade 8	-1.61	0.21	0.13	0.17	0.23
Grade 11	-1.90	-0.03	0.12	0.82	0.57

The inconsistency across grades in the relationships between accommodations and KIRIS performance for mildly retarded students was more apparent when all four subjects are viewed together. The association between dictation and scores was strongly positive in Grades 4 and 11, although markedly less so in fourth-grade mathematics (Figure 6). In contrast, the relationship between dictation and scores was much more modest in all subjects in Grade 8. Cueing showed only modest relationships in Grades 4 and 8 but was strongly related to scores in Grade 11. The relationships between performance and both

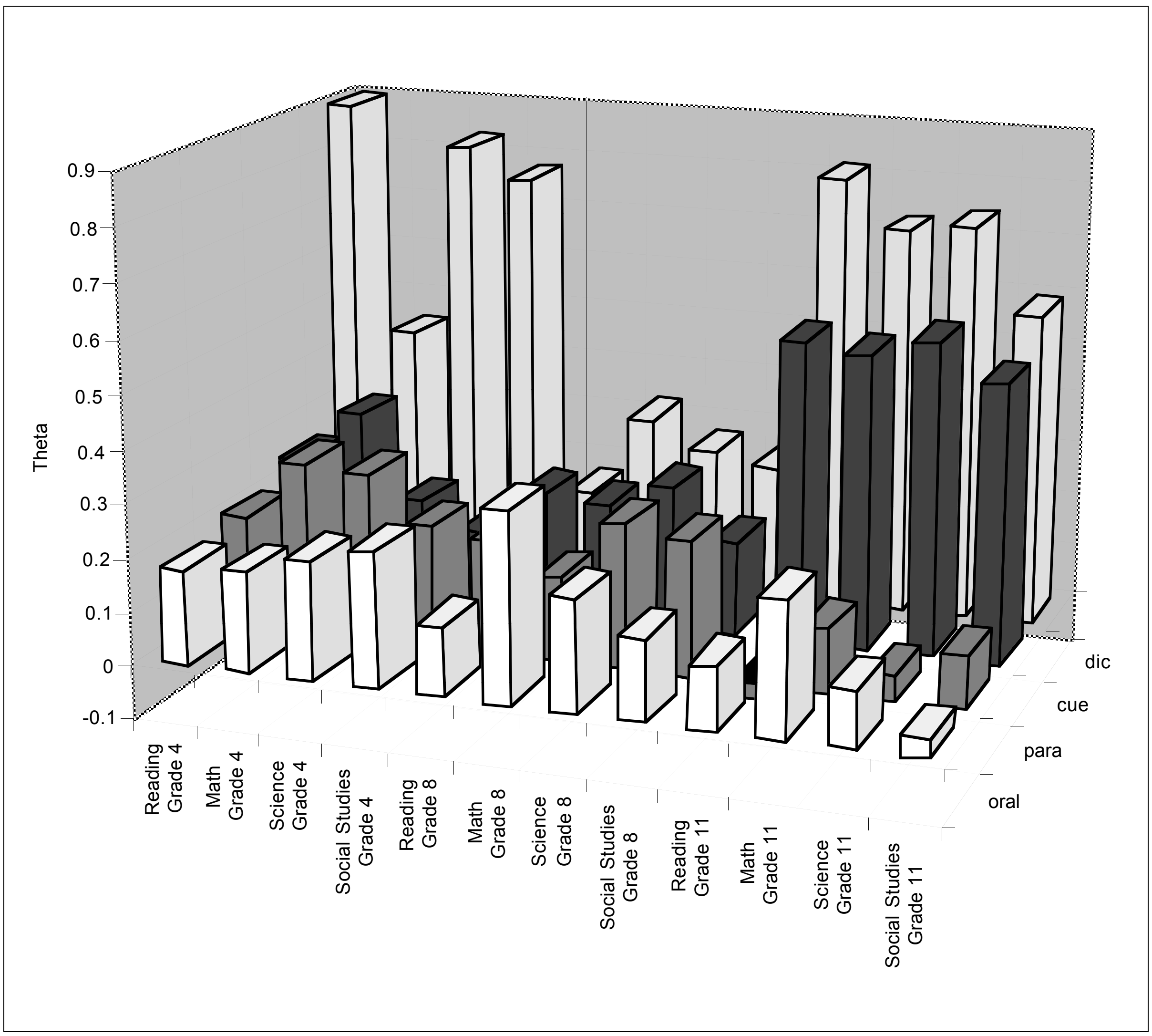


Figure 6. Mentally retarded student test scores, by grade, subject, and accommodation: Estimated effects of individual accommodation.

paraphrasing and oral presentation were less variable across grades. In most cases, their relationships with scores were small to moderate, and they were smaller in Grade 11 than in the other two grades.

These analyses of individual accommodations appear to explain the results reported earlier for mutually exclusive categories of accommodations, including the very high scores received by students receiving certain combinations of accommodations. They indicate that providing students with the opportunity to dictate responses accounts for much of the performance gain of students receiving multiple accommodations. For example, in science, mildly retarded fourth-grade students who received the combination of oral presentation, paraphrasing, and dictation scored 0.11 standard deviation above the mean for students without disabilities (Figure 1). These analyses of individual accommodations accurately predict that result. They predict that these students would get a boost of about 0.8 standard deviation from dictation, 0.3 standard deviation from paraphrasing, and 0.2 from oral presentation, combining to provide a predicted mean score 0.15 standard deviation above the mean for students without disabilities—almost exactly their observed mean score. Similarly, for fourth-grade learning-disabled students receiving these three accommodations, the model predicts a gain in science of about 0.6 standard deviation from dictation, 0.2 from paraphrasing, and 0.1 from oral presentation, combining to give a predicted mean score 0.57 standard deviation above the mean for students without disabilities. Their observed mean score was nearly exactly that: 0.54 standard deviation above the mean of students without disabilities.

Accommodations provided to fourth graders with emotional or behavioral disturbances showed somewhat similar associations with performance on KIRIS, although the small number of students receiving certain accommodations makes these findings more uncertain. Omitting the cueing variable from the analysis (because only 32 students with this disability received cueing, either alone or in combination with other accommodations), dictation was again associated with large increases in scores, ranging from 0.7 to 1.0 standard deviation.[9] In this group, paraphrasing and oral presentation were associated with small changes in scores that were neither consistent in direction nor statistically significant. In the

---

[9] With cueing added to the analysis, the estimates for dictation were slightly larger. Cueing was associated with a statistically nonsignificant increase of about 0.2 standard deviation in all subjects other than mathematics. In mathematics, cueing was associated with a significant increase of 0.6 standard deviation.

case of eighth-grade students with emotional or behavioral disturbances, dictation was associated with more modest increases in scores, ranging from 0.2 to 0.4 standard deviation, but only 33 of these 413 students received dictation, either alone or in combination with other accommodations, so these estimates are unreliable. The models for students with emotional or behavioral disturbances could not be verified using observed scores because too few of these students fall into most mutually exclusive categories of accommodations.

Item-Level Assessment Results for Students With Disabilities

Important indications of the quality of KIRIS assessment results for students with disabilities can be found in the responses of students to individual test items. Patterns that appear consistently among many of the assessment's items provide evidence about the quality of measurement for the groups in question. Patterns that appear only for a modest number of items may have relatively little effect on overall measurement quality but may nonetheless provide useful hints about the functioning of the assessment. For example, evidence of bias that appears only for a few items may suggest types of items that are potentially problematic for these students.

Three types of item analysis were conducted using KIRIS data. *Simple descriptive statistics* include means on each open-response item, the percentage of students leaving the item blank, and the percentage scoring receiving a score of zero on the 0-to-4 scale used to score these items. *Item-to-total-score correlations* are the correlations between scores on each item, on the 0-to-4 scale, and total scores for the appropriate subject area. Analyses of *differential item difficulty* explore whether students with and without disabilities who obtained similar total scores in the appropriate subject area performed similarly on specific test items.

These analyses were carried out for all common items (those administered to all students, regardless of test form) in mathematics and reading in Grades 4 and 8. Because students with disabilities tested with accommodations scored on average much higher than those tested with no accommodations, these analyses were conducted separately for disabled students tested with and without accommodations.

Taken together, these item-level analyses reveal a complex and mixed view. Some of the results show no indication of weakness in the results for students with disabilities, while others provide grounds for concern.

Item-Level Descriptive Statistics

For the sake of simplicity, many of the findings pertaining to item difficulty, the percentage of students obtaining zero scores, and the percentage of students leaving items blank are presented here only as the means of the results, averaged across all common items in a grade and subject. The results for all individual items are provided in Appendix B.

As one would expect, the common items were more difficult for students with disabilities than for other students. In reading in both grades, students with disabilities who were tested without accommodations scored on average more than half a score point lower (on the 0-to-4 scale) than students without disabilities (Table 18). These differences were larger in mathematics—particularly in Grade 8, in which disabled students without accommodations scored on average a full point lower than students without disabilities.

The pattern of relative item difficulties is more complex in the case of students with disabilities who were tested with accommodations. As one would expect from the analyses of total scores reported above, the common items were easier for disabled students who had received accommodations than for those who had not. This difference was smaller in Grade 8 than in Grade 4—particularly in the case of reading, where the difference between eighth-grade disabled students with and without accommodations was a mere 0.1 scale point. The smaller difference in eighth grade between students with disabilities with and without accommodations may reflect in part the mix of accommodations offered in each

Table 18

Mean Scores on Common Reading Items, by Disability Status and Accommodations, Grade 4 (Means of Item-Level Means)

	Grade 4 reading	Grade 8 reading	Grade 4 math	Grade 8 math
No disability	2.7	2.5	2.4	2.0
Students with disabilities, no accommodations	2.1	1.8	1.7	1.0
Students with disabilities, with accommodations	2.5	1.9	2.0	1.2

grade. In the fourth grade, 61% of students offered accommodations were permitted dictation, which had by far the largest positive effects on scores. In eighth grade, only 21% of students with accommodations were provided with dictation.

Perhaps more important than mean differences on these items are differences in the percentage of students either omitting the item or obtaining a score of 0, which is defined in the Kentucky General Scoring Guide as an answer that "is completely wrong or has nothing to do with the question." Large percentages of zeros or omits would suggest that the items have little measurement value for many students with disabilities.

In reading in both Grades 4 and 8, scores of zero were rare among students without disabilities and considerably more common among students with disabilities assessed without accommodations (Table 19). Among disabled students assessed without accommodations, the percentage scoring zero on reading items ranged from 4% to 21% (Appendix B, Tables B5 and B6). Students with disabilities assessed with accommodations had fewer zeros than unaccommodated students but more than students without disabilities.

In mathematics, zero scores were much more common in all groups, and the performance of students with disabilities was strikingly worse (Table 19). Averaging across the common items, about one fourth of students without disabilities received scores of zero. Of students with disabilities assessed without accommodations, nearly half of the fourth graders and over half of the eighth graders received scores of zero. Among these students, only two mathematics items showed fewer than 25% receiving zeros: one eighth-grade item on which 13% obtained a zero, and one fourth-grade item on which 22% scored zero. From 30% to

Table 19

Mean Percent of Students Receiving a Score of Zero on Common Items, by Disability Status and Accommodations

	Grade 4 reading	Grade 8 reading	Grade 4 math	Grade 8 math
No disability	5	2	25	24
Students with disabilities, no accommodations	15	12	46	57
Students with disabilities, with accommodations	8	9	43	50

73% received scores of zero on the other mathematics items. Three of the six Grade 8 mathematics items showed zero scores for half or more of disabled students assessed without accommodations (Appendix B, Tables B7 and B8).

Students with disabilities who received accommodations did marginally better on mathematics items, but even among those students, half of the eighth graders and over 40% of the fourth graders received scores of zero. The most extreme cases were two items in Grade 8 on which 59% and 65% of disabled students assessed with accommodations obtained zero scores (Appendix B, Tables B7 and B8).

Few nondisabled students omitted any of the common items. The highest omit rate for nondisabled students was 1.8% (for an eighth-grade mathematics item; see Appendix B), and the mean across items ranged from about half a percent in reading to a bit over 1% in mathematics (Table 20). Omit rates for disabled students assessed with accommodations were only moderately higher, but those for disabled students assessed without accommodations were substantially higher, particularly in Grade 8 mathematics. Even in Grade 8 mathematics, however, the percentage of students omitting items was on average modest, and the effects of these omissions on the aggregate scores used in KIRIS would presumably be small.

Item-to-Total-Score Correlations

Correlations between performance on a given item and a test in its entirety are commonly examined as an indicator of item discrimination—that is, the degree to which performance on an item differentiates between high- and low-scoring students. A very low item-to-total-score correlation may indicate that the item is

Table 20

Mean Percent of Students Leaving Common Items Blank, by Disability Status and Accommodations

	Grade 4 reading	Grade 8 reading	Grade 4 math	Grade 8 math
No disability	0.4	1.1	0.5	1.3
Students with disabilities, no accommodations	2.1	4.2	2.7	6.2
Students with disabilities, with accommodations	0.4	1.7	0.7	2.1

not measuring the attribute that the test as a whole is intended to measure. Low correlations for a large number of items suggest a low level of internal consistency and cast doubt on the quality of the assessment as a whole.

In this case, however, the primary question is not the overall quality of KIRIS or of particular items, but rather how well the items and the assessment as a whole function for students with disabilities, in comparison to students without disabilities. Comparing item-to-whole-test correlations for students with and without disabilities can pinpoint individual items that are not differentiating as well for students with disabilities and can help judge the overall quality of their assessment results. For example, Anderson et al. (n.d.) found that in the 1995 NAEP field test, the correlations between items and total scores were often lower for students with disabilities than for other students, indicating that the assessment was less discriminating for students with disabilities. If these correlations are low for many of the tests' items, the quality of measurement of the test as a whole is called into question for those students.

To explore this question, we calculated the point-polyserial correlations between scores on KIRIS common items and students' performance on KIRIS in that subject area (on the theta scale, restandardized to mean 0 and standard deviation 1 in the nondisabled population).[10] This was done separately for several combinations of disability status and accommodations: (a) nondisabled students; (b) all students with disabilities; (c) students with disabilities who received no accommodations; (d) students with disabilities who received any accommodations; and (e) students with disabilities who received dictation, alone or in combination with any other accommodation. In addition, these analyses were repeated separately for students with learning disabilities. Dictation was singled out for this analysis because it had the strongest relationship to performance. Accordingly, if the use of accommodations were to alter the relationship between performance on items and the test as a whole, one might expect that effect to be particularly pronounced in the case of accommodations that included dictation.

---

[10] The point-polyserial correlation is analogous to the conventional point-biserial correlation, but with one variable polytomous rather than dichotomous. That is, it is the sample correlation between a continuous variable (total performance on KIRIS for a given subject area, on a theta scale) and a polytomous variable (performance on a KIRIS item, on a 0-4 scale). In contrast, the analog of the biserial correlation is the polyserial correlation, which estimates the correlation between the continuous variable and a latent normal variable assumed to underlie the polytomous variable (Drasgow, 1985). Polyserial correlations were not estimated because they are burdensome to calculate and the results presented here suggest that they would not be informative in this case.

These correlations were all quite high and showed essentially no variation across disability groups or accommodations in either grade or subject.[11] There was no evidence that any of the common items in mathematics or reading in Grades 4 or 8 were less discriminating for students with disabilities. Moreover, the results were similar for the very heterogeneous group of all students with disabilities and for the more homogenous group of students classified as learning disabled. The use of accommodations in general, or of those involving dictation in particular, had no effect on these relationships. This is illustrated in Table 21 with Grade 4 mathematics. (Grade 4 mathematics was chosen for illustration because it is the case in which poor item performance might be most likely and the consistency of correlations is therefore particularly noteworthy. The impact of accommodations was also particularly pronounced in Grade 4, and other evidence of item bias for students with disabilities was more common in mathematics than reading.) The correlations for any one of the five common items in mathematics, arrayed in the columns of Table 21, are all very similar.

Table 21

Correlations Between Common-Item Performance and Total Score (theta), Grade 4 Mathematics, by Disability and Accommodation

	Item 14	Item 17	Item 35	Item 41	Item 44
Nondisabled	.73	.63	.73	.65	.70
All students with disabilities (SWD)	.77	.63	.78	.70	.75
SWD, no accommodations	.76	.65	.76	.71	.72
SWD, any accommodations	.77	.63	.79	.69	.75
SWD, dictation	.76	.64	.79	.69	.74
All learning-disabled students (LD)	.76	.61	.77	.68	.74
LD, no accommodations	.76	.63	.71	.67	.75
LD, any accommodations	.76	.61	.78	.68	.73
LD, dictation	.76	.62	.78	.68	.73

[11] The correlations obtained in this manner are biased upward because performance on each item contributes to the total test score to which it is being compared. That is, because the total test score (theta) is based in part on each item with which it is compared, the observed point-polyserial correlation will be positive even when the correlation between that item and the sum of the other items on the test is zero. KIRIS includes relatively few items per subject area, which makes this confounding greater than it is in the case of traditional tests that include a larger number of items. However, this confounding is not critical here, where the issue is not the overall size of each correlation, but rather how correlations compare across groups.

Differential Item Difficulty

An important concern in evaluating assessments for groups that may perform differently from the norm is the possibility of differential item difficulty—that is, items that are either harder or easier for a specific group than one would predict on the basis of their performance on the assessment as a whole. Differential difficulty may be a sign of item bias. For example, some test-takers with visual impairments maintain that on a timed test, items that entail long reading passages are biased against students who use Braille because Braille cannot be skimmed (Willingham et al., 1988). Differential difficulty can also reflect other factors, however, such as differences in curriculum exposure, and whether it reflects bias in such cases is a matter of interpretation. Differential difficulty is often called *differential item functioning*, or *DIF*.

Because students receiving testing accommodations performed markedly differently in terms of total scores from students with disabilities who received no accommodations, DIF was examined separately for students with disabilities with and without accommodations. DIF was evaluated using the logistic discriminant function analysis (DFA) technique of Miller and Spray (1993), which tests whether scores on a particular item predict membership in the target group (students with disabilities) after holding constant differences in total scores in that subject area. This reverses the logic of most DIF methods, which test whether membership in the group predicts performance on a given item after holding constant differences in total scores. However, there is evidence that the two approaches yield similar results when both are applicable, and the DFA approach has several advantages (Miller & Spray, 1993). Most importantly, while many of the conventional methods require dichotomous items, the DFA approach easily accommodates the polytomous items (scored on a 0-to-4 scale) in KIRIS, and it can test for both uniform DIF (DIF that appears across the range of proficiency) and non-uniform DIF (DIF that appears in specific ranges of proficiency).

In order to examine DIF visually, a method commonly used with dichotomous test items was adapted to the polytomous items used in KIRIS. This conventional approach displays the probability that students will answer an item correctly as a function of their total scores on the relevant portion of the test. If the observed probabilities of a correct response increase with total scores in the pattern predicted by the model used to scale the test, the item is considered to fit the model. If the item is differentially difficult for the target group, the curve of observed probabilities will be offset to the right of the theoretical curve; if it is differentially easy, it will be offset to the left.

KIRIS items are scored on a 5-point scale, and collapsing this scale to a dichotomous "success/failure" scale in order to apply this method would lose information. Instead, mean scores on the common items were displayed as a function of total score in that subject area. Because the number of students with disabilities was in some cases very small, students were divided into a small number of categories of total scores, each of which spanned a wide range of performance (0.5 standard deviation) on the assessment. (Even using these broad categories, the number of students with disabilities within some groups was small, and apparent differences in performance among them should be interpreted cautiously.) Means for students within these ranges were calculated for all mathematics and reading common items for Grades 4 and 8, separately for nondisabled students and for disabled students assessed with and without accommodations. These means were plotted against total scores in that subject area (theta scores) to obtain curves showing the change in mean scores accompanying each increase of 0.5 standard deviation in scores on the KIRIS assessment.

Statistical tests of DIF. In the case of students with disabilities who received no accommodations, statistically significant indications of DIF appeared in 5 of 22 items, while a 6th item missed the criterion for statistical significance by a trivial amount.[12] That is, on these 5 or 6 items, students with disabilities performed either better or worse than their total scores on KIRIS would have predicted. Four of the 6 items were in mathematics (2 in each grade). The DIF was inconsistent in direction: 3 of the items were differentially difficult for students with disabilities, and 3 were differentially easy.

DIF was both more common and generally much larger in the case of students with disabilities who did receive testing accommodations. Across both grades and subjects, 13 of the 22 common items showed significant uniform DIF for students tested with accommodations. In all but one case, the DIF was statistically highly significant. Among this group as well, the DIF was inconsistent in direction: 7 items were differentially difficult for students with disabilities, while

---

[12] Because 22 items were examined for DIF for each group of students with disabilities (10 in Grade 4 and 12 in Grade 8), there was an appreciable chance of finding apparently significant DIF by chance alone. Accordingly, all tests of DIF were evaluating using a critical level of p = .0005 per item, which is equivalent to providing a test for the full set of 22 items at p = .01.

6 were differentially easy. DIF appeared in all four combinations of subject and Grade: 4 of 5 items in fourth-grade mathematics; 3 of 5 items in fourth-grade reading; and 3 out of 6 items in both subjects in the eighth grade. Of the 13 items subjects that showed uniform DIF, 9 showed non-uniform DIF as well—that is, particularly pronounced differential difficulty for students with disabilities whose total scores fell within a specific range.[13]

Visual inspection of the size of DIF. These statistical analyses, however, do not provide a straightforward indication of the substantive importance of these instances of DIF. Their statistical significance is sensitive to sample size; given the large number of students tested in KIRIS, some instances of DIF might be statistically significant even if they are substantively too small to be important. Furthermore, direct interpretation of the coefficients from the logistic discriminant function analysis is difficult.[14] On the other hand, the plots of item means described earlier provide an easily interpreted view of the severity of DIF.

In reading, visual inspection showed that differences in difficulty were modest for most of the items that showed statistically significant DIF. In contrast, several of the mathematics items showed marked differences in difficulty between nondisabled students and disabled students assessed with accommodations.

In reading, the most substantial instance of DIF was modest; it appeared in fourth-grade item 20 (Figure 7).[15] The item was consistently more difficult for disabled students assessed with accommodations, but the difference was not large. Averaged across the theta scale, accommodated students with disabilities scored roughly 0.2 score points lower than students without disabilities whose total scores were within the same range. While this DIF is only moderate in size, it accounts for most of the difference in performance on this item between these two groups of students. The simple mean difference on this item between students

---

[13] Statistically significant interactions between item performance and test score (indicating non-uniform DIF) were ignored when the main effect of the item (indicating overall or uniform DIF) was not statistically significant.

[14] In addition to the usual complexities inherent in interpreting coefficients from a logistic model, these models pose the additional difficulty that they estimate the probability of being disabled given patterns of performance, not the probability of a given score given disability.

[15] In Figure 7 and all comparable figures, the x-axis is labeled with the bottom of each range of scores. Thus, the points at x = 0 represent the range from zero to .5 standard deviation above the mean. The item numbers used in these analyses were assigned by RAND to give each item used in 1992, 1993, 1994, or 1995 a unique identifier and do not correspond to item numbers in materials published by the Kentucky Department of Education.

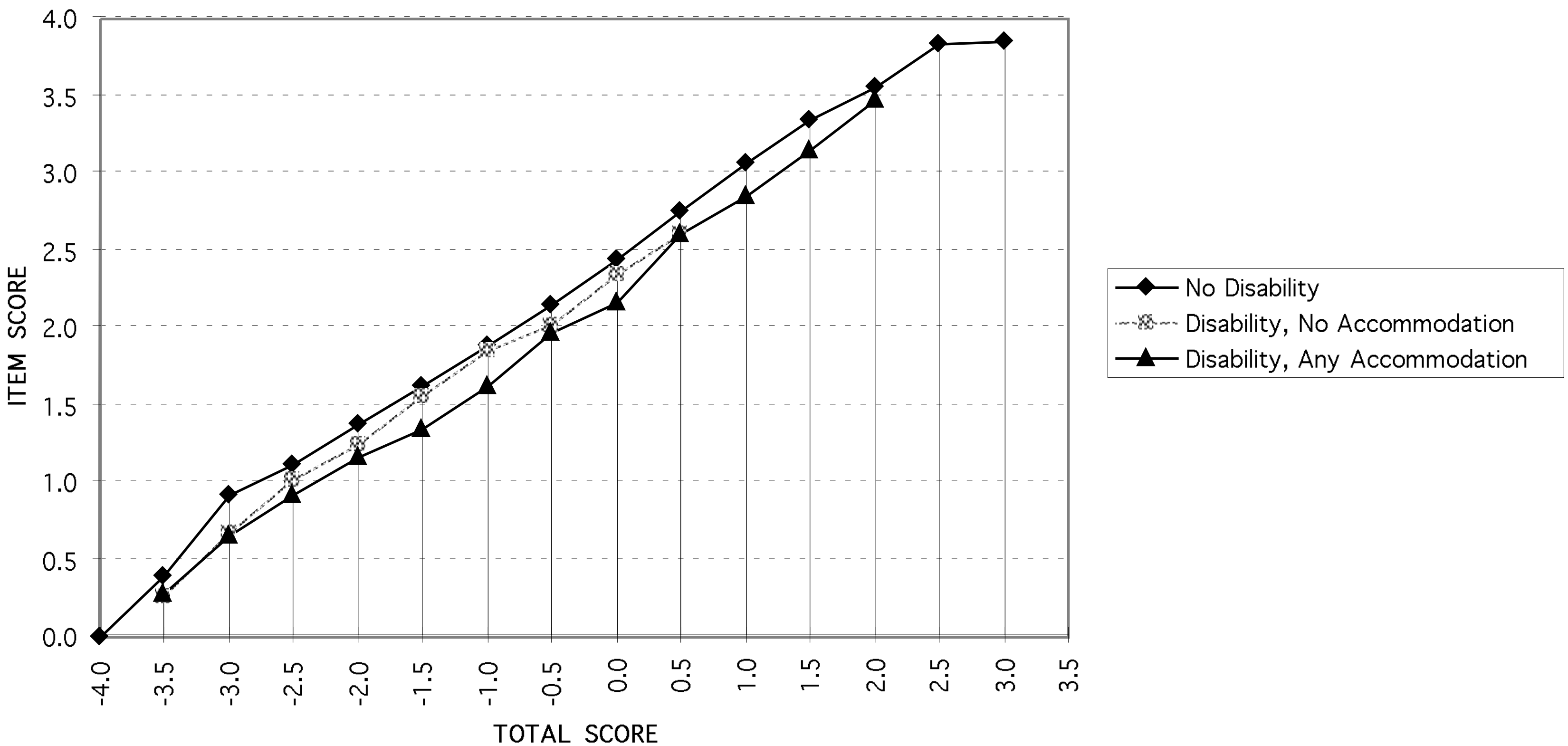


Figure 7. Mean scores of students with disabilities, by accommodation and total score, fourth-grade reading item 20.

without disabilities and disabled students with accommodations, not controlling for differences in total scores, was about 0.3 points (see Appendix B, Table B1).

In the case of all of the five other reading items that showed statistically significant DIF, visual inspection revealed very small or even trivial group differences. For example, Grade 4 reading item 44 was statistically significantly easier for disabled students with accommodations, but the differences between them and students without disabilities were trivial at almost all points in the score range (Figure 8).

Visual inspection showed little DIF in reading for students who had been assessed without accommodations. On item 20, disabled students without accommodations scored lower than nondisabled students with similar total scores, but only by a very small margin (Figure 7). Consistent with the statistical tests described above, this pattern—lesser differences in difficulty, or none whatever, for disabled students assessed without accommodations—appeared in the plots of most items that showed DIF for disabled students assessed with accommodations.

Mathematics yielded a very different picture: Many of the items showing statistically significant DIF showed sizable differences in performance when examined visually. Only one of these seven items (the one showing the statistically weakest evidence of DIF) showed an insubstantial difference between groups on visual inspection. Two showed modest differences, roughly comparable to the largest difference found in reading. The remaining four items showed considerably larger differences.

An example of a math item showing modest DIF is Grade 4 math item 41 (Figure 9). Averaged across the range of theta scores, disabled students assessed with accommodations averaged roughly 0.2 points lower on this item than students without disabilities with total scores in the same range. Overall, without controlling for differences in total scores, disabled students with accommodations scored about .4 points lower than nondisabled students on this item. These results suggest that roughly half of the difference between these groups on this item was attributable to DIF.

An example of an item that showed larger DIF is Grade 4 mathematics item 17, which was considerably more difficult for disabled students receiving accommodations than for nondisabled students (Figure 10). The difference on this

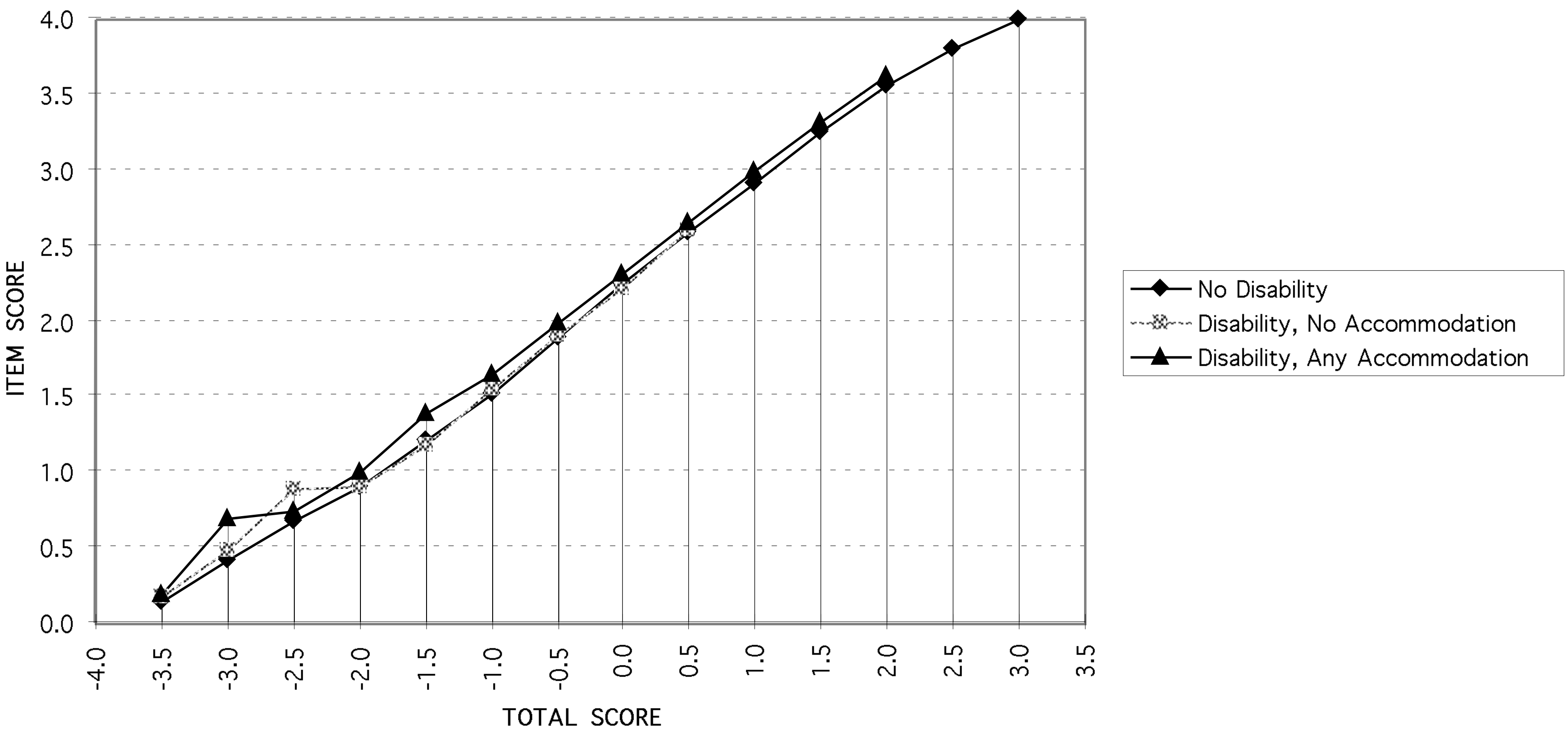


Figure 8. Mean scores of students with disabilities, by accommodation and total score, fourth-grade reading item 44.

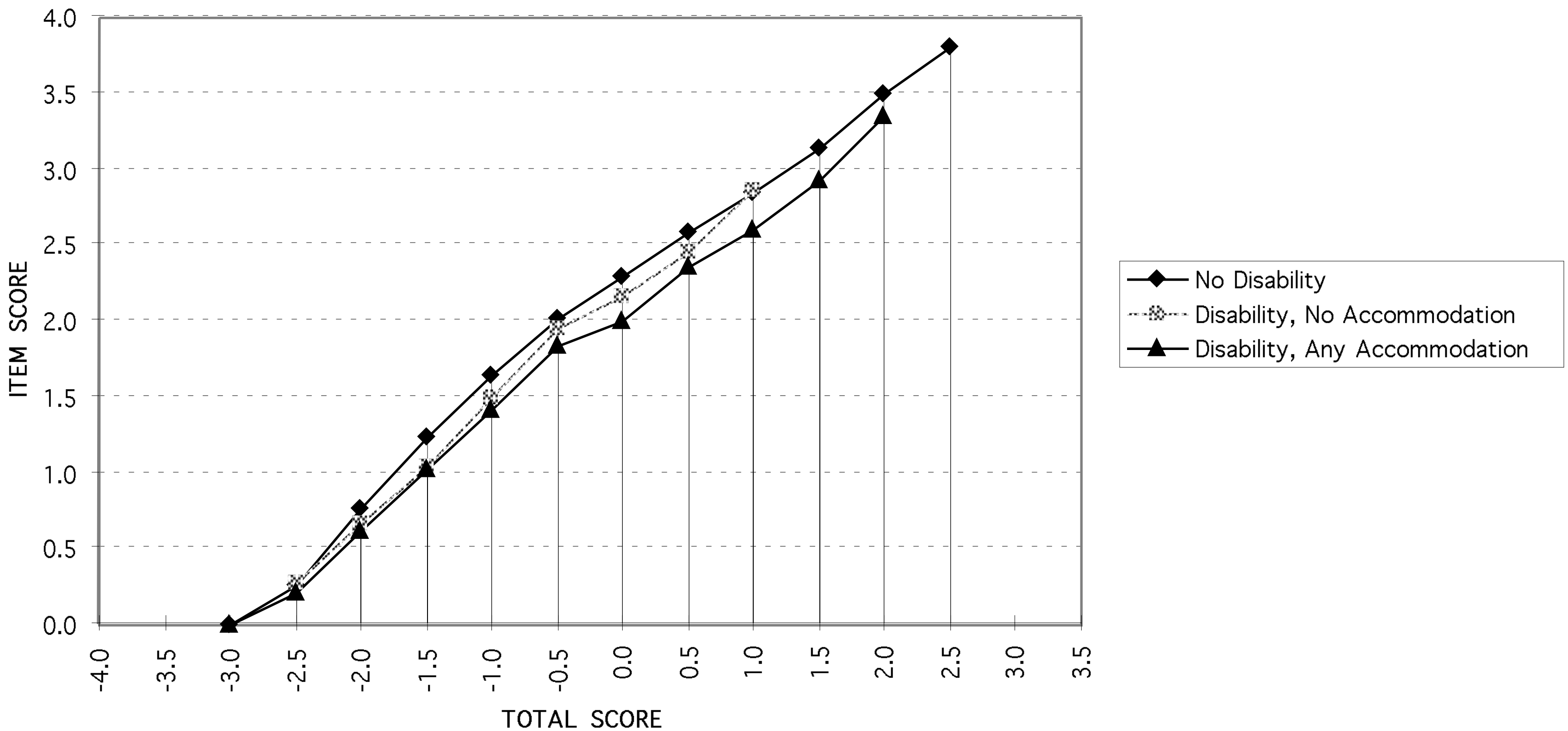


Figure 9. Mean scores of students with disabilities, by accommodation and total score, fourth-grade mathematics item 41.

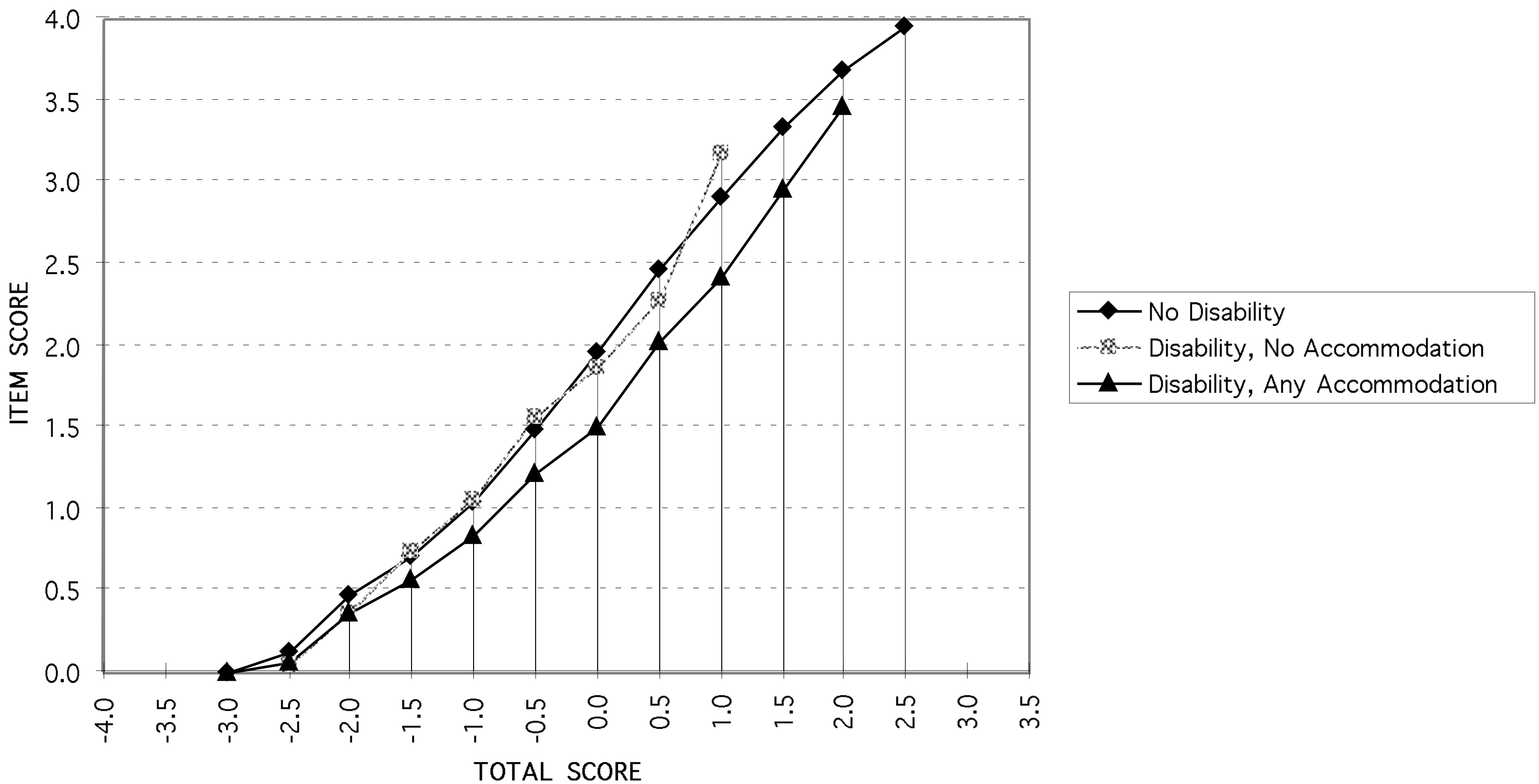


Figure 10. Mean scores of students with disabilities, by accommodation and total score, fourth-grade mathematics item 17.

item between these two groups was small for students scoring well below average in mathematics. This presumably reflects a floor effect, as about a third of the nondisabled students and about half of the disabled students tested with accommodations received scores of zero on this item (Appendix B, Table B7). For students with mathematics scores above .5 standard deviation below the mean, however, differences in performance on this item were sizable. For students in a range from the average to about one and a half standard deviations above the average, disabled students with accommodations scored about a half point lower than nondisabled students on this item. In contrast, the performance of disabled students receiving no accommodations showed little DIF; their performance mirrored that of nondisabled students very closely.

Equally striking DIF, but in the opposite direction, was shown by Grade 8 mathematics item 33 (Figure 11). This item also showed little difference between the groups for students with low total scores in mathematics, again because of a floor effect: 41% of nondisabled students and fully 59% of disabled students scored zero on this item (Appendix B, Table B8). However, the difference gradually widened as students' scores exceeded 1 standard deviation below the mean in mathematics. For students in the broad middle range of total scores, disabled students with accommodations averaged half a point *higher* on this item than did students without disabilities. Among those with higher total scores, disabled students with accommodations outscored nondisabled students by more than .75 point on this item. This is an example of non-uniform DIF: DIF became larger as total scores increased, even apart from the impact of large numbers of zero scores. The corresponding results for disabled students without accommodations showed no DIF at most levels and more modest DIF at the highest score intervals for that group, but this limited evidence of DIF is of limited value. Fully 72% of disabled students assessed with no accommodations scored zero on this item (Appendix B, Table B8), and the two intervals showing DIF for them in Figure 11 together included only 144 disabled students.

Because total scores differed markedly depending on the specific types of accommodations students were provided, it would be potentially useful to explore the extent of DIF as a function of specific accommodations. For example, because dictation generally had the strongest relationship to total scores, it could be useful to explore DIF specifically for students with specific combinations of accommodations that include dictation. Plots of item-level performance were

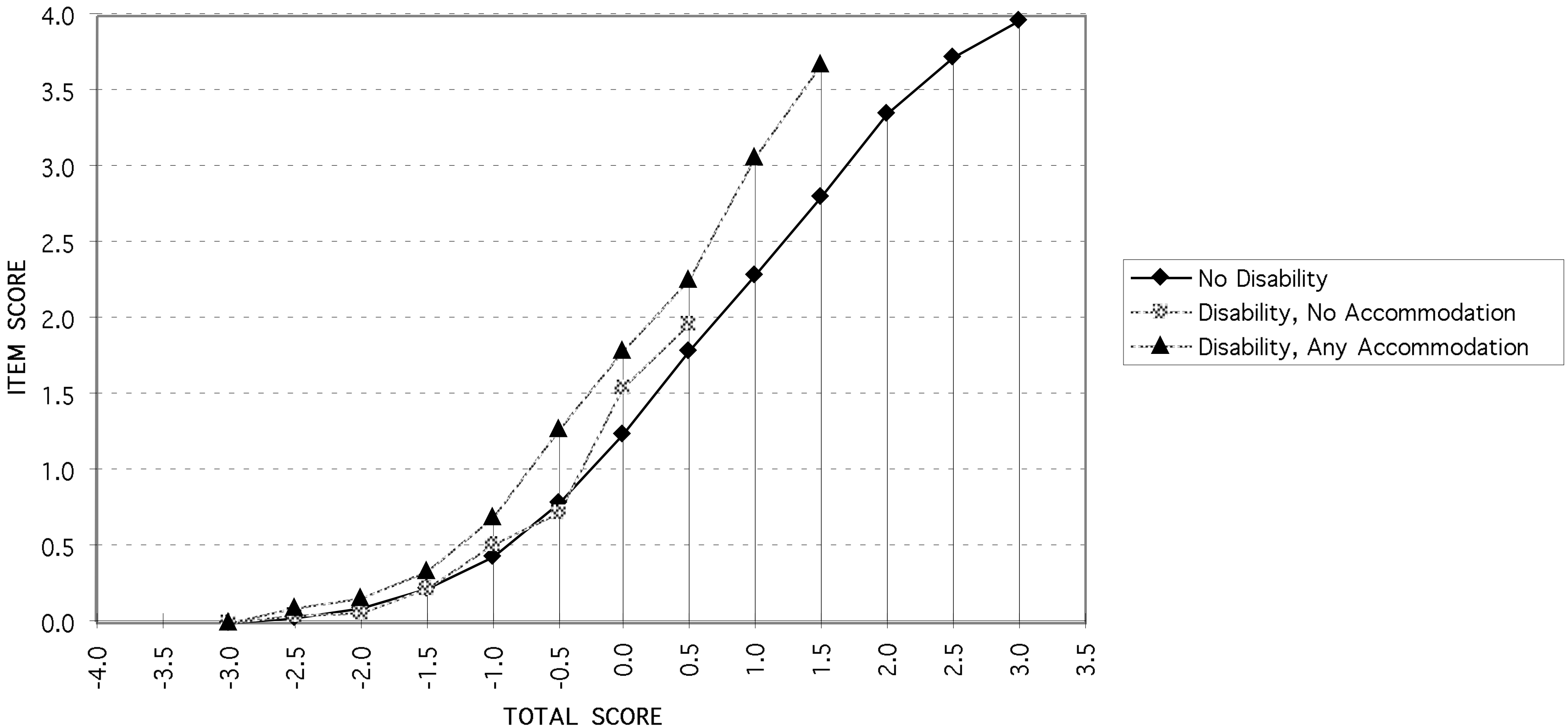


Figure 11. Mean scores of students with disabilities, by accommodation and total score, eighth-grade mathematics item 33.

created for specific accommodations, using both simple and mutually exclusive categorizations of accommodations. When students with disabilities are classified simultaneously by accommodations and total scores, however, the resulting groups are often very small, and the plots for these groups are quite unstable. No consistent difference in DIF among accommodations group was apparent, but it is possible that larger samples—for example, results from larger states—would should some clearer relationships between specific accommodations and DIF.

Interpreting DIF for students with disabilities. At first glance, some of these results seem counterintuitive: Why would some items be differentially *easy* for students with disabilities? A simple explanation seems plausible, however, when one considers that in almost all instances, appreciable DIF appeared only for students who had received accommodations.

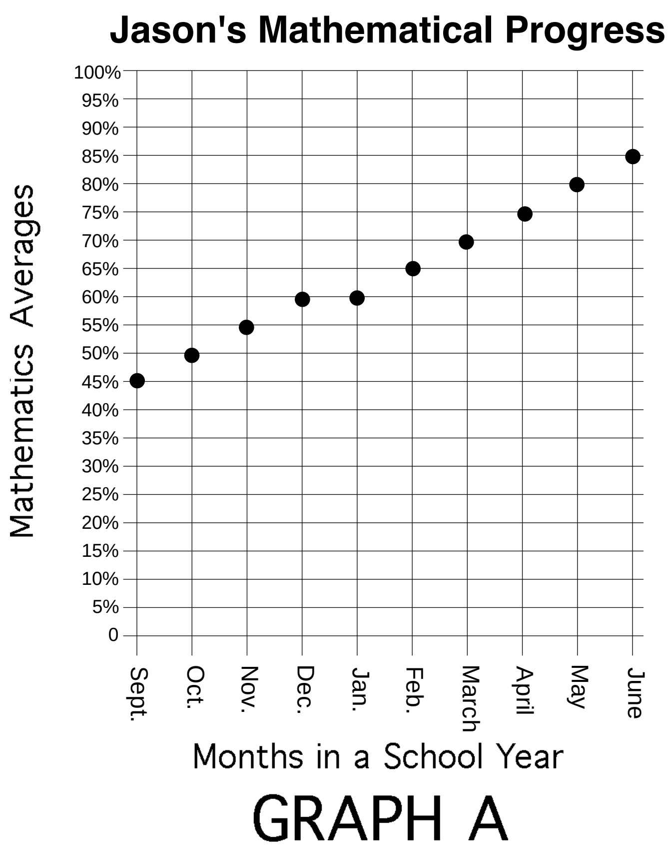
A possible explanation for these findings is that the direction of DIF may reflect the relative impact of accommodations on performance on a given item. If accommodations had relatively little impact on a given item, the item would appear to be differentially difficult for students with accommodations, because performance would be boosted less on that item than on many of the others that contribute to total scores. Conversely, an item that is more affected by accommodations than most would appear differentially easy.

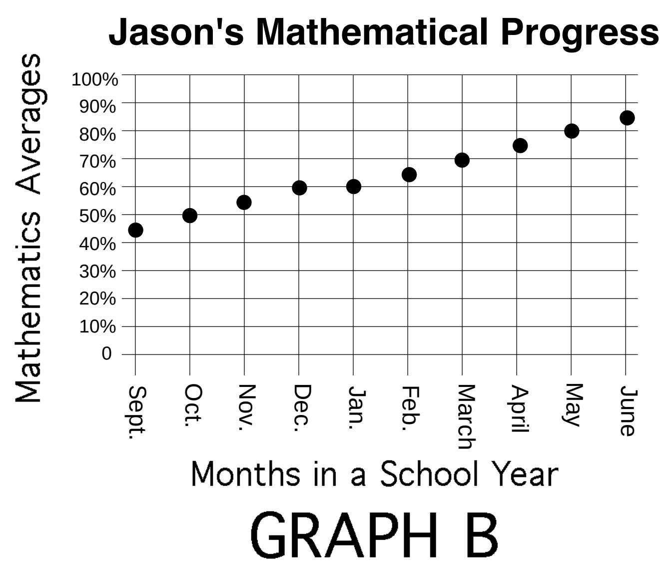
If this explanation is correct, an important question would be what are the attributes of the items performance on which is affected unusually much or little by accommodations. The small number of common items used in mathematics precludes drawing firm conclusions about this, but a suggestive pattern appears when the items with DIF favoring students with disabilities are compared to those favoring students without disabilities.

Across Grades 4 and 8, four of 11 common mathematics items showed sizable DIF favoring students without disabilities. That is, those four items were harder for students with disabilities assessed with accommodations than for nondisabled students with similar total mathematics scores. The four items, three from Grade 4 and one from Grade 8, are shown in Figure 12. One entails interpreting and explaining a simple graph; one requires solving some simple arithmetic problems based on patterns displayed in tabular form; a third requires drawing and explaining inferences about probabilities based on graphical representations of relative probabilities; and the fourth requires that students

Grade 8 Item 17

Last December, Jason's parents said that if his grades in mathematics significantly improved by June, he could go to Disney World during summer vacation. Jason decided he would figure his average for each month. He tried different ways to graph the information. He graphed the information on two different graphs as shown below.





a. Does each graph show the same information? Explain your reasoning.

b. Which graph would Jason show to his parents to convince them that his mathematics grades had improved? Explain your reasoning.

c. Give several reasonable claims that Jason could use to convince his parents that his grades improved.

Figure 12. Mathematics items differentially difficult for accommodating students with disabilities.

Continued on next page.

Grade 4 Item 35

The libraries at Lincoln School and King School both charge their students fines for overdue books.

Fines at Lincoln School

Number of Days Overdue	1	2	3	4	and so on
Fine	15¢	23¢	31¢	39¢	and so on

Fines at King School

Number of Days Overdue	1	2	3	4	and so on
Fine	1¢	2¢	4¢	8¢	and so on

a. If you returned a book to Lincoln School that was 5 days overdue, how much would your fine be? Explain how you figured this out.

b. If you returned a book to King School that was 5 days overdue, how much would your fine be? Explain how you figured this out.

c. Kendra and Brian are each returning a book that is 8 days overdue. Kendra's book is from the Lincoln School library, while Brian's book is from the King School library. Who will pay the greater fine. Explain how you figured this out.

BE SURE TO LABEL YOUR RESPONSES (a), (b), AND (c).

Figure 12 (continued).

Continued on next page.

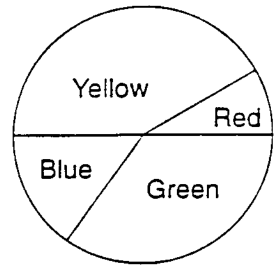
Grade 8 Item 41

3. Here are some spinner games for you to think about. For each game, the first person to get ten points wins.

GAME 1:

This is the spinner for Game 1.

Each player chooses one color and gets one point if the spinner lands on that color.

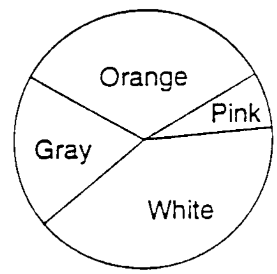


a. If you were playing Game 1, which color would you choose so that you would have the best chance of winning? Explain your answer.

GAME 2:

This is the spinner for Game 2.

Each player chooses a first color and a second color. The player gets one point if the spinner lands on the first color chosen, but loses one point if it lands on the second color chosen.

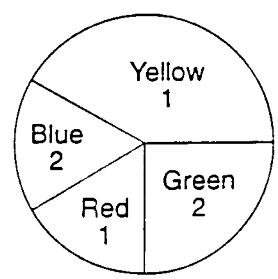


b. If you were playing Game 2, which colors would you choose as your first color and second color so that you would have the best chance of winning? Explain your answer.

GAME 3:

This is the spinner for Game 3.

Each player chooses one color. The player gets the number of points shown below the color if the spinner lands on the color he or she chose.



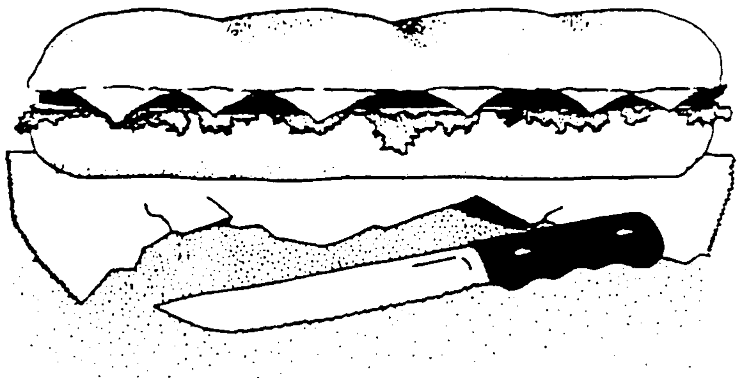
c. If you were playing Game 3, which color would you choose so that you would have the best chance of winning? Explain your answer.

BE SURE TO LABEL YOUR RESPONSES (a), (b), AND (c).

Figure 12 (continued).

Continued on next page.

Grade 4 Item 17



4. Hank got a submarine sandwich for dinner. Hank is going to share the sandwich with his brother, Bill. This is what they say.

Hank: Would you rather have $\frac{1}{3}$ or $\frac{1}{6}$ of the sandwich?

Bill: I'm really hungry, so I want $\frac{1}{6}$ because that's more than $\frac{1}{3}$.

Hank: Wait! That doesn't make sense.

Bill: Yes, it does! Anybody knows that 6 is more than 3, so $\frac{1}{6}$ of a sandwich has to be more than $\frac{1}{3}$ of it.

Tell who is correct and explain why.

Figure 12 (continued).

know and be able to explain the fact that the size of a fractional quantity is inversely related to the size of the denominator.

Three common mathematics items showed DIF favoring students with disabilities. In other words, if the hypothesized explanation is correct, accommodations improved performance on these items more than they did on mathematics items generally. Performance on two of these three items must be interpreted cautiously, because they are the items on which the largest percentage of disabled students assessed with accommodations had blank or zero scores. These three items are shown in Figure 13. Grade 4 item 44 is a fairly simple measurement task requiring the use of a ruler and a printed pattern. Grade 8 item 33 (with a combined blank or zero rate of 62% of disabled students assessed with accommodations) is a multistep problem, presented without diagrams or charts. It is arithmetically simple, but it entails a somewhat complex written array of information. Grade 8 item 37 (with a combined blank or zero rate of 67% of disabled students assessed with accommodations) is a multistep problem, presented with a brief written explanation and a diagram, that requires that students remember the relationship between the circumference and diameter of circles.

Two of the three items that were differentially easy for accommodated students with disabilities—Grade 4 item 44 and Grade 8 item 33—appear to share one characteristic: The difficulty of these items (especially Grade 8 item 33) appears to arise to an unusual degree from the required reading, and they do not require using, interpreting, or producing graphs, tables, or other nonverbal representations. For example, Grade 8 item 33 requires that one extract and interpret information from a block of 7 sentences. Once that interpretation is completed, both the arithmetic and the exposition required by the item are quite simple. Grade 8 item 37, however is an exception: Its directions are verbally simple, and it does require use of a figure.

Discussion

This study addressed two of the primary issues raised by the assessment of students with disabilities: the feasibility of including the majority of these students in statewide assessments, and the quality of performance information provided by their participation.

Grade 4 Item 44

Please answer question 5 in the space on page 9 of your Student Response Booklet.

Use the ruler you have been given to answer question 5.

5. Tim's class is making 5 posters for a pep rally using the pattern on the following page. The students will copy the pattern onto bright colored paper and cut out the posters. Then they will put fancy narrow border tape around the edges of each poster. They are having difficulty figuring out how much border tape they will need.

   Write a note to the class telling them how many INCHES of border tape they will need for the 5 POSTERS. Explain the following in your note:

   a. what measurements you made, and

   b. what you did with the measurements to come up with your answer.

   Be sure to show ALL your work.

Figure 13. Mathematics items differentially easy for accommodating students with disabilities.

Continued on next page.

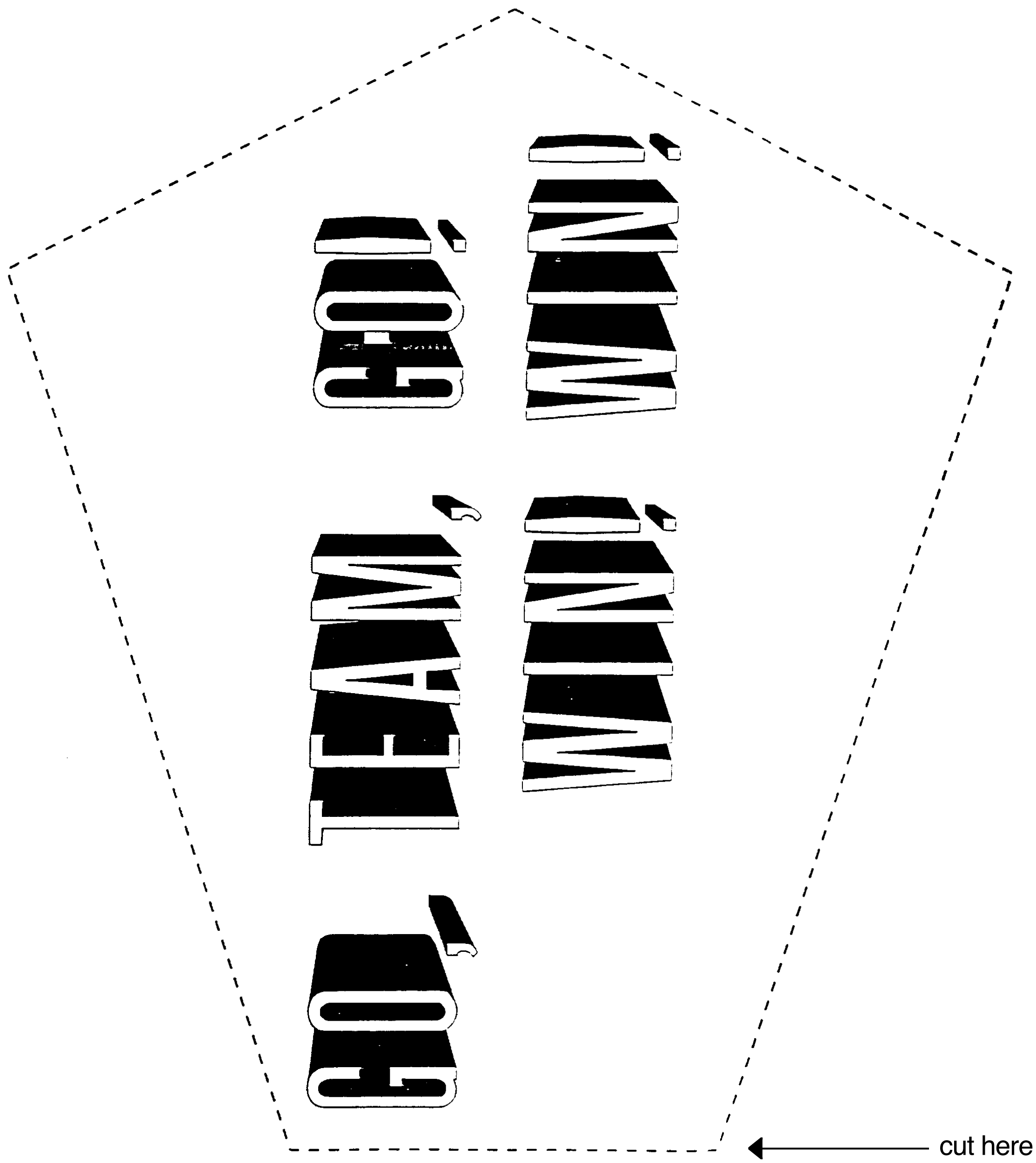


Figure 13 (continued).

Continued on next page.

Grade 8 Item 33

Please answer question 6 in the space on page 9 of your Student Response Booklet.

6. A hot new rock group is coming to Kentucky for two performances. The concerts will be held at Rupp Arena, which has a seating capacity of 24,000, and Freedom Hall, which has a seating capacity of 19,000. The group needs to make $150,000 from each concert to cover their expenses. They would like to make a total profit of at least $110,000. It is predicted that both concerts will be sold out. The cost of the tickets needs to be the same at both arenas. What would be the minimum cost of the tickets, to the nearest dollar? Show how you arrived at your answer.

Grade 8 Item 37

4. A middle school has a running track with semicircular ends as shown below. Each straight-away of the track is 100 feet long. The field that is surrounded by the track is 25 feet across.

   One mile equals 5,280 feet. In order to jog one mile, how many laps—rounded to the nearest lap—would a person have to jog? Explain your reasoning. (Assume the person jogs along the inside edge of the track.)

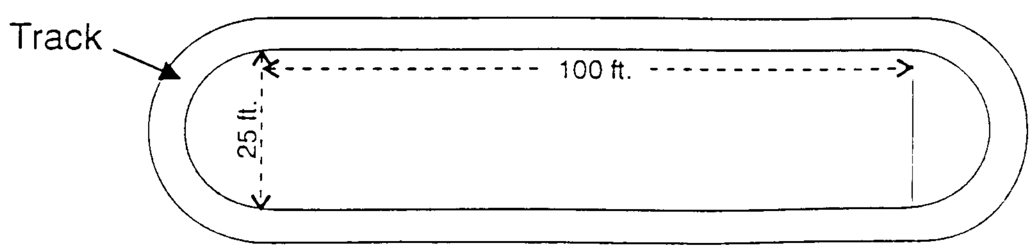


Figure 13 (continued).

The Kentucky experience suggests that it is feasible to include many students with disabilities in some large-scale assessments. This finding is consistent with the 1995 NAEP field test, which found that it was feasible to increase substantially the participation of students with disabilities in NAEP (Anderson et al., n.d.). Although precise and comparable counts are unavailable, it appears that Kentucky has at least approximately met its goal of including the large majority of identified students with disabilities in its regular assessment.

The results pertaining to the quality of measurement, however, were mixed. While some results are encouraging, others suggest that Kentucky's very high level of participation brought with it substantial problems in terms of quality of measurement. The problems that appeared were not uniform. They varied across parts of the assessment and grade levels. The problems also varied with administrative conditions: With the exception of item difficulty, the measurement problems appeared primarily among students assessed with accommodations. These measurement problems also probably vary much more among groups with different disabilities than the results here showed, because the small numbers of students in some groups made separate analysis of them impractical. Some of these measurement problems may be remedied by technical or administrative changes, but others may reflect fundamental tensions among the goals of the assessment program.

To clarify these points, it is helpful to summarize the array of good and bad news about the quality of assessment results, focusing on four areas: the assessment's level of difficulty; the use of accommodations; the apparent effects of accommodations on scores; and item-level indicators of assessment quality.

Level of Difficulty

One indicator of difficulty is the percentage of students either leaving an item blank or scoring zero on it, which is defined in the KIRIS assessment as a response that is irrelevant or totally wrong. In reading, the percentages of students with blank or zero responses were generally small to moderate. In the fourth grade, the percentage of students with disabilities assessed without accommodations with a blank or zero score ranged from 10% to 20%; in eighth grade, the corresponding percentages ranged from 7% to 27%. The corresponding percentages were much lower for students who did receive accommodations, ranging from 4% to 12% in the fourth grade and from 5% to 17% in the eighth grade.

In mathematics, however, the percentage of students scoring either blank or zero was often high, suggesting that this portion of the assessment was too difficult for many students with disabilities. Among fourth-grade students with disabilities assessed without accommodations, the percentages with blank or zero responses to the common mathematics items ranged from 25% to 49%. Among eighth-grade students with disabilities assessed without accommodations, from 19% (one item) to 80% (two items) either left the item blank or scored zero on it, and the percentage was 40% or higher for every item but one. The percentages for students with disabilities assessed with accommodations were modestly better, particularly in eighth grade, but were nonetheless very high, ranging from 18% to 67%.

The Use of Accommodations

The available data offer little opportunity to evaluate directly whether accommodations were provided either consistently with KDE's guidelines or in a manner that is likely to offset biases caused by disabilities without introducing new ones. For example, there is no information on the basis of proctors' decisions about the use of accommodations, and there is no evidence indicating the extent to which accommodations such as paraphrasing and oral presentation were used in a manner consistent with KDE's guidelines. Moreover, since KIRIS is an operational assessment, no effort was made to control the use of accommodations or to retest students under different accommodation conditions, so direct comparisons of scores from accommodated and unaccommodated administrations are clouded.

However, the sheer pervasiveness of accommodations is reason for concern. Over 80% of students with disabilities in the 4th grade and two thirds or more in the 8th and 11th grades were assessed with accommodations, and most of these were provided with two or more accommodations. This frequency appears inconsistent with the detailed and seemingly restrictive guidelines for the use of accommodations issued by the Kentucky Department of Education. For example, the guidelines present paraphrasing as an intrusive technique, and they admonish educators to use the least intrusive accommodations possible. Nonetheless, almost three fourths of the students with disabilities in the 4th grade and about half in the secondary grades were provided with paraphrasing. More important, the frequency of these accommodations raises concerns that they may be used more than measurement considerations (that is, the biases in scores caused by disabilities) would warrant and that they therefore might be biasing scores upward.

The differences between grades in the use of accommodations also raises important questions about the appropriateness of their use. The percentage of students with disabilities assessed without any accommodations increases with age (from 19% in 4th grade to 39% in the 11th grade). The rate with which certain accommodations were used was quite stable across the grades, but the use of two of them declined markedly with age. Almost three fourths of 4th-grade students with disabilities were provided with oral presentation, in contrast to a little over half in Grade 8 and under half in Grade 11. Changes in the use of dictation were far more striking and potentially more important. Half of students with disabilities in the 4th grade, but only 14% in the 8th grade and 5% in the 11th grade, were permitted to dictate responses. Given the very large association between the use of dictation and student's scores, differences between grades in the use of this accommodation presumably help to explain the finding that the gap in scores between accommodated and unaccommodated students was much larger in the 4th grade than in the secondary grades.

Taken together, these findings suggest that particularly in the fourth grade, accommodations may be used in part to offset poor reading and writing skills rather than disabilities *per se*. This would be unsurprising for the same reason it may be difficult to address: the distinction between poor reading skills and reading disabilities is ambiguous. It has long been known that the proportion of students identified as learning disabled varies markedly. In addition, classifications by school personnel are often inconsistent with research-based classifications (Shepard, 1989). Factors such as insufficient training and variations in diagnostic criteria undoubtedly contribute to these problems of classification, but recent research indicates that classification may necessarily be ambiguous. As Lyon (1966) wrote:

> The concept of LD focuses on the notion of a discrepancy between a child's academic achievement and his or her apparent capacity to learn. Recent research indicates, however, that disability in basic reading skills is primarily caused by deficits in phonological awareness, which is independent of any achievement-capacity discrepancy. . . . There is no clear demarcation between students with normal

> reading abilities and those with mild reading disability. The majority of children with reading disabilities have relatively mild reading disabilities. (pp. 54-55)

Thus, it appears unrealistic to expect school personnel to distinguish reliably between poor readers and LD students. Moreover, the lack of a clear empirical distinction between them suggests the need to rethink which test-based inferences warrant specific accommodations for students who have mild difficulties with reading.

The Apparent Effects of Accommodations on Scores

The striking associations between accommodations and test scores also raise doubts about the quality of assessment results for some students with disabilities. As noted earlier, there is one very important uncertainty in these findings: Absent additional information, it remains unclear how much the accommodations *per se* contributed to these disparities. To some degree, students offered different accommodations might have scored differently from each other even if they had been given the same accommodations. Nonetheless, the magnitude of some of the differences is grounds for concern and warrants further investigation.

One indication of potential problems is the large variation in scores among students receiving different accommodations. This appears when students are placed into mutually exclusive groups based on the accommodations they were provided, and the lowest-scoring group (always either the students with no accommodations or those offered only oral presentation) was contrasted to the highest-scoring group (always one in which students were provided with multiple accommodations). Among fourth-grade students with mild mental retardation, the difference between the lowest- and highest-scoring combinations of accommodations was a full 1.3 standard deviation, and the corresponding difference for learning-disabled fourth and eighth graders was 0.8 standard deviation.

Perhaps more telling are the comparisons between the highest-scoring groups of students with disabilities and students with no disabilities. The highest-scoring group of mildly mentally retarded students (those assessed with oral presentation, paraphrasing, and dictation) scored near the mean for nondisabled students in all subjects other than mathematics—an implausible result, given that these students have generalized cognitive deficits. Similarly, the two highest-scoring groups of fourth-grade learning-disabled students scored above the average for nondisabled students in 7 of 8 comparisons, and far above the average in science. While it is not clear what the "correct" scores are, these findings also seem implausible. That is, while one might argue that many learning-disabled students would show a normal distribution of performance if the effects of their disabilities were offset, there is no reason to expect that such students are in any real sense fully half a standard deviation above nondisabled students in science performance.[16]

The key to these differences in scores may lie in disentangling the effects of individual accommodations. In the two groups large enough for this analysis—learning-disabled and mildly retarded students—the independent associations between accommodations and scores varied markedly among the accommodations about which Kentucky recorded information. In most comparisons, dictation had by far the strongest association with scores. In some groups, cueing—which was far less frequent—also had a strong positive association with scores. Several of the accommodations commonly used elsewhere, such as extended time, were not tracked in the KIRIS data.

Item-Level Indicators of Assessment Quality

In addition to simple indicators of item difficulty (the percentages leaving items blank or scoring zero), this study used several other aspects of item-level performance to explore the quality of KIRIS results for students with disabilities. The results of these analyses were mixed.

On the positive side, correlations between item-level performance and total scores were not appreciably different for students with disabilities than for others, regardless of subject, grade level, or the use of accommodations. This finding applied to learning-disabled students as well as to the entire population of students with disabilities. This suggests that KIRIS items in mathematics and reading differentiate between high and low achievers as well for students with disabilities as for other students. This positive finding, however, is subject to one important

---

[16] Whether it would be reasonable to expect students with learning disabilities to obtain scores even comparable to those of nondisabled students is arguable and would depend on the particular inferences the assessment is used to support. For example, students with reading disabilities might be expected (if taught effectively) to have proficiency in arithmetic computation similar to that of nondisabled students, and accommodations that raised scores by removing the effect of reading disabilities might therefore lead to a more valid inference about this particular set of skills. On the other hand, some students with reading disabilities may have a lower true level of proficiency on mathematical tasks that require substantial reading or writing.

caveat: The confidence one can have in the correlations in mathematics is undermined by the very large percentages of students scoring zero or leaving the items blank. These positive findings stand in stark contrast to the results of the National Assessment pilot study of including more students with disabilities in NAEP (Anderson et al., n.d.), which found substantially lower item-score correlations for students with disabilities than for other students.

Similarly, except for groups that were too small to provide reliable estimates, the performance of students with disabilities on individual KIRIS items was always found to increase with total scores, and these increases generally paralleled quite closely those observed among students without disabilities. These findings are consistent with the correlational results. This also contrasts with the results of the NAEP field test, which found that trends in performance on individual items for students with disabilities often differed from those for nondisabled students and were even non-monotonic in many cases. (That is, the performance of students with disabilities on some NAEP items actually decreased as their total scores increased.)

Finally, instances of DIF (differential item functioning), which might indicate item bias, were both few and generally minor for students with disabilities assessed *without* accommodations.

Offsetting these positive findings from item-level analyses were findings showing frequent and often large DIF for students with disabilities who were provided with assessment accommodations. For these students, this study found statistically significant instance of DIF on more than half of the common items in reading and mathematics in Grades 4 and 8. In terms of substantive size, the substantial instances of DIF were almost all in mathematics. The small numbers of students who received certain combinations of accommodations made it impossible to determine clearly whether DIF is strongly associated with the specific accommodations students were given, but no consistent differences among accommodations were apparent.

Implications for Further Research

The shortage of research on the assessment of students with disabilities in the elementary and secondary grades remains severe. This report is one of the first studies of the assessment of a nearly representative group of students with disabilities in a large-scale elementary and secondary assessment of achievement and is the first of its kind in the context of a statewide, high-stakes assessment. Additional research is needed to test patterns found in this study and to explore the many questions it left unaddressed. The stark contrasts between the findings reported here and several of the findings of the 1995 NAEP field trial (Anderson et al., n.d.) underscore the importance of additional research.

Given the heterogeneity of the population of students with disabilities, it may prove important to focus some of the future research on specific and reasonably homogeneous subgroups, to the extent that samples are sufficiently large. Breaking the population into smaller groups may help interpret findings—as in this study, in which the patterns shown among students with mental retardation and learning disabilities helped clarify the implications of the large associations between accommodations and scores. In addition, the best ways of assessing students may vary markedly among subgroups of students with disabilities. For example, it is possible that a given accommodation may increase the validity of an assessment for visually disabled students but decrease it for many learning-disabled students. In such a case, the much larger number of students with learning disabilities would obscure the positive effects for the visually disabled students unless the research distinguished among students with different disabilities. Focusing on differences among groups runs counter to a currently a widespread effort to downplay the distinctions among disability groups because classification has been ambiguous and error-prone and because classifications are often not useful in selecting educational interventions. However, as Shepard (1989) pointed out, classification for research purposes need not mirror classification for decisions about instructional placement or services, and meaningful classifications among groups of students with disabilities are needed for some types of research even when they are not useful in making decisions about services. In some instances, it may be necessary for researchers to reclassify students rather than rely on schools' own classifications, because the classification by school personnel is inconsistent across jurisdictions and often inconsistent with research-based classifications (e.g., Shepard, 1989).

In the light of the findings reported here, several types of additional research appear particularly important and are discussed next.

Descriptive studies in additional settings. Descriptive studies similar to this one but in different settings and with different assessments are needed to explore the generalizability of the findings reported here. Results in other settings could differ for any number of reasons, including the definitions of disabilities used to classify students, the difficulty and format of the assessments, aspects of test administration (e.g., time limits), the consequences for test scores, and patterns in the use of accommodations. As more stringent and consistent rules for the exclusion of students with disabilities from large-scale assessments gradually become more common, descriptive studies of this sort will become both more feasible and more important. In the interim, descriptive studies that analyze the exclusion of students as well as the test performance of those included would be useful.

Further research on the use of accommodations. A clear and pressing need is for additional research on the use of accommodations. While extant studies show that the use of accommodations (or at least the guidelines for them) varies among jurisdictions, relatively little is known about variations in their use within an assessment program or about the factors that are associated with that variation. More information is needed about the process by which decisions about accommodations are made: whether state guidelines are known, how they are interpreted, and whether they are followed; what educators understand to be the functions of accommodations; and what student characteristics and other factors educators take into account in making decisions about accommodations.

Further research on the effects of accommodations. The shortage of research exploring the effects of accommodations on scores and on their validity is severe, and the need for additional research is pressing (e.g., National Research Council, 1997). The findings reported here are only a first step in that direction and indeed underscore the need for additional research. For example, there is a need for research examining accommodations not monitored in Kentucky, such as the provision of extra time. There is also a need for research that tests the generalizability of findings reported here and that employs methods different from those used in this study to examine the impact of accommodations.

The difficulties noted above in evaluating scores obtained with accommodations point to the need for experimental tests of accommodations. The likely confounding of decisions about accommodations with both performance levels and characteristics of disabilities, and the dearth of information about the bases for these decisions, limit the usefulness of nonexperimental research that relies, as this report did, on variations in accommodations that arise naturally as a result of decentralized decision making. Planned variations in the accommodations provided and more detailed information about the students assessed are needed to more firmly identify the effects of accommodations on scores and on validity.

The results presented here also point to a specific hypothesis that warrants further exploration: Some of the accommodations used most in Kentucky—oral presentation, paraphrasing, and dictation—may be affecting primarily the reading difficulty of the assessment. This might explain several of the findings reported above. If this is so, mathematics items the difficulty of which hinges markedly on reading demands would tend to be more affected by accommodations, thus becoming differentially easy. Items the difficulty of which depends relatively little on reading would generally be less affected by accommodations and thus would appear differentially difficult. This might also explain why accommodations appear to generate less DIF in the reading assessment; their impact on performance in reading may be more consistent across items. If accommodations are having this effect, the subsequent question would be what are the circumstances under which this effect is appropriate in terms of its impact on the validity of scores.

Further research on the validity of scores. Research on the validity of scores obtained by students with disabilities—apart from the more specific question of the effects of accommodation—is also scanty. Here again, the research reported here sheds some light but also highlights the need for additional work. Moreover, to obtain a clearer view of the validity of the scores of students with disabilities may require not only additional research, but also new methods of validation and greater conceptual clarity about the inferences assessments are used to support.

One limitation of the validity evidence that is currently available for elementary and secondary assessments is the insufficient criteria to which scores can be compared. The analyses reported here focused on internal criteria of validity—that is, aspects of validity evidence internal to KIRIS itself, such as DIF and point-polyserial correlations. The findings with respect to these internal criteria are informative, but they were inconsistent and would be only an incomplete view of validity in any case. Research that focuses on testing for admissions screening, such as the SAT and the Graduate Record Examination (Willingham et al., 1988) and the Law School Admissions Test (Wightman, 1993), can use an external, predictive criterion—later performance, such as freshman grades—as an imperfect but very useful additional source of validity evidence. No comparable predictive criterion, however, is available for validating performance on KIRIS or on most other large-scale elementary and secondary achievement tests. Concurrent external criteria, such as class grades or performance on other tests, are likely to be ambiguous in the case of many students with disabilities. For example, a low correlation between grades and a test could be positive validity evidence if teachers were providing disabled students with inadequate opportunities to display their competence, but it could be negative evidence if it indicated that the test is less fair to students with disabilities than are teachers' grades.

The use of any validity evidence, however, requires clarity about the inference itself, and that clarity may be still be wanting in the case of some students with disabilities—particularly the especially numerous students with learning or cognitive disabilities. For example, if a reading-disabled student finds it difficult to read science problems or to write required answers, is that student's score made more or less valid by presenting the items orally or by allowing the student to dictate responses? Given the results reported here, these appear to be particularly pressing questions, but the answers hinge in part on the nature of the inference the scores will support. If the inference is primarily about mastery of scientific knowledge, these accommodations may increase the validity of scores. But if the inference extends to real-world applications of scientific knowledge and skills and to the ability to communicate scientific information, these accommodations may weaken the validity of scores. Moreover, even accommodations that increase the validity of scores for some learning-disabled students may lessen validity in the case of students with mental retardation. That is, students with learning disabilities are thought to have a disability that makes it difficult for them to manifest their cognitive skills, and some accommodations might therefore increase the validity of their scores by offsetting that obstruction. In contrast, students with retardation are by definition those with cognitive deficits, so to some degree, their poor performance on tests is likely to be accurate.

What is needed is therefore a broad effort that explores the utility of a range of research methods and validation criteria for evaluating the quality of assessment-based information about the performance of students with disabilities. Over the long run, this effort will require not only the analysis of conventional internal and external criteria, but also primary data collection, perhaps including the retesting of students with various types and degrees of accommodation.

Research on the effects of assessing students with disabilities. Inclusion of students with disabilities in large-scale assessments is seen by many advocates not as an end in itself, but rather as a means to improve the educational opportunities provided to them. Research on the diverse effects of inclusion is therefore critically important.

Implications for Policy and Practice

Years will be required to build a more adequate foundation of research to guide policies pertaining to the assessment of students with disabilities. In the meantime, educators and policymakers need to make decisions about how to proceed. Despite the many questions that remain unanswered, the present study does offer a detailed view of the assessment of these students that has potential implications for policy and practice.

The implications of this study, however, may not apply equally to all groups of students with disabilities. Students with disabilities constitute a highly diverse group. Some have severe cognitive deficits but no physical disabilities; others have severe physical limitations but very high levels of intellectual performance; some have severe psychological conditions, such as autism, that make their underlying level of intellectual performance difficult to ascertain; and yet others have face difficulties that might be deemed a learning disability in one jurisdiction but simply low ability or poor performance in another. It is not sensible to expect that all of these highly diverse students have similar needs in terms of either instruction or assessment simply by virtue of having one or another condition that policymakers currently subsume under the rubric of disability, and it is not likely that the results presented here apply similarly to all of them.

The students for whom the implications of these findings are clearest are those with cognitive and learning disabilities: learning-disabled students, who constitute nearly half of all of the fourth-grade students with disabilities served under IDEA in Kentucky, and mildly retarded students, who constitute roughly another fourth. By virtue of their numbers, these students had a disproportionate influence on the results for all students with disabilities, and they were also the specific focus of some of the analyses. Students with emotional/behavioral or speech/communication disorders were much less numerous, and the extent to which their characteristics and performance are reflected in these findings is less clear.

The other identified disabilities have such low prevalence rates that they are not reflected to any appreciable degree in the findings reported here, although some of the implications of this study—such as the importance of closely monitoring the use of accommodations—may apply to them nonetheless. Ironically, these relatively rare disabilities include the physical disabilities that were a primary focus of early efforts to establish policies and practices for the assessment of individuals with disabilities (Phillips, 1996). For example, the need to provide special forms of tests, such as large type or Braille, to students who have severe visual impairments has long been recognized, and the potential advantages and disadvantages of various approaches to assessing these students has often been discussed (e.g., Fischer, 1994). Students whose primary disability is a visual impairment, however, constituted only about 0.5% of the identified population of students with disabilities ages 6 through 21 served under IDEA Part B nationwide in 1994-95. Similarly, students with orthopedic impairments, a subset of whom may need alternative modes of producing and recording answers, constituted only 1.2% of the served population (U.S. Department of Education, 1996).

Monitoring the assessment of students with disabilities. Apart from the research efforts noted above, routine monitoring of the assessment of students with disabilities as part of ongoing assessment programs could provide valuable information for refining policy while simultaneously helping to minimize unintended poor practice. Monitoring at intervals of several years might suffice, as some aspects of the assessment of these students are likely to change only slowly.

The results reported here suggest the potential usefulness of monitoring three aspects of the assessment of students with disabilities: the difficulty of the assessments for these students, the use of accommodations, and simple aggregate statistics on the performance of students with disabilities. To track difficulty, three simple statistics—mean score, percent blank, and percent zero—calculated at the level of individual items for all students with disabilities or for important subgroups would suffice for most purposes. (Similar monitoring for other low-achieving groups would also be useful.) Routine monitoring of the use of accommodations would provide policymakers with an indication of whether practice is consistent with their intent and with the guidelines provided to educators. In Kentucky, most of the basic information is already collected from proctors routinely, although it may be useful to add additional questions, such as questions on the use of special settings and extra time. Routine or at least periodic calculation of some of the simpler aggregate statistics presented here, such as mean scores for students assessed with various combinations of accommodations, would alert policymakers to possible problems that warrant further investigation.

Reconciling inclusiveness and high standards. The results presented here may be one indication of a tension between two the goals of many current assessment systems—the focus on high standards on the one hand, and the goal of including and raising the performance of low-achieving students on the other.

Before the current wave of education reform, this tension was less important to many large-scale assessments. Minimum-competency tests, for example, were constructed to be appropriately difficult for most low-achieving students. Some other tests, such as the NAEP Trial State Assessment, were designed primarily to produce efficient estimates of performance for the student population as a whole, and their limitations for assessing performance nearer the extremes were therefore not a pressing concern.

In contrast, KIRIS, like some other current large-scale assessments, is intended both to focus on relatively high levels of performance and to improve performance of students at the low end of the distribution (among others). The tension between these two goals is apparent in some of the results for students with disabilities, which is unsurprising given that the identified population of students with disabilities is dominated by students with cognitive and learning disabilities. But the conflict is by no means a matter of disability *per se* or of the policy decision to increase the inclusion of students with disabilities in the assessment. Rather, it is an issue that is likely to arise in the assessment of low-achieving students more generally.

There are a variety of approaches for addressing this tension, but all entail trade-offs.

One option is simply to accept the costs of weak measurement for some low-achieving students. For example, a state might decide that including all students in a single assessment sends a desirable signal to schools, even if it comes at the cost of using a test that is too difficult for some students—as appears to be true in the case of the KIRIS mathematics assessment. Clearly, the higher the target level of the assessment, the more severe this measurement cost is likely to be, not only because the mismatch will be greater for a given student, but also because the number of students for whom the assessment is poorly suited will increase. However, the importance of these problems will also hinge on the uses of scores. The costs are least serious when scores are used only to gauge the average performance of entire schools, districts, or states. In that case, the impact of errors in measurement of the performance of low-scoring students may be small—even inconsequential, if the affected proportion of the student population is small enough. The consequences will be more substantial if scores are reported at the level of smaller aggregates (for example, means for all students with disabilities, or the proportion of students passing a relatively low standard), because the students for whom measurement is weak will be a larger share of the whole. The importance of poor measurement would be greatest, of course, if scores were reported at the level of the individual student—particularly if the scores carried with them consequences for the students.

It is important to recognize, however, that the potential costs of using an overly difficult assessment, just like the potential benefits, could go beyond measurement *per se* and are at present poorly understood. For example, this approach might increase the opportunities afforded to students with disabilities, but it could also distort incentives and degrade the instruction offered to some. That is, if test items are aimed too far above students' current level of performance, teachers may be encouraged to use inappropriate forms of test preparation rather than strengthened instruction. Similarly, the use of overly hard items may discourage students. The mix of positive and negative effects of using overly difficult assessments remains an open, empirical question.

Another option for addressing the tension between inclusion and high standards, when it appears to be severe, would be to use an assessment system designed specifically to provide high-quality measurement over a broad range of performance, rather than a single assessment for each subject area. For example, states could offer several tests of a given content area pegged to different levels of difficulty, using grades, coursework, or a screening test to direct students to a particular assessment. In order to avoid unintended tracking, students who pass a lower level test could be given the opportunity to sit for a more difficult one later. This would provide better measurement for students at diverse levels of performance, but it would impose significant administrative and financial costs and might undermine the state's efforts to extend high standards to the largest possible number of students. Computer-adaptive testing could also be employed for the same purpose. This approach selects tasks for students during testing based on their performance on previous items in the test. This in effect creates a range of tests of varying difficulty out of a single, broad item bank. This approach, however, would entail sizable financial and administrative costs, particularly if large numbers of computers had to be purchased, housed, and maintained. If constructed response items were to be used, it would be necessary to ensure that all students had sufficient typing skill to avoid confounding differences in typing skill with differences in the abilities being measured, and alternative input options would be needed for some disabled students. This approach would also require that policymakers and the public accept a system in which students are not administered all the same tests, and some methods of adaptive testing are difficult to explain to lay audiences (e.g., Stocking, 1996).

A third option would be to use assessment accommodations to lessen the tension between inclusiveness and high standards. (Indeed, some advocates now suggest offering accommodations not only to students with disabilities, but to all low-performing students who would score better with them.) The extremely widespread use of accommodations with disabled students in Kentucky suggests that many Kentucky educators have *de facto* chosen this option. The use of accommodations to lessen this tension, however, is appropriate in only some instances and even then raises a number of difficult issues. Some of these are discussed in the following section.

Regulating the use of accommodations. Policymakers and educators face serious uncertainties in deciding how to regulate the use of assessment accommodations. This study suggests that accommodations may have been used more in KIRIS than either KDE's guidelines or measurement considerations would have warranted, and the use of some accommodations was associated with a number of substantial measurement problems. On the other hand, it is not clear how much further restricting the use of accommodations would have reduced these measurement problems; some portion of the problems may have stemmed from the characteristics of the particular students given those accommodations or other factors rather than from the use of accommodations *per se*. Moreover, research clarifying the effects of accommodations for various types of students and assessments remains in very short supply.

Nonetheless, there are steps policymakers could consider for regulating the use of accommodations. Deciding among these options may require compromising between measurement goals and other educational goals, especially in the light of the current dearth of guidance from research. For example, the desire to maximize inclusion in the assessment program—as a tool for making educators more accountable for the education of students with disabilities—may argue for a permissive policy with respect to accommodations, while the desire for meaningful scores may argue for a restrictive policy.

For purposes of measurement, the function of accommodations is to offset biases caused by a disability. As Willingham noted, "in the case of disabled students, some aspects of standardization are breached in the interest of reducing sources of *irrelevant* difficulty that might otherwise lower scores artificially" (Willingham, 1988, p. 12, emphasis added). The purpose of an accommodation is not to raise scores *per se*, but rather to make them fairer and more accurate. To do this, one must be able to identify the difficulties caused by the disability that are *irrelevant* to the construct measured and to design accommodations that will offset those specific difficulties.

In some cases, the bias caused by a disability is unambiguous, and the nature of appropriate accommodations seems reasonably clear. For example, a test administered only in printed form cannot provide a valid measure of the academic performance of students with severe visual impairments. A test that requires writing with a pencil cannot provide a valid measure of the academic performance of a student with physical disabilities that impede or prevent the use of a pencil. In response to these biases, tests are altered in some way—by presenting them in different form, by administering them under different circumstances, or by providing alternative response modes—in order to obtain a more accurate view of the student's performance in the domain in question.

Most students with disabilities, however, have either learning or cognitive disabilities, and the appropriate accommodations for these students are often not apparent. That is, for these students, it is often difficult to distinguish difficulty on the assessment that is *irrelevant* to the construct measured—and that is therefore an appropriate target for accommodations—from difficulties that accurately reflect these students' weak mastery of the content or skills at issue. This difficulty is compounded by the ambiguous classification of students—in particular, the inconsistent classification of students as learning disabled—and sometimes by a lack of clarity about the specific inferences the test is intended to support.

Despite these uncertainties, and in part to address them, policymakers could consider the following several steps:

Clarifying the intended inferences. The great majority of instances of accommodations in KIRIS involved the use of language: oral presentation, paraphrasing, and dictation. Moreover, the largest single group of students with disabilities are learning disabled, most of whom have some language-related difficulty. Therefore, further clarifying the role of decoding, reading comprehension, writing, and other expression in the inferences the test is intended to support (distinguishing between reading and other subjects) could be a large step toward more consistent accommodation practices.

Because of the prevalence of reading disabilities, one might expect accommodations that lessen the need to read (oral presentation) or potentially make comprehension easier (paraphrasing) would be particularly frequent, and the results of this study confirmed that expectation. Surprisingly, however, these accommodations were not those with particularly strong associations with scores. The far stronger association of dictation with scores and the frequency of its use in the fourth grade suggest the importance of considering productive rather than only receptive language skills in specifying intended inferences.

Clarifying guidelines for accommodations. Crafting clear guidelines for accommodations appears to be insufficient. As noted earlier, KDE's guidelines are clear and seemingly restrictive, but the use of accommodations appears to have been excessive nonetheless. Active dissemination of guidelines and incorporation of information about accommodations in professional development activities could improve educators' awareness and understanding of the issues involved. In addition, it might prove helpful to focus professional development and documentation in part on the basic logic and purposes of accommodation, explaining to educators that the goal is not simply to raise scores, but rather to offset biases caused by specific disabilities.

Monitoring the use of accommodations by individual schools. If routine monitoring of accommodations is undertaken, in line with the suggestion above, it might be practical to use those data to explore variations among schools in the use of accommodations. Currently, the use of accommodations is so extensive in

Kentucky, particularly in the fourth grade, that comparisons among schools would likely be uninformative. Over time, however, these comparisons might be able to highlight schools with unusually permissive or restrictive practices that warrant additional investigation.

Periodic self-reports and audits. Periodic surveys and audits could be undertaken both to obtain more detailed information about the uses of accommodations and to signal to educators the importance of keeping practice consistent with published guidelines. Educators could be questioned, for example, about their reasons for providing accommodations, and they could be asked to provide concrete examples of their practices.

Next Steps

The national movement toward greater inclusion of students with disabilities in large-scale assessments has only begun. By taking the lead among states in including most students with disabilities in its statewide assessment program, Kentucky has also placed itself in the forefront of efforts to explore the feasibility and impact of this effort. While Kentucky's initial experience is instructive, the experiences in other jurisdictions, with different assessments, testing policies, and populations, may be quite different. The shortage of both practical experience and relevant research suggests that the trend toward greater inclusion will require careful monitoring and many midcourse corrections over a period of years.

REFERENCES

Anderson, N. E., Jenkins, F. F., & Miller, K. E. (n.d.) *NAEP inclusion criteria and testing accommodations: Findings from the NAEP 1995 Field Test in Mathematics*. Princeton, NJ: Educational Testing Service.

Drasgow, F. (1985). Polychoric and polyserial correlations. In S. Kotz & N. Johnson (Eds.), *The encyclopedia of statistical sciences*. New York: Wiley.

Erickson, R. N., Thurlow, M. L., & Thor, K. (1995). *1994 state special education outcomes*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Fischer, R. J. (1994). The Americans With Disabilities Act: Implications for measurement. *Educational Measurement: Issues and Practice, 13*(3), 17-26, 37.

Kentucky Department of Education. (1996). [Draft] *Procedures for inclusion of all students in the KIRIS accountability and state-required national-reference assessments*. Frankfort, KY: Author.

Koretz, D., Barron, S., Mitchell, K., & Stecher, B. (1996). *The perceived effects of the Kentucky Instructional Results Information System (KIRIS)* (MR-792-PCT/FF). Santa Monica, CA: RAND.

Lyon, G. R. (1996). Learning disabilities. *The future of children, 6*(1), 54-76. (Available online at: http://www.ldonline.org/ld_indepth/general_info/future_children.html).

McGrew, K. S., Thurlow, M. L., Shriner, J. G., & Spiegel, A. N. (1992). *Inclusion of students with disabilities in national and state data collection programs*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Miller, T. R., & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement, 30,* 107-122.

National Research Council, Committee on Goals 2000 and the Inclusion of Students With Disabilities. (1997). *Educating one and all: Students with disabilities and standards-based reform*. Washington, DC: National Academy Press.

Phillips, S. E. (1996). Legal defensibility of standards: Issues and policy perspectives. *Educational Measurement: Issues and Practice, 15(*2), 5-13, 19.

Shepard, L. A. (1989). Identification of mild handicaps. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 545-572). New York: Macmillan.

Shriner, J. G., & Thurlow, M. L. (1992). *State special education outcomes 1991*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

Stocking, M. (1996). An alternative method for scoring adaptive tests. *Journal of Educational and Behavioral Statistics, 21,* 365-389.

U.S. Department of Education. (1995). *To assure the free appropriate public education of all children with disabilities: Seventeenth annual report to Congress on the implementation of the Individuals With Disabilities Education Act*. Washington, DC: Author.

U.S. Department of Education. (1996). *To assure the free appropriate public education of all children with disabilities: Eighteenth annual report to Congress on the implementation of the Individuals With Disabilities Education Act*. Washington, DC: author.

Wightman, L. F. (1993). *Test takers with disabilities: A summary of data from special administrations of the LSAT* (Research Report 93-03). Newton, PA: Law School Admission Council.

Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). *Testing handicapped people*. Boston: Allyn and Bacon.

APPENDIX A

REGRESSION ANALYSES OF ACCOMMODATIONS

Grade 4 Learning-Disabled Students

8 case(s) deleted due to missing data.

Dep Var: READING   N: 2150   Multiple R: 0.371   Squared multiple R: 0.138

Adjusted squared multiple R: 0.136   Standard error of estimate: 0.946

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.523	0.053	0.0	.	-9.844	0.000
PARA	0.122	0.043	0.059	0.903	2.818	0.005
ORAL	0.008	0.059	0.003	0.875	0.139	0.889
DIC	0.705	0.043	0.345	0.899	16.328	0.000
CUE	0.336	0.070	0.100	0.930	4.810	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	306.577	4	76.644	85.721	0.000
Residual	1917.871	2145	0.894		

---

8 case(s) deleted due to missing data.

Dep Var: MATH   N: 2150   Multiple R: 0.251   Squared multiple R: 0.063

Adjusted squared multiple R: 0.061   Standard error of estimate: 1.103

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.516	0.062	0.0	.	-8.332	0.000
PARA	0.226	0.051	0.098	0.903	4.468	0.000
ORAL	0.053	0.068	0.017	0.875	0.773	0.440
DIC	0.429	0.050	0.188	0.899	8.514	0.000
CUE	0.351	0.082	0.093	0.930	4.310	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	175.628	4	43.907	36.117	0.000
Residual	2607.621	2145	1.216		

---

8 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 2150   Multiple R: 0.333   Squared multiple R: 0.111

Adjusted squared multiple R: 0.110   Standard error of estimate: 1.018

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.311	0.057	0.0	.	-5.443	0.000
PARA	0.167	0.047	0.077	0.903	3.586	0.000
ORAL	0.069	0.063	0.024	0.875	1.091	0.276
DIC	0.648	0.046	0.299	0.899	13.945	0.000
CUE	0.250	0.075	0.070	0.930	3.324	0.001

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	277.826	4	69.457	67.081	0.000
Residual	2220.954	2145	1.035		

----

8 case(s) deleted due to missing data.

Dep Var: SOCST   N: 2150   Multiple R: 0.357   Squared multiple R: 0.127

Adjusted squared multiple R: 0.126   Standard error of estimate: 0.962

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.422	0.054	0.0	.	-7.798	0.000
PARA	0.116	0.044	0.056	0.903	2.636	0.008
ORAL	-0.073	0.060	-0.026	0.875	-1.216	0.224
DIC	0.710	0.044	0.344	0.899	16.171	0.000
CUE	0.298	0.071	0.088	0.930	4.191	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	289.526	4	72.381	78.222	0.000
Residual	1984.852	2145	0.925		

Grade 8 Learning-Disabled Students

11 case(s) deleted due to missing data.

Dep Var: READING   N: 2107   Multiple R: 0.197   Squared multiple R: 0.039

Adjusted squared multiple R: 0.037   Standard error of estimate: 1.027

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.075	0.035	0.0	.	-30.548	0.000
PARA	0.099	0.053	0.047	0.704	1.853	0.064
ORAL	-0.023	0.052	-0.011	0.752	-0.439	0.661
DIC	0.605	0.074	0.181	0.925	8.150	0.000
CUE	0.062	0.074	0.019	0.851	0.829	0.407

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	89.826	4	22.456	21.306	0.000
Residual	2215.451	2102	1.054		

----

11 case(s) deleted due to missing data.

Dep Var: MATH   N: 2107   Multiple R: 0.267   Squared multiple R: 0.072

Adjusted squared multiple R: 0.070   Standard error of estimate: 0.959

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.999	0.033	0.0	.	-30.389	0.000
PARA	0.118	0.050	0.059	0.704	2.361	0.018
ORAL	0.023	0.048	0.011	0.752	0.472	0.637
DIC	0.665	0.069	0.209	0.925	9.576	0.000
CUE	0.270	0.070	0.089	0.851	3.888	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	148.943	4	37.236	40.470	0.000
Residual	1934.032	2102	0.920		

----

11 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 2107   Multiple R: 0.250   Squared multiple R: 0.062

Adjusted squared multiple R: 0.060   Standard error of estimate: 0.975

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.963	0.033	0.0	.	-28.830	9.9E-16
PARA	0.081	0.051	0.040	0.704	1.605	0.109
ORAL	0.090	0.049	0.045	0.752	1.836	0.066
DIC	0.708	0.071	0.220	0.925	10.038	9.9E-16
CUE	-0.002	0.071	-0.001	0.851	-0.031	0.975

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	132.556	4	33.139	34.890	9.99201E-16
Residual	1996.491	2102	0.950		

---

11 case(s) deleted due to missing data.

Dep Var: SOCST   N: 2107   Multiple R: 0.254   Squared multiple R: 0.065

Adjusted squared multiple R: 0.063   Standard error of estimate: 0.929

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.977	0.032	0.0	.	-30.677	0.000
PARA	0.065	0.048	0.034	0.704	1.348	0.178
ORAL	-0.061	0.047	-0.032	0.752	-1.308	0.191
DIC	0.755	0.067	0.246	0.925	11.225	0.000
CUE	0.095	0.067	0.032	0.851	1.405	0.160

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	125.374	4	31.343	36.309	0.000
Residual	1814.542	2102	0.863		

Grade 11 Learning-Disabled Students

6 case(s) deleted due to missing data.

Dep Var: READING   N: 1055   Multiple R: 0.183   Squared multiple R: 0.034

Adjusted squared multiple R: 0.030   Standard error of estimate: 0.939

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.450	0.042	0.0	.	-34.281	0.000
PARA	0.134	0.069	0.070	0.707	1.950	0.052
ORAL	0.152	0.068	0.079	0.727	2.219	0.027
DIC	0.504	0.156	0.100	0.960	3.224	0.001
CUE	0.049	0.098	0.016	0.876	0.500	0.617

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	32.204	4	8.051	9.127	0.000
Residual	926.231	1050	0.882		

----

6 case(s) deleted due to missing data.

Dep Var: MATH   N: 1055   Multiple R: 0.353   Squared multiple R: 0.125

Adjusted squared multiple R: 0.122   Standard error of estimate: 0.818

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.234	0.037	0.0	.	-33.496	0.000
PARA	0.310	0.060	0.177	0.707	5.167	0.000
ORAL	0.121	0.060	0.069	0.727	2.025	0.043
DIC	0.799	0.136	0.173	0.960	5.866	0.000
CUE	0.319	0.085	0.115	0.876	3.743	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	100.298	4	25.075	37.477	0.000
Residual	702.515	1050	0.669		

----

6 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 1055   Multiple R: 0.243   Squared multiple R: 0.059

Adjusted squared multiple R: 0.056   Standard error of estimate: 0.904

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.287	0.041	0.0	.	-31.626	0.000
PARA	0.134	0.066	0.072	0.707	2.029	0.043
ORAL	0.116	0.066	0.062	0.727	1.762	0.078
DIC	0.842	0.151	0.171	0.960	5.594	0.000
CUE	0.167	0.094	0.057	0.876	1.769	0.077

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	53.969	4	13.492	16.513	0.000
Residual	857.938	1050	0.817		

----

6 case(s) deleted due to missing data.

Dep Var: SOCST   N: 1055   Multiple R: 0.266   Squared multiple R: 0.071

Adjusted squared multiple R: 0.067   Standard error of estimate: 0.859

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.249	0.039	0.0	.	-32.285	0.000
PARA	0.225	0.063	0.127	0.707	3.576	0.000
ORAL	0.051	0.063	0.028	0.727	0.814	0.416
DIC	0.797	0.143	0.169	0.960	5.571	0.000
CUE	0.203	0.090	0.072	0.876	2.264	0.024

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	59.005	4	14.751	19.974	0.000
Residual	775.450	1050	0.739		

Grade 4 Mildly Mentally Retarded Students

16 case(s) deleted due to missing data.

Dep Var: READING   N: 1327   Multiple R: 0.432   Squared multiple R: 0.186

Adjusted squared multiple R: 0.184   Standard error of estimate: 1.076

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.305	0.090	0.0	.	-14.498	0.000
PARA	0.206	0.064	0.086	0.872	3.234	0.001
ORAL	0.178	0.104	0.047	0.816	1.707	0.088
DIC	0.897	0.067	0.367	0.829	13.485	0.000
CUE	0.254	0.098	0.067	0.904	2.585	0.010

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	350.912	4	87.728	75.739	0.000
Residual	1531.262	1322	1.158		

---

16 case(s) deleted due to missing data.

Dep Var: MATH   N: 1327   Multiple R: 0.317   Squared multiple R: 0.101

Adjusted squared multiple R: 0.098   Standard error of estimate: 1.100

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.395	0.092	0.0	.	-15.164	0.000
PARA	0.319	0.065	0.136	0.872	4.886	0.000
ORAL	0.189	0.107	0.051	0.816	1.774	0.076
DIC	0.454	0.068	0.191	0.829	6.681	0.000
CUE	0.351	0.100	0.096	0.904	3.502	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	179.263	4	44.816	37.045	0.000
Residual	1599.292	1322	1.210		

---

16 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 1327   Multiple R: 0.424   Squared multiple R: 0.180

Adjusted squared multiple R: 0.177   Standard error of estimate: 1.087

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.220	0.091	0.0	.	-13.421	0.000
PARA	0.311	0.064	0.129	0.872	4.820	0.000
ORAL	0.223	0.105	0.058	0.816	2.119	0.034
DIC	0.831	0.067	0.338	0.829	12.370	0.000
CUE	0.191	0.099	0.050	0.904	1.926	0.054

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	342.114	4	85.529	72.424	0.000
Residual	1561.215	1322	1.181		

----

16 case(s) deleted due to missing data.

Dep Var: SOCST   N: 1327   Multiple R: 0.393   Squared multiple R: 0.155

Adjusted squared multiple R: 0.152   Standard error of estimate: 1.081

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.396	0.090	0.0	.	-15.440	0.000
PARA	0.225	0.064	0.095	0.872	3.518	0.000
ORAL	0.252	0.105	0.067	0.816	2.405	0.016
DIC	0.773	0.067	0.321	0.829	11.566	0.000
CUE	0.144	0.099	0.039	0.904	1.460	0.144

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	282.698	4	70.675	60.484	0.000
Residual	1544.725	1322	1.168		

Grade 8 Mildly Mentally Retarded Students

31 case(s) deleted due to missing data.

Dep Var: READING N: 1215 Multiple R: 0.209 Squared multiple R: 0.044

Adjusted squared multiple R: 0.040 Standard error of estimate: 1.021

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.607	0.060	0.0	.	-26.931	0.000
PARA	0.211	0.069	0.100	0.744	3.061	0.002
ORAL	0.130	0.074	0.055	0.805	1.754	0.080
DIC	0.167	0.075	0.066	0.896	2.219	0.027
CUE	0.233	0.086	0.082	0.862	2.715	0.007

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	57.420	4	14.355	13.784	0.000
Residual	1260.162	1210	1.041		

---

31 case(s) deleted due to missing data.

Dep Var: MATH N: 1215 Multiple R: 0.323 Squared multiple R: 0.105

Adjusted squared multiple R: 0.102 Standard error of estimate: 0.900

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.719	0.053	0.0	.	-32.650	0.000
PARA	0.157	0.061	0.081	0.744	2.582	0.010
ORAL	0.355	0.066	0.164	0.805	5.407	0.000
DIC	0.324	0.067	0.140	0.896	4.876	0.000
CUE	0.222	0.076	0.086	0.862	2.925	0.004

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	114.584	4	28.646	35.354	0.000
Residual	980.421	1210	0.810		

---

31 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 1215   Multiple R: 0.303   Squared multiple R: 0.092

Adjusted squared multiple R: 0.089   Standard error of estimate: 0.934

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.691	0.055	0.0	.	-30.961	0.000
PARA	0.268	0.063	0.135	0.744	4.259	0.000
ORAL	0.209	0.068	0.094	0.805	3.068	0.002
DIC	0.272	0.069	0.114	0.896	3.947	0.000
CUE	0.266	0.079	0.100	0.862	3.387	0.001

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	106.671	4	26.668	30.585	0.000
Residual	1055.015	1210	0.872		

----

31 case(s) deleted due to missing data.

Dep Var: SOCST   N: 1215   Multiple R: 0.257   Squared multiple R: 0.066

Adjusted squared multiple R: 0.063   Standard error of estimate: 0.932

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.624	0.055	0.0	.	-29.790	0.000
PARA	0.250	0.063	0.128	0.744	3.979	0.000
ORAL	0.151	0.068	0.069	0.805	2.216	0.027
DIC	0.252	0.069	0.107	0.896	3.663	0.000
CUE	0.176	0.078	0.067	0.862	2.249	0.025

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	74.514	4	18.628	21.428	0.000
Residual	1051.919	1210	0.869		

Grade 11 Mildly Mentally Retarded Students

15 case(s) deleted due to missing data.

Dep Var: READING N: 583 Multiple R: 0.319 Squared multiple R: 0.101

Adjusted squared multiple R: 0.095 Standard error of estimate: 0.925

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.903	0.067	0.0	.	-28.589	0.000
PARA	-0.027	0.090	-0.014	0.732	-0.294	0.769
ORAL	0.116	0.093	0.058	0.731	1.253	0.211
DIC	0.816	0.156	0.212	0.943	5.228	0.000
CUE	0.566	0.128	0.183	0.910	4.431	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	55.922	4	13.981	16.322	0.000
Residual	495.076	578	0.857		

---

15 case(s) deleted due to missing data.

Dep Var: MATH N: 583 Multiple R: 0.453 Squared multiple R: 0.205

Adjusted squared multiple R: 0.200 Standard error of estimate: 0.703

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.622	0.051	0.0	.	-32.075	0.000
PARA	0.119	0.069	0.075	0.732	1.738	0.083
ORAL	0.250	0.071	0.154	0.731	3.547	0.000
DIC	0.727	0.119	0.234	0.943	6.127	0.000
CUE	0.549	0.097	0.220	0.910	5.661	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	73.670	4	18.418	37.270	0.000
Residual	285.630	578	0.494		

---

15 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 583   Multiple R: 0.348   Squared multiple R: 0.121

Adjusted squared multiple R: 0.115   Standard error of estimate: 0.840

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.810	0.060	0.0	.	-29.946	0.000
PARA	0.048	0.082	0.027	0.732	0.587	0.558
ORAL	0.106	0.084	0.057	0.731	1.254	0.210
DIC	0.744	0.142	0.211	0.943	5.246	0.000
CUE	0.583	0.116	0.206	0.910	5.032	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	56.293	4	14.073	19.923	0.000
Residual	408.282	578	0.706		

----

15 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 583   Multiple R: 0.348   Squared multiple R: 0.121

Adjusted squared multiple R: 0.115   Standard error of estimate: 0.840

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.810	0.060	0.0	.	-29.946	0.000
PARA	0.048	0.082	0.027	0.732	0.587	0.558
ORAL	0.106	0.084	0.057	0.731	1.254	0.210
DIC	0.744	0.142	0.211	0.943	5.246	0.000
CUE	0.583	0.116	0.206	0.910	5.032	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	56.293	4	14.073	19.923	0.000
Residual	408.282	578	0.706		

Grade 4 Students With Emotional/Behavioral Disturbances

9 case(s) deleted due to missing data.

Dep Var: READING   N: 319   Multiple R: 0.401   Squared multiple R: 0.161

Adjusted squared multiple R: 0.153   Standard error of estimate: 1.081

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.929	0.098	0.0	.	-9.472	0.000
PARA	0.126	0.142	0.054	0.739	0.893	0.373
ORAL	-0.057	0.162	-0.024	0.581	-0.350	0.727
DIC	0.929	0.149	0.392	0.674	6.231	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	70.338	3	23.446	20.076	0.000
Residual	367.879	315	1.168		

----

9 case(s) deleted due to missing data.

Dep Var: MATH   N: 319   Multiple R: 0.316   Squared multiple R: 0.100

Adjusted squared multiple R: 0.091   Standard error of estimate: 1.119

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.753	0.102	0.0	.	-7.409	0.000
PARA	0.182	0.147	0.077	0.739	1.243	0.215
ORAL	-0.163	0.168	-0.068	0.581	-0.969	0.333
DIC	0.752	0.154	0.317	0.674	4.873	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	43.613	3	14.538	11.609	0.000
Residual	394.470	315	1.252		

----

9 case(s) deleted due to missing data.

Dep Var: SCIENCE N: 319 Multiple R: 0.402 Squared multiple R: 0.161

Adjusted squared multiple R: 0.153 Standard error of estimate: 1.192

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.889	0.108	0.0	.	-8.215	0.000
PARA	0.015	0.156	0.006	0.739	0.093	0.926
ORAL	-0.041	0.179	-0.015	0.581	-0.227	0.821
DIC	1.067	0.164	0.408	0.674	6.490	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	86.055	3	28.685	20.180	0.000
Residual	447.763	315	1.421		

---

9 case(s) deleted due to missing data.

Dep Var: SOCST N: 319 Multiple R: 0.376 Squared multiple R: 0.142

Adjusted squared multiple R: 0.133 Standard error of estimate: 1.142

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-0.907	0.104	0.0	.	-8.749	0.000
PARA	0.254	0.150	0.103	0.739	1.700	0.090
ORAL	-0.229	0.171	-0.092	0.581	-1.336	0.182
DIC	0.935	0.158	0.378	0.674	5.938	0.000

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	67.748	3	22.583	17.308	0.000
Residual	410.986	315	1.305		

Grade 8 Students With Emotional/Behavioral Disturbances

12 case(s) deleted due to missing data.

Dep Var: READING   N: 401   Multiple R: 0.116   Squared multiple R: 0.014

Adjusted squared multiple R: 0.006   Standard error of estimate: 1.207

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.495	0.079	0.0	.	-18.900	0.000
PARA	0.015	0.148	0.006	0.726	0.098	0.922
ORAL	0.150	0.154	0.060	0.669	0.977	0.329
DIC	0.341	0.236	0.078	0.865	1.447	0.149

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	7.921	3	2.640	1.812	0.144
Residual	578.589	397	1.457		

---

12 case(s) deleted due to missing data.

Dep Var: MATH   N: 401   Multiple R: 0.182   Squared multiple R: 0.033

Adjusted squared multiple R: 0.026   Standard error of estimate: 1.034

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.313	0.068	0.0	.	-19.394	0.000
PARA	0.182	0.127	0.083	0.726	1.430	0.154
ORAL	0.198	0.132	0.091	0.669	1.502	0.134
DIC	0.234	0.202	0.061	0.865	1.159	0.247

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	14.524	3	4.841	4.532	0.004
Residual	424.087	397	1.068		

---

12 case(s) deleted due to missing data.

Dep Var: SCIENCE   N: 401   Multiple R: 0.185   Squared multiple R: 0.034

Adjusted squared multiple R: 0.027   Standard error of estimate: 1.083

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.338	0.071	0.0	.	-18.859	0.000
PARA	0.019	0.133	0.008	0.726	0.145	0.885
ORAL	0.267	0.138	0.117	0.669	1.938	0.053
DIC	0.407	0.212	0.102	0.865	1.925	0.055

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	16.500	3	5.500	4.688	0.003
Residual	465.745	397	1.173		

----

12 case(s) deleted due to missing data.

Dep Var: SOCST   N: 401   Multiple R: 0.156   Squared multiple R: 0.024

Adjusted squared multiple R: 0.017   Standard error of estimate: 1.000

Effect	Coefficient	Std Error	Std Coef	Tolerance	t	P(2 Tail)
CONSTANT	-1.442	0.065	0.0	.	-22.012	0.000
PARA	0.155	0.123	0.073	0.726	1.260	0.209
ORAL	0.160	0.127	0.076	0.669	1.258	0.209
DIC	0.191	0.195	0.052	0.865	0.980	0.327

Analysis of Variance

Source	Sum-of-Squares	DF	Mean-Square	F-Ratio	P
Regression	9.937	3	3.312	3.315	0.020
Residual	396.690	397	0.999		

APPENDIX B

ITEM-LEVEL ASSESSMENT RESULTS FOR STUDENTS WITH DISABILITIES

Descriptive analyses of item-level performance included means, percent of students leaving each item blank, and the percent of students scoring zero on each item. A zero score was defined in the Kentucky General Scoring Guide as an answer that "is completely wrong or has nothing to do with the question." These statistics were calculated for all mathematics and reading common items in Grades 4 and 8. The results are presented in the following tables.

Mean Scores

Table B1

Mean Scores on Common Reading Items, by Disability Status and Accommodations, Grade 4

	Item 3	Item 20	Item 41	Item 42	Item 44
No disability	2.4	2.3	1.9	1.9	2.1
Students with disabilities, no accommodations	2.0	1.8	1.5	1.4	1.6
Students with disabilities, with accommodations	2.3	2.0	1.9	1.6	2.1

Table B2

Mean Scores on Common Reading Items, by Disability Status and Accommodations, Grade 8

	Item 12	Item 19	Item 21	Item 31	Item 38	Item 41
No disability	2.0	2.2	1.7	1.8	2.3	2.4
Students with disabilities, no accommodations	1.5	1.6	1.1	1.3	1.8	1.8
Students with disabilities, with accommodations	1.5	1.6	1.3	1.4	1.8	1.9

Table B3

Mean Scores on Common Mathematics Items, by Disability Status and Accommodations, Grade 4

	Item 14	Item 17	Item 35	Item 41	Item 44
No disability	2.1	1.8	1.7	2.1	1.8
Students with disabilities, no accommodations	1.4	1.3	1.1	1.5	1.4
Students with disabilities, with accommodations	1.7	1.3	1.3	1.7	1.9

Table B4

Mean Scores on Common Mathematics Items, by Disability Status and Accommodations, Grade 8

	Item 17	Item 20	Item 29	Item 33	Item 35	Item 37
No disability	2.0	2.2	1.6	1.2	1.9	1.0
Students with disabilities, no accommodations	1.3	1.2	0.8	0.5	0.9	0.4
Students with disabilities, with accommodations	1.2	1.4	1.0	0.9	1.1	0.5

Percent of Students Receiving a Score of Zero

Table B5

Percent of Students Receiving a Score of Zero on Common Reading Items, by Disability Status and Accommodations, Grade 4

	Item 33	Item 20	Item 41	Item 42	Item 44
No disability	2	2	6	7	4
Students with disabilities, no accommodations	9	9	16	18	10
Students with disabilities, with accommodations	4	5	7	12	4

Table B6

Percent of Students Receiving a Score of Zero on Common Reading Items, by Disability Status and Accommodations, Grade 8

	Item 12	Item 19	Item 21	Item 31	Item 38	Item 41
No disability	1	1	5	2	0	0
Students with disabilities, no accommodations	11	10	21	10	4	5
Students with disabilities, with accommodations	8	7	15	10	4	4

Table B7

Percent of Students Receiving a Score of Zero on Common Mathematics Items, by Disability Status and Accommodations, Grade 4

	Item 14	Item 17	Item 35	Item 41	Item 44
No disability	27	31	16	8	19
Students with disabilities, no accommodations	44	46	39	22	30
Students with disabilities, with accommodations	42	48	39	21	21

Table B8

Percent of Students Receiving a Score of Zero on Common Mathematics Items, by Disability Status and Accommodations, Grade 8

	Item 17	Item 20	Item 29	Item 33	Item 35	Item 37
No disability	2	10	11	41	17	39
Students with disabilities, no accommodations	13	34	41	72	50	73
Students with disabilities, with accommodations	16	30	34	59	48	65

Percent of Students Leaving Items Blank

Table B9

Percent of Students Omitting Common Reading Items, by Disability Status and Accommodations, Grade 4

	Item 3	Item 20	Item 41	Item 42	Item 44
No disability	0.3	0.2	0.5	0.6	0.3
Students with disabilities, no accommodations	2.9	1.4	3.5	0.6	2.0
Students with disabilities, with accommodations	0.4	0.3	0.6	0.5	0.3

Table B10

Percent of Students Omitting Common Reading Items, by Disability Status and Accommodations, Grade 8

	Item 12	Item 19	Item 21	Item 31	Item 38	Item 41
no disability	1.0	0.9	1.4	1.3	0.9	0.9
Students with disabilities, no accommodations	3.8	3.4	5.9	5.9	3.2	3.3
Students with disabilities, with accommodations	1.7	1.2	2.4	2.7	1.2	1.3

Table B11

Percent of Students Omitting Common Mathematics Items, by Disability Status and Accommodations, Grade 4

	Item 14	Item 17	Item 35	Item 41	Item 44
No disability	0.4	0.5	0.5	0.3	0.8
Students with disabilities, no accommodations	1.7	3.0	2.6	2.6	3.7
Students with disabilities, with accommodations	0.7	0.7	0.6	0.5	1.1

Table B12

Percent of Students Omitting Common Mathematics Items, by Disability Status and Accommodations, Grade 8

	Item 17	Item 20	Item 29	Item 33	Item 35	Item 37
No disability	1.1	1.5	1.0	1.8	1.4	1.3
Students with disabilities, no accommodations	5.9	5.9	5.2	8.4	5.5	6.6
Students with disabilities, with accommodations	1.9	1.7	1.6	3.3	2.0	2.3