Examining the Validity of GED Tests Scores with Scheduling and Setting Accommodations

GED Testing Service Research Studies, 2004-1

General Educational Development Testing Service of the American Council on Education

GED® Testing Service
One Dupont Circle NW, Suite 250
Washington, DC 20036-1163
(202) 939-9490
Fax: (202) 659-8875
www.gedtest.org

Running head: TEST SCORES AND ACCOMMODATIONS

Examining the Validity of GED® Tests Scores with Scheduling and Setting Accommodations

Carol E. George-Ezzelle

GED® Testing Service of the American Council on Education

Gary Skaggs

Virginia Polytechnic Institute and State University

Paper presented at the annual conference of the National Council on Measurement in Education (NCME) held from April 13 to 15, 2004, in San Diego, CA.

Examining the Validity of Test Scores with Scheduling and Setting Accommodations

In an attempt to accommodate the testing needs of students with disabilities (SWDs) or English language learners (ELLs), tests or test administration conditions are often altered. The intention of the alterations, or accommodations, is to remove any barriers that prevent those students from demonstrating their "true" knowledge, skills, and abilities. Although accommodations are "heralded as promoting equity in assessment" (Sireci, Li, & Scarpati, 2003, p. 3), debate arises over whether test scores obtained under standard versus accommodated testing conditions actually have the same meaning and are, therefore, comparable, and whether the validity of test score interpretations varies across testing conditions for general population students and SWDs or ELLs.

Current testing standards call for test developers to provide evidence that testing procedures and test scores, and the inferences made based on the test scores, show evidence of validity and are comparable across subpopulations (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). Evidence of the comparability of test validity across subpopulations can be collected through examination of (a) the representation of the content domain being tested, (b) the relationship of test performance to other variables, (c) the internal structure of the test, (d) the response processes of examinees, and (e) the consequences of test score use. This research focuses on examining the internal structure of the test and the response processes of examinees with and without accommodations of test scheduling (e.g., extra time or breaks) and physical setting (private room).

Numerous studies on the comparability of tests across examinee subpopulations using the above approaches have been conducted. Specific subpopulations researched include those based on

gender, ethnicity, primary language, cultural backgrounds, and the use of test accommodations. Summaries of the research on the effect of test accommodations on test performance have yielded inconsistent findings. “One thing that is clear from our review is that there are no unequivocal conclusions that can be drawn regarding the effects, in general, of accommodations on students’ test performance. The literature is clear that accommodations and students are both heterogeneous” (Sireci et al., 2003, p. 16).

In an effort to contribute to the research on the effect of test accommodations on test performance, the current study examines the comparability of academic achievement test performance across examinees who did not receive any test accommodation and examinees who received an accommodation of either (a) extended time only, (b) extended time and private room only, or (c) extended time, private room, and supervised breaks only. These test accommodations represent situations in which there is a change in the scheduling/timing or physical setting of the exam versus the presentation (e.g., audio) or response format (dictated response) of the exam. The equivalence of an academic achievement test’s psychometric properties across accommodated and non-accommodated examinees will be examined through the calculation of group descriptive statistics, reliability estimates, standard errors of measurement, and differential item functioning (DIF).

Literature Review

Several summaries of test accommodation research have been published within the past three years (Abedi, Hofstetter, & Lord, 2003; Sireci et al., 2003; Thompson, Blount, & Thurlow, 2002). All of the summaries underscore the inconclusiveness of the many studies on the effects of test accommodations. Thompson et al. (2002) found in their review of test accommodation research from 1999 to 2001 that the number of studies on accommodations increased steadily, but findings

on the effects of test accommodations remained inconsistent. Sireci et al. (2003) also found inconsistent findings on the effects of test accommodations across their review of more than 150 studies. Abedi et al.'s (2003) review of test accommodations for English language learners (ELLs) questioned whether accommodations for ELLs benefited not only ELLs but also non-ELLs. Past accommodation research indicates a level of generalizability that is limited not only by inconsistent results across similar studies, but also by confounding variables such as the presence of multiple accommodations, small sample sizes, research design limitations, varied results by disability type, and impact of language proficiency level (Thurlow, McGrew, Tindal, Thompson, Ysseldyke, & Elliott, 2000). Thus, the level of generalizability of past accommodation research warrants further research in an attempt to ascertain the appropriateness of test accommodations.

There have been numerous studies pertaining to test scheduling accommodations, particularly extended time. Thompson et al. (2002), in a summary of research, found that the number of published research studies from 1999 through 2001 on timing/scheduling accommodations (17) was second only to the number of studies on oral administration (22). This predominance of studies on oral and extended time accommodations was also found in Sireci et al.'s (2003) review. Both research reviews summarized the effect of extended time on test performance as inconclusive. Thompson et al. found that extended time had positive effects on the performance of students with disabilities in four of the studies; however, an additional four studies did not find an effect. With respect to research specifically on test accommodations using extended time, Sireci et al. found that extended time generally improved the performance of both students with and without disabilities although it improved the performance of students with disabilities to a greater extent.

Fuchs (2000) used differential item functioning (DIF) to examine the comparability of test responses across several test accommodations for students with and without learning disabilities on statewide mathematics and reading tests. For computation items on the mathematics test, the extra time accommodation resulted in a greater number of items exhibiting substantial DIF in favor of students without disabilities than in favor of SWDs, thus providing no support for the use of extra time for SWDs. Similarly, no evidence of DIF was found on the reading test for either group of students. However, for concept and application items on the mathematics test, the extra time accommodation resulted in a greater number of items exhibiting substantial DIF in favor of the SWDs, thereby supporting the extra time accommodation for SWDs on these types of math items.

However, Zuriff (2000) and Abedi et al. (2003) concluded that evidence of the comparability of test scores under extra time accommodations is not consistent because several studies have found that the accommodation also improved the performance of non-learning-disabled students. Similar increased test score results were found for both ELLs and English-proficient students in a study by Abedi, Lord, Hofstetter, and Baker (2000). Results indicating invalidly inflated scores (that is, where both the general population of students and SWDs groups stand to gain in performance regardless of whether they need the accommodation) call into question the validity of the accommodation and the subsequent unfairness to students who did not receive the accommodation.

In a study that examined whether extended time would produce benefits for student with and without disabilities, Overall, Marquart (2000) found that students achieved similar scores under standard and extended time testing conditions. Further, similar effects were found not only for students with and without disabilities, but also for comparisons of students who were previously categorized as performing either at or below grade level. Marquart hypothesized that the lack of

performance differences may have been due to no apparent need for additional testing time (students finished well within the time limits even under standard time testing conditions). However, a more positive testing experience under extended time conditions was reflected in student questionnaires.

The College Board has played a prominent role in exploring the effects of extended time and has published various reports. In one report, Cahalan, Mandinach, and Camera (2002) examined the predictive validity of the SAT I: Reasoning Test scores under extended time accommodations for a group of students with learning disabilities. Adjusted correlations between the SAT scores and first year grade point average (FGPA) were positive with SWDs. However, with students without disabilities, the correlation between there two factors was not strong. The addition of high school grade point average (HSGPA), a variable generally thought to improve the prediction of FGPA, did not improve the prediction of FGPA for females with disabilities. The researchers cautioned that small sample sizes and sample selection procedures limit the generalizability of their research. In another report from the College Board, Camera and Schneider (2000) investigated the test-retest (junior and senior years) performance of four groups of students:

1. students without disabilities who took the SAT I twice under standard testing conditions,
2. students with a disability who took the SAT I twice with an extended time accommodation,
3. students with a disability who took the SAT I first under standard testing conditions and then with an extended time accommodation, and
4. students with a disability who took the SAT I first with an extended time accommodation and then under standard testing conditions.

The test score gains were highest for group 3—up to three times greater than gains reported for groups 1 and 2. However, once extra time limits increased beyond time and a half or double time, test score gains generally decreased.

Also categorized under "scheduling" accommodations is multiple-day test administration. Walz, Albus, Thompson, and Thurlow (2000) presented initial research on multiple-day accommodation effects for students with and without disabilities testing over three days vs. one day. No significant effects on test performance for student group, testing condition, or interaction were found. A slight increase in average performance was seen in the one-day test administration condition. The researchers recommended further research of this nature, using larger sample sizes and better control of language proficiency levels across student groups.

Studies reporting the use of an accommodation of physical setting, such as the use of a private room, appear to be limited to research undertaken as part of a multiple-accommodation research or meta analyses. Chiu and Pearson (1999) and Elliott, Kratochwill, and McKevitt (2001) reported improved test scores of SWDs using multiple accommodations. Multiple-accommodation research findings, excluding the two most frequently researched accommodations of extra time and oral presentation, report an increase in test scores of SWDs (Schulte, Elliot, and Kratochwill, 2001).

This study presents additional research in the area of scheduling and setting test accommodations, particularly the accommodation of extended time, with tests of general academic achievement. This research also aims to provide information on accommodation use across a variety of subject areas (writing, reading, social studies, and science). Furthermore, the current study explored test item response differences between examinees using a sample atypical of the population addressed by the majority of test accommodation research—that is, an adult population.

In conclusion, this research serves to augment the numerous previous test accommodation research studies as well as to expand the research by its use of examinee populations beyond high school.

Method

Participants

The data analyzed are from the 2002 examination cycle of the Tests of General Educational Development (GED®) (American Council on Education, 2002a). At the time of analysis, the database contained test and candidate data from GED administrations in 48 states and the District of Columbia (Ohio and Connecticut data are not included). Test and candidate data from the English version of the GED Tests administered in the U.S. during the 2002 examination cycle were the base source of the study's sample. This study will refer to the three operational forms as Form 1, Form 2, and Form 3. In 2002, approximately 140,000 to 160,000 candidates were administered each form of each test.

Within this base source, a small sample of candidates requested and received some form of accommodation in the administration of the tests. Prior to testing, candidates who requested test accommodations were required to complete a form to include documentation of his or her disability. Approval was granted after review by state-level GED administrators or GED Testing Service staff. The accommodation sample consists of test and candidate data from candidates who took the GED Tests with the following test scheduling and/or setting accommodations: (a) extended time only, or (b) extended time and private room only, or (c) extended time, private room, and supervised breaks only. The number of candidates receiving these accommodations is presented in Table 1.

Instruments

The study utilized test data obtained from the Tests of General Educational Development, a battery of tests designed "to measure academic achievement in a four-year program of high school education in the core content areas of U.S. and Canadian high school curricula" (American Council

on Education, 2002b, p. 4). The GED Tests consist of five tests in the following content areas, Language Arts; Writing; Social Studies; Science; Language Arts, Reading; and Mathematics. The Mathematics Test was not used in this study. Except for the Language Arts, Reading Test, all of the tests used in this study consisted of 50 multiple-choice items. The Language Arts, Reading Tests each consist of 40 items. The Language Arts, Writing and Reading Tests are structured around reading passages with related sets of items. Part of the Science and Social Studies Tests are also passage-based.

Table 1

Number of GED® Candidates Receiving Scheduling/Setting Accommodations

Test Form	Extended Time Only	Extended Time & Private Room Only	Extended Time, Private Room, & Supervised Breaks Only	Total
Language Arts, Writing Form 1	60	42	22	124
Language Arts, Writing Form 2	51	34	7	92
Language Arts, Writing Form 3	52	26	9	87
Social Studies Form 1	60	49	18	127
Social Studies Form 2	56	26	7	89
Social Studies Form 3	53	39	7	99
Science Form 1	49	40	17	106
Science Form 2	41	28	7	76
Science Form 3	50	34	7	91
Language Arts, Reading Form 1	71	48	18	137
Language Arts, Reading Form 2	61	36	7	104
Language Arts, Reading Form 3	64	42	13	119

In addition to test scores from the GED Tests, the study also accessed demographic information provided by each candidate. Such demographic information included age, gender, race, geographic region of residence, and highest level of education completed. Descriptive statistics on

demographic characteristics of the accommodated and non-accommodates samples are presented in Table 2. The non-response rate for the demographic questions was high (41% to 56%) for the accommodated sample vs. the non-accommodated sample (1% to 12%). Based on all responses, candidates in the accommodated sample were, on average, younger, less likely to be African American or of Hispanic descent, and likely to have an educational level lower than candidates in the non-accommodated sample.

Table 2

Sample Demographic Characteristics

	Sample Group	
Demographic	Accommodated	Non-Accommodated
Age		
16-20 years	34%	17%
20-24 years	12%	48%
25-29 years	3%	13%
30-34 years	3%	6%
35-39 years	2%	7%
40-49 years	3%	6%
50-59 years	2%	2%
60+ years	<1%	1%
Missing/Invalid	42%	<1%
Gender		
Male	42%	57%
Female	16%	41%
Missing	42%	2%
Ethnicity/Race		
Hispanic origin or descent	3%	12%
American Indian or Alaskan native	<1%	2%
Asian	<1%	1%
Black/African American	5%	21%
Native Hawaiian or Pacific Islander	<1%	<1%
White	41%	51%
Missing	50%	12%
Highest Educational Level		
None	<1%	0%
K-6th grade	<1%	1%
7th grade	5%	1%
8th grade	8%	7%
9th grade	14%	16%
10th grade	10%	25%
11th grade	4%	32%
12th grade	1%	6%
Missing	56%	10%
Geographic Region		
Northeast	23%	20%
Midwest	16%	16%
South	18%	41%
West	2%	21%
Missing	41%	2%

Differential Item Functioning

DIF analyses were conducted to evaluate whether individual items or groups of items performed differentially for accommodated vs. non-accommodated candidates. The SIBTEST procedure (Shealy & Stout, 1993) was used for all DIF analyses. SIBTEST evaluates differences in item functioning between two groups: the reference group and the focal group. For this study, the focal group consisted of those candidates who received the specified scheduling or setting accommodations. The reference group was a sample of candidates who did not receive any accommodations. Because more than 140,000 candidates per test form received no accommodations, a random sample of 500 candidates was selected from each test form to make the group sizes more comparable. SPSS Version 12.0 (SPSS, 2003) was used to carry out the sampling.

SIBTEST conducts a DIF analysis on a *suspect subtest* containing one or more items. Most traditional DIF analyses focus on a suspect subtest of one item or in other words, an individual item analysis. However, the suspect subtest can consist of groups of items. In this study, both types of analyses were carried out. Subtest groupings of items were based on item content; for example, poetry items in the Language Arts, Reading Tests were analyzed as a group. SIBTEST uses a *valid subtest* to match examinee ability levels. The valid subtest is a group of items that are assumed to be free of DIF. In this study, the valid subtest for each SIBTEST run consisted of all the items that were not part of the suspect subtest. In other words, in individual item analyses, the valid subtest comprised the other 49 (or 39 for Language Arts, Reading) items on the test. For content subtest analyses, the valid subtest consisted of the remaining items on the test.

The end product of a SIBTEST analysis is the calculation of a statistic, β_{UNI}. β_{UNI} has the following form:

$$\beta_{UNI} = \sum_{i=1}^{k} p_i \left(\overline{Y}_{Ri}^{*} - \overline{Y}_{Fi}^{*} \right)$$

where k is the number of items on the valid subtest, p_i is the proportion of focal group candidates obtaining raw score i, and $\overline{Y}_{Ri}^{*}$ and $\overline{Y}_{Fi}^{*}$ are the mean raw scores on the suspect subtest for the reference and focal groups, respectively, with raw score i on the valid subtest. The means are adjusted by a regression correction that effectively controls for an inflation to the Type I error rate that would occur due to measurement error. A useful feature of the β_{UNI} statistic is that its sign indicates the direction of DIF. A positive value favors the reference group, and a negative value favors the focal group. In addition, an asymptotic standard error is available for β_{UNI}. Dividing β_{UNI} by its standard error yields a z statistic that is normally distributed, thus providing a statistical test for the significance of the magnitude of β_{UNI}.

Results

Raw score descriptive statistics, Kuder-Richardson formula #20, and the standard error of measurement (SEM) were calculated for the accommodated and non-accommodated samples. These results are shown below in Table 3. In all 12 test forms, the candidates testing under standard administration had higher mean raw scores than the candidates receiving scheduling or setting accommodations. The differences in mean raw scores between the two groups ranged approximately from one-tenth to one-third of a standard deviation. The K-R 20s and SEMs were about the same between the two samples across test forms, providing evidence of equal reliability of the tests for both groups.

Table 3

Raw Score Descriptive Statistics for Accommodated and Non-Accommodated Samples

	n	Mean	Median	*SD*	Min	Max	K-R 20	SEM
Lang. Arts, Writing Form 1								
Accommodated	124	31.07	32	8.99	8	48	.89	3.0
Non-accommodated	500	34.45	37	9.65	6	50	.91	2.9
Lang. Arts, Writing Form 2								
Accommodated	92	32.01	33	7.78	14	50	.85	3.0
Non-accommodated	500	35.83	37	8.20	6	50	.88	2.8
Lang. Arts, Writing Form 3								
Accommodated	87	31.32	32	9.29	10	50	.90	3.0
Non-accommodated	500	34.67	36	9.16	7	50	.90	2.9
Social Studies Form 1								
Accommodated	127	32.76	35	10.06	4	49	.92	2.9
Non-accommodated	500	35.47	37	8.60	6	50	.89	2.9
Social Studies Form 2								
Accommodated	89	32.64	34	9.14	6	49	.90	2.9
Non-accommodated	500	33.77	35	10.09	1	50	.92	2.9
Social Studies Form 3								
Accommodated	99	31.12	32	9.98	6	49	.91	3.0
Non-accommodated	500	35.29	37	9.31	4	50	.91	2.8
Science Form 1								
Accommodated	106	33.68	36	10.68	6	50	.93	2.9
Non-accommodated	500	37.07	39	9.33	5	50	.92	2.6
Science Form 2								
Accommodated	76	35.04	35	7.93	11	50	.87	2.9
Non-accommodated	500	35.54	37	8.34	10	50	.88	2.9
Science Form 3								
Accommodated	91	33.98	35	9.78	8	48	.91	2.9
Non-accommodated	500	36.54	39	9.19	6	50	.91	2.8
Lang. Arts, Reading Form 1								
Accommodated	137	28.18	29	7.46	6	40	.89	2.5
Non-accommodated	500	29.86	32	7.88	8	40	.91	2.4
Lang. Arts, Reading Form 2								
Accommodated	104	28.26	30	6.93	7	39	.86	2.6
Non-accommodated	500	31.04	33	7.25	6	40	.90	2.3
Lang. Arts, Reading Form 3								
Accommodated	119	27.82	29	7.14	3	39	.87	2.6
Non-accommodated	500	29.53	31	6.70	5	40	.86	2.5

DIF for Individual Items

The next 12 tables show the *individual item* SIBTEST results for the 12 test forms. Table 4 shows the results for Language Arts, Writing Test, Form 1; Table 5 the results for Language Arts, Writing Test, Form 2; and so forth. In order to control for the Type I error rate within each form, the Bonferroni correction was used; an item was referred to content specialists for further review if the *p*-value for β_{UNI} was less than .05 divided by the number of items. Using this criterion, 11 items across the 12 test forms were identified as exhibiting substantial DIF. Table 16 lists the items exhibiting substantial DIF and contains item descriptions.

GED Tests content specialists examined the flagged items listed in Table 16 for any plausible reason to explain why the items favored the indicated group. For the three items from the Language Arts, Writing Tests, the specialists could identify no characteristic of the items that would have provided an advantage to one group over another. The single social studies item had a map with symbols that required analysis and evaluation. The specialists speculated that candidates with time or setting accommodations might have benefited by using extra time on this question to gain a better understanding of the map and symbols.

Two science items were flagged in favor of candidates with accommodations. One of the items had a food chain graphic, required analysis and inference, and was located toward the end of the test (number 40 out of 50). As speculated for the Social Studies Test item, candidates with time or setting accommodations may have benefited by using extra time to fully understand the graphic. The second science item had a topographic map with several legends; however, the cognitive requirements of the item appeared to be very low. No characteristic of the item that would have been advantageous to candidates with scheduling or setting test accommodations or disadvantageous to candidates under standard test administration was identified.

Table 4

SIBTEST Results for Individual Items: Language Arts, Writing, Form 1

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	-0.008	-0.133	0.894	26	-0.034	-0.919	0.358
2	-0.013	-0.366	0.714	27	-0.076	-3.593	0.000
3	0.032	0.558	0.577	28	0.051	0.950	0.342
4	0.001	0.031	0.976	29	-0.047	-1.313	0.189
5	0.007	0.230	0.818	30	-0.057	-1.135	0.256
6	0.060	1.324	0.186	31	0.118	2.697	0.007
7	-0.071	-1.692	0.091	32	-0.017	-0.369	0.712
8	-0.038	-0.759	0.448	33	0.036	0.690	0.490
9	0.062	1.167	0.243	34	0.050	0.942	0.346
10	-0.029	-0.564	0.572	35	-0.010	-0.191	0.848
11	-0.019	-0.500	0.617	36	0.008	0.135	0.893
12	0.038	0.634	0.526	37	0.003	0.062	0.950
13	-0.068	-2.410	0.016	38	0.014	0.254	0.799
14	0.031	0.587	0.557	39	-0.022	-0.766	0.443
15	-0.047	-1.763	0.078	40	-0.006	-0.132	0.895
16	-0.026	-0.440	0.660	41	-0.012	-0.229	0.819
17	0.024	0.598	0.550	42	-0.024	-0.456	0.648
18	-0.010	-0.198	0.843	43	-0.037	-0.860	0.390
19	0.101	1.917	0.055	44	-0.031	-0.572	0.567
20	0.084	2.062	0.039	45	0.077	1.279	0.201
21	-0.016	-0.553	0.580	46	0.053	1.091	0.275
22	0.031	0.650	0.515	47	-0.055	-1.776	0.076
23	0.047	1.079	0.281	48	0.016	0.454	0.650
24	-0.080	-1.524	0.128	49	0.094	1.893	0.058
25	0.024	0.454	0.650	50	0.069	1.125	0.261

Table 5

SIBTEST Results for Individual Items: Language Arts, Writing, Form 2

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.001	0.022	0.982	26	0.047	1.040	0.298
2	0.042	0.755	0.450	27	0.124	2.416	0.016
3	0.046	0.675	0.500	28	0.003	0.055	0.956
4	-0.003	-0.183	0.855	29	0.070	1.584	0.113
5	0.019	0.621	0.534	30	0.050	0.687	0.492
6	-0.097	-3.465	0.001	31	0.032	0.669	0.504
7	-0.013	-0.247	0.805	32	-0.003	-0.054	0.957
8	-0.052	-0.966	0.334	33	-0.003	-0.076	0.940
9	0.101	1.571	0.116	34	0.089	1.794	0.073
10	0.028	0.733	0.463	35	0.044	0.965	0.335
11	-0.003	-0.068	0.946	36	0.028	0.672	0.502
12	-0.035	-0.841	0.401	37	-0.016	-0.409	0.682
13	-0.084	-1.518	0.129	38	0.019	0.332	0.740
14	0.058	1.693	0.091	39	-0.092	-1.736	0.083
15	-0.054	-0.882	0.378	40	-0.069	-1.254	0.210
16	0.121	2.346	0.019	41	-0.045	-0.650	0.516
17	-0.086	-1.420	0.156	42	0.143	2.196	0.028
18	0.071	1.347	0.178	43	-0.117	-2.028	0.043
19	0.116	1.744	0.081	44	-0.082	-2.121	0.034
20	0.042	0.759	0.448	45	0.030	0.509	0.611
21	0.030	0.599	0.549	46	-0.014	-0.384	0.701
22	-0.006	-0.103	0.918	47	-0.007	-0.150	0.881
23	-0.107	-1.938	0.053	48	0.071	1.476	0.140
24	-0.007	-0.240	0.810	49	0.060	1.236	0.216
25	-0.037	-1.084	0.278	50	0.092	1.924	0.054

Table 6

SIBTEST Results for Individual Items: Language Arts, Writing, Form 3

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	-0.032	-1.363	0.173	26	0.046	0.711	0.477
2	-0.110	-1.676	0.094	27	0.034	0.547	0.585
3	-0.078	-0.992	0.321	28	0.005	0.177	0.859
4	-0.014	-0.404	0.686	29	-0.033	-0.554	0.580
5	-0.048	-1.486	0.137	30	0.126	2.037	0.042
6	-0.028	-0.445	0.656	31	0.039	0.666	0.506
7	0.020	0.408	0.683	32	0.006	0.129	0.897
8	-0.033	-0.693	0.488	33	0.016	0.457	0.648
9	-0.010	-0.215	0.830	34	0.031	0.637	0.524
10	-0.017	-0.337	0.736	35	-0.134	-2.266	0.023
11	-0.024	-0.356	0.722	36	-0.002	-0.024	0.981
12	-0.015	-0.766	0.444	37	0.011	0.219	0.827
13	0.001	0.022	0.983	38	0.057	1.166	0.244
14	0.002	0.027	0.979	39	-0.070	-1.135	0.257
15	-0.099	-1.702	0.089	40	0.070	1.030	0.303
16	-0.075	-1.104	0.270	41	-0.070	-1.288	0.198
17	-0.066	-1.364	0.173	42	-0.083	-1.262	0.207
18	0.145	3.278	0.001	43	0.007	0.116	0.907
19	-0.043	-0.670	0.503	44	0.175	2.714	0.007
20	0.060	0.922	0.356	45	-0.003	-0.051	0.959
21	0.002	0.028	0.977	46	-0.011	-0.172	0.863
22	0.048	0.921	0.357	47	-0.028	-0.456	0.648
23	0.042	0.834	0.404	48	-0.002	-0.037	0.970
24	-0.002	-0.176	0.860	49	0.001	0.011	0.991
25	-0.042	-0.719	0.472	50	0.069	1.125	0.261

Table 7

SIBTEST Results for Individual Items: Social Studies, Form 1

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.024	0.568	0.570	26	-0.089	-1.616	0.106
2	-0.001	-0.042	0.967	27	0.048	0.832	0.405
3	0.084	1.684	0.092	28	-0.051	-1.062	0.288
4	0.054	1.284	0.199	29	0.061	1.150	0.250
5	0.046	0.917	0.359	30	-0.004	-0.106	0.915
6	0.020	0.521	0.602	31	0.028	0.765	0.444
7	0.027	0.756	0.450	32	-0.027	-0.682	0.495
8	-0.013	-0.280	0.779	33	0.065	1.313	0.189
9	0.153	2.858	0.004	34	-0.029	-0.642	0.521
10	-0.006	-0.095	0.925	35	0.025	0.563	0.573
11	-0.035	-1.387	0.165	36	0.045	0.965	0.334
12	-0.042	-1.417	0.157	37	-0.029	-0.659	0.510
13	0.050	1.178	0.239	38	0.012	0.250	0.802
14	0.084	1.467	0.142	39	-0.080	-1.381	0.167
15	-0.058	-1.467	0.142	40	0.098	1.636	0.102
16	-0.017	-0.384	0.701	41	0.062	1.265	0.206
17	0.019	0.503	0.615	42	0.027	0.555	0.579
18	0.022	0.418	0.676	43	-0.090	-1.778	0.075
19	-0.030	-0.890	0.374	44	0.126	2.391	0.017
20	-0.036	-1.901	0.057	45	-0.022	-0.429	0.668
21	0.106	2.428	0.015	46	-0.110	-2.226	0.026
22	0.005	0.097	0.923	47	0.004	0.080	0.937
23	-0.007	-0.132	0.895	48	-0.021	-0.368	0.713
24	0.059	1.124	0.261	49	-0.115	-2.258	0.024
25	-0.009	-0.192	0.848	50	0.035	0.689	0.491

Table 8

SIBTEST Results for Individual Items: Social Studies, Form 2

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.003	0.104	0.917	26	0.013	0.231	0.817
2	-0.028	-1.059	0.290	27	0.008	0.153	0.879
3	-0.001	-0.024	0.981	28	0.049	1.121	0.262
4	0.082	1.443	0.149	29	-0.048	-0.947	0.343
5	-0.010	-0.282	0.778	30	0.061	1.007	0.314
6	-0.031	-1.052	0.293	31	0.147	2.333	0.020
7	0.055	1.555	0.120	32	-0.012	-0.241	0.809
8	0.100	1.678	0.093	33	0.043	1.491	0.136
9	-0.030	-0.679	0.497	34	0.083	1.260	0.208
10	-0.016	-0.240	0.811	35	-0.050	-0.957	0.339
11	0.040	0.957	0.339	36	0.108	1.677	0.094
12	0.060	1.591	0.112	37	0.002	0.051	0.959
13	0.062	1.260	0.208	38	-0.002	-0.033	0.973
14	-0.011	-0.226	0.821	39	-0.081	-1.850	0.064
15	0.077	1.497	0.134	40	-0.008	-0.164	0.870
16	-0.159	-3.257	0.001	41	-0.080	-1.884	0.060
17	-0.094	-1.572	0.116	42	-0.066	-0.982	0.326
18	-0.090	-1.997	0.046	43	-0.058	-0.998	0.318
19	-0.092	-2.003	0.045	44	-0.055	-1.063	0.288
20	-0.014	-0.348	0.728	45	0.086	1.279	0.201
21	0.053	1.153	0.249	46	0.034	0.610	0.542
22	0.026	0.497	0.619	47	0.008	0.122	0.903
23	0.013	0.203	0.839	48	0.060	0.936	0.349
24	-0.011	-0.209	0.835	49	0.022	0.370	0.711
25	0.088	1.589	0.112	50	0.073	1.215	0.224

Table 9

SIBTEST Results for Individual Items: Social Studies, Form 3

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.040	1.435	0.151	26	0.132	2.465	0.014
2	-0.015	-0.333	0.739	27	-0.101	-2.504	0.012
3	0.027	0.488	0.625	28	-0.029	-0.502	0.616
4	-0.054	-1.049	0.294	29	0.019	0.275	0.783
5	-0.028	-0.889	0.374	30	0.024	0.437	0.662
6	-0.027	-0.562	0.574	31	0.044	1.199	0.230
7	0.051	1.742	0.081	32	-0.079	-1.494	0.135
8	0.013	0.303	0.762	33	0.012	0.209	0.834
9	0.043	0.875	0.382	34	0.129	2.230	0.026
10	0.029	0.477	0.634	35	0.024	0.449	0.653
11	-0.074	-1.350	0.177	36	0.039	0.858	0.391
12	0.134	2.234	0.025	37	0.015	0.252	0.801
13	0.031	0.946	0.344	38	-0.103	-1.881	0.060
14	-0.111	-2.424	0.015	39	0.002	0.033	0.974
15	0.174	3.085	0.002	40	-0.072	-1.185	0.236
16	-0.045	-1.509	0.131	41	-0.133	-2.062	0.039
17	-0.058	-0.936	0.349	42	-0.132	-2.252	0.024
18	0.026	0.620	0.535	43	-0.053	-0.818	0.413
19	0.060	1.249	0.212	44	0.039	0.619	0.536
20	0.040	0.843	0.399	45	0.035	0.518	0.605
21	0.041	0.772	0.440	46	0.023	0.388	0.698
22	0.025	0.486	0.627	47	0.043	0.732	0.464
23	0.060	1.000	0.317	48	0.020	0.367	0.714
24	0.044	1.021	0.307	49	-0.028	-0.485	0.627
25	0.034	0.638	0.523	50	0.047	0.706	0.480

Table 10

SIBTEST Results for Individual Items: Science, Form 1

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	-0.026	-0.821	0.412	26	0.045	0.736	0.462
2	0.013	0.387	0.699	27	0.012	0.260	0.795
3	0.138	2.351	0.019	28	-0.027	-0.784	0.433
4	0.061	1.164	0.245	29	-0.006	-0.106	0.916
5	-0.020	-0.347	0.728	30	-0.063	-1.111	0.267
6	-0.039	-1.373	0.170	31	0.136	2.381	0.017
7	-0.005	-0.499	0.618	32	0.172	2.553	0.011
8	0.104	1.696	0.090	33	0.033	0.509	0.611
9	0.071	1.482	0.138	34	-0.104	-2.082	0.037
10	0.015	0.322	0.748	35	0.067	1.039	0.299
11	0.034	0.685	0.493	36	0.010	0.155	0.877
12	-0.058	-1.199	0.230	37	0.074	1.348	0.178
13	0.035	0.888	0.374	38	0.066	1.016	0.309
14	0.022	0.515	0.607	39	0.042	0.619	0.536
15	0.017	0.410	0.682	40	-0.164	-3.507	0.000
16	-0.031	-0.678	0.498	41	-0.032	-0.456	0.648
17	0.027	0.525	0.599	42	0.049	0.778	0.436
18	0.001	0.022	0.982	43	-0.029	-0.515	0.607
19	0.027	0.579	0.563	44	0.080	1.493	0.136
20	0.049	0.790	0.429	45	-0.052	-0.978	0.328
21	0.086	1.477	0.140	46	-0.050	-1.002	0.316
22	0.032	0.767	0.443	47	0.114	1.579	0.114
23	-0.046	-1.288	0.198	48	0.048	1.065	0.287
24	-0.035	-0.837	0.403	49	-0.018	-0.284	0.777
25	0.008	0.184	0.854	50	-0.060	-1.200	0.230

Table 11

SIBTEST Results for Individual Items: Science, Form 2

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.035	0.679	0.497	26	0.078	1.247	0.212
2	0.019	0.664	0.507	27	-0.054	-1.057	0.291
3	-0.042	-1.433	0.152	28	-0.018	-0.291	0.771
4	-0.006	-0.218	0.827	29	0.013	0.224	0.823
5	-0.065	-1.795	0.073	30	-0.060	-1.102	0.271
6	0.022	0.478	0.632	31	0.092	1.602	0.109
7	0.074	1.466	0.143	32	0.150	2.191	0.028
8	-0.037	-1.285	0.199	33	-0.003	-0.055	0.956
9	-0.022	-0.731	0.465	34	0.026	0.384	0.701
10	0.027	0.888	0.375	35	0.093	1.321	0.187
11	0.028	0.691	0.490	36	-0.058	-1.145	0.252
12	0.025	0.439	0.661	37	0.010	0.153	0.878
13	0.025	0.407	0.684	38	-0.073	-1.175	0.240
14	-0.017	-0.311	0.756	39	0.060	1.024	0.306
15	0.030	0.468	0.640	40	0.077	1.189	0.234
16	0.005	0.263	0.793	41	-0.052	-0.841	0.400
17	0.056	1.821	0.069	42	0.039	0.630	0.529
18	0.002	0.049	0.961	43	-0.051	-0.716	0.474
19	-0.003	-0.040	0.968	44	0.054	1.111	0.267
20	-0.039	-0.871	0.384	45	-0.011	-0.160	0.873
21	-0.041	-0.753	0.452	46	0.004	0.056	0.955
22	-0.037	-0.676	0.499	47	0.036	0.567	0.571
23	-0.069	-1.617	0.106	48	0.010	0.152	0.879
24	0.083	1.326	0.185	49	-0.099	-1.609	0.108
25	-0.010	-0.175	0.861	50	-0.145	-2.405	0.016

Table 12

SIBTEST Results for Individual Items: Science, Form 3

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.050	1.711	0.087	26	0.002	0.033	0.974
2	0.006	0.252	0.801	27	0.116	2.136	0.033
3	0.020	0.465	0.642	28	0.044	0.832	0.405
4	-0.020	-0.562	0.574	29	-0.120	-2.789	0.005
5	0.025	0.478	0.633	30	0.112	1.844	0.065
6	0.001	0.011	0.991	31	0.076	1.642	0.101
7	-0.029	-0.701	0.483	32	-0.027	-0.574	0.566
8	0.046	1.737	0.082	33	-0.022	-0.434	0.664
9	0.006	0.219	0.827	34	-0.095	-3.197	0.001
10	-0.013	-0.218	0.827	35	-0.105	-1.893	0.058
11	-0.042	-0.784	0.433	36	-0.037	-0.706	0.480
12	0.044	1.007	0.314	37	-0.047	-1.737	0.082
13	0.022	0.376	0.707	38	-0.063	-1.305	0.192
14	-0.008	-0.292	0.771	39	-0.017	-0.282	0.778
15	0.043	0.641	0.521	40	-0.063	-0.996	0.319
16	-0.051	-1.436	0.151	41	0.117	1.954	0.051
17	0.016	0.298	0.765	42	0.138	2.322	0.020
18	-0.008	-0.244	0.807	43	0.127	2.399	0.016
19	0.030	0.729	0.466	44	-0.063	-0.953	0.341
20	-0.045	-0.997	0.319	45	0.032	0.493	0.622
21	0.027	0.791	0.429	46	0.036	0.532	0.594
22	0.047	1.203	0.229	47	-0.106	-1.950	0.051
23	-0.017	-0.562	0.574	48	-0.060	-1.218	0.223
24	-0.009	-0.172	0.864	49	-0.074	-1.203	0.229
25	-0.039	-1.078	0.281	50	-0.026	-0.450	0.653

Table 13

SIBTEST Results for Individual Items: Language Arts, Reading, Form 1

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.012	0.292	0.770	21	0.173	3.464	0.001
2	-0.016	-0.450	0.653	22	0.036	0.741	0.459
3	0.017	0.429	0.668	23	-0.068	-2.052	0.040
4	0.054	1.423	0.155	24	0.042	1.306	0.192
5	-0.005	-0.544	0.586	25	0.015	0.413	0.679
6	0.024	0.667	0.505	26	0.000	0.014	0.989
7	0.029	0.672	0.501	27	-0.036	-2.043	0.041
8	0.005	0.154	0.878	28	-0.027	-0.570	0.569
9	-0.046	-1.719	0.086	29	-0.005	-0.129	0.897
10	-0.011	-0.389	0.697	30	-0.066	-1.839	0.066
11	0.029	0.719	0.472	31	0.004	0.101	0.919
12	0.022	0.377	0.706	32	0.083	1.762	0.078
13	0.070	1.354	0.176	33	0.041	1.294	0.196
14	0.020	0.462	0.644	34	0.057	1.091	0.275
15	-0.053	-1.552	0.121	35	-0.035	-0.839	0.401
16	-0.007	-0.179	0.858	36	-0.040	-1.427	0.154
17	-0.038	-0.993	0.321	37	-0.041	-1.351	0.177
18	-0.012	-0.360	0.719	38	0.027	0.607	0.544
19	-0.106	-3.224	0.001	39	0.054	0.885	0.376
20	-0.014	-0.317	0.751	40	-0.111	-4.588	0.000

Table 14

SIBTEST Results for Individual Items: Language Arts, Reading, Form 2

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.053	1.126	0.260	21	0.014	0.267	0.789
2	0.004	0.099	0.921	22	0.117	1.661	0.097
3	0.036	0.895	0.371	23	0.052	1.149	0.251
4	0.098	1.867	0.062	24	-0.008	-0.301	0.763
5	0.014	0.318	0.751	25	-0.038	-0.935	0.350
6	-0.058	-4.163	0.000	26	-0.028	-0.608	0.543
7	0.006	0.115	0.909	27	-0.055	-1.633	0.102
8	-0.026	-0.772	0.440	28	-0.008	-0.155	0.877
9	0.006	0.138	0.890	29	0.069	1.401	0.161
10	0.016	0.275	0.783	30	0.081	1.223	0.221
11	0.024	0.628	0.530	31	0.124	2.202	0.028
12	0.017	0.346	0.729	32	-0.003	-0.105	0.916
13	0.101	1.431	0.153	33	0.048	0.711	0.477
14	0.035	0.886	0.376	34	-0.086	-1.668	0.095
15	-0.033	-0.873	0.383	35	0.036	0.627	0.531
16	0.039	1.001	0.317	36	-0.032	-0.919	0.358
17	0.134	3.254	0.001	37	0.068	1.320	0.187
18	0.037	0.858	0.391	38	0.041	0.616	0.538
19	0.008	0.146	0.884	39	0.058	1.224	0.221
20	0.019	0.402	0.687	40	0.076	1.133	0.257

Table 15

SIBTEST Results for Individual Items: Language Arts, Reading, Form 3

Item	β_{UNI}	β_{UNI} Z	*p*	Item	β_{UNI}	β_{UNI} Z	*p*
1	0.008	0.266	0.790	21	-0.062	-1.246	0.213
2	0.037	0.880	0.379	22	0.036	0.878	0.380
3	-0.055	-0.816	0.415	23	0.084	1.452	0.146
4	-0.007	-0.122	0.903	24	0.018	0.380	0.704
5	-0.026	-0.599	0.549	25	0.080	1.528	0.126
6	0.018	0.687	0.492	26	0.018	0.365	0.715
7	0.028	0.569	0.570	27	0.052	0.896	0.370
8	0.162	2.972	0.003	28	0.067	1.466	0.143
9	0.020	0.652	0.514	29	-0.124	-2.515	0.012
10	-0.038	-0.719	0.472	30	0.013	0.207	0.836
11	0.008	0.274	0.784	31	-0.041	-1.175	0.240
12	0.007	0.588	0.556	32	-0.005	-0.101	0.920
13	0.007	0.141	0.888	33	-0.048	-1.632	0.103
14	-0.005	-0.164	0.870	34	-0.113	-2.566	0.010
15	0.038	0.917	0.359	35	-0.031	-0.640	0.522
16	-0.028	-0.808	0.419	36	0.052	0.796	0.426
17	0.055	1.242	0.214	37	-0.116	-2.483	0.013
18	0.035	0.691	0.490	38	-0.063	-1.755	0.079
19	0.102	2.105	0.035	39	0.000	-0.003	0.998
20	0.016	0.397	0.691	40	-0.054	-1.256	0.209

Table 16

Individual Items Flagged for DIF by SIBTEST

Test/Form	Item	Group Favored	Description of Item
Writing Form 1	27	Accommodated	Usage
Writing Form 2	6	Accommodated	Usage
Writing Form 3	18	Non-accommodated	Mechanics
Social Studies Form 2	16	Accommodated	Analyze map
Science Form 1	40	Accommodated	Make inference about graphic
Science Form 3	34	Accommodated	Analyze map
Reading Form 1	19	Accommodated	Reading comprehension
Reading Form 1	21	Non-accommodated	Analysis
Reading Form 1	40	Accommodated	Extended synthesis
Reading Form 2	6	Accommodated	Reading comprehension
Reading Form 2	17	Non-accommodated	Analysis

The Language Arts, Reading Test had the greatest number of flagged items; three items were flagged in favor of candidates with accommodations, and two items were flagged in favor of candidates testing under standard administration. All flagged items were attached to works of prose fiction (versus non-fiction or poetry). Two of the three questions that favored examinees with accommodations were reading comprehension items; the other was an extended synthesis question that required examinees to use additional information in the item stem and synthesize it with the passage information in order to arrive at a correct answer. Test specialists speculated that candidates with extended time might have been more likely to go back through the items and check the accuracy of comprehension and analysis items. However, the two items in favor of non-accommodated examinees were also analysis items, suggesting that additional time may have resulted in examinees' mistrusting or second-guessing their first interpretation of the passage and/or their initial answer to the item.

DIF by Subtest

In addition to running the SIBTEST procedure on individual items, DIF analyses were extended to clusters of items (subtests) grouped by content areas. Because the same content areas were covered by each form within a test, it was possible to examine the consistency of subtest DIF results across test forms. Tables 17 to 20 show the results of these analyses for each of the four tests.

The one consistent finding was DIF in favor of non-accommodated candidates for the Mechanics subtest in all three forms of the Language Arts, Writing Test. Test specialists hypothesized that candidates testing under standard administration may be more accustomed to the requirements of unassisted editing for capitalization, spelling, and punctuation. There were several other statistically significant, but inconsistent, findings. In Social Studies Form 1, the Civics and

Government subtest exhibited DIF in favor of non-accommodated candidates. In Science Form 2 the Life Science subtest exhibited DIF in favor of accommodated candidates, and in Science Form 3, the Physical Science subtest exhibited DIF in favor of non-accommodated candidates. In Language Arts, Reading Form 3, post-1960 fiction favored accommodated candidates, and drama favored non-accommodated candidates. The only pattern in the Language Arts, Reading Test DIF results was that the direction of DIF favored non-accommodated candidates in all three forms for poetry and drama, although the only statistically significant result was for drama on Form 3.

Table 17

SIBTEST Results for Subtests on the Language Arts, Writing Test

Subtest	Form	No. items	β_{UNI}	β_{UNI} Z	*p*
Mechanics	Form 1	11	.487	2.663	.008
	Form 2	9	.707	2.934	.003
	Form 3	11	.519	2.241	.025
Organization	Form 1	7	.014	.087	.931
	Form 2	5	-.176	-1.309	.190
	Form 3	7	-.050	-.217	.828
Sentence Structure	Form 1	17	.080	.305	.760
	Form 2	18	-.025	-.079	.937
	Form 3	17	.266	.928	.353
Usage	Form 1	15	-.133	-.609	.542
	Form 2	18	-.063	-.240	.810
	Form 3	15	-.471	-1.738	.082

Table 18

SIBTEST Results for Subtests on the Social Studies Test

Subtest	Form	No. items	β_{UNI}	β_{UNI} Z	*p*
Civics & Government	Form 1	11	.331	2.224	.026
	Form 2	11	-.026	-.135	.892
	Form 3	8	.111	.699	.485
Economics	Form 1	11	-.168	-1.128	.259
	Form 2	10	.169	.956	.339
	Form 3	14	-.161	-.753	.452
Geography	Form 1	8	.009	.071	.943
	Form 2	8	-.211	-1.097	.273
	Form 3	7	.218	1.432	.152
National History	Form 1	12	.244	1.203	.229
	Form 2	15	.197	.885	.376
	Form 3	12	-.073	-.267	.789
World History	Form 1	8	-.211	-1.413	.158
	Form 2	8	-.247	-1.743	.081
	Form 3	9	.166	.962	.336

Table 19

SIBTEST Results for Subtests on the Science Test

Subtest	Form	No. items	β_{UNI}	β_{UNI} Z	*p*
Earth Science	Form 1	10	.302	1.902	.057
	Form 2	9	.082	.596	.551
	Form 3	10	-.271	-1.516	.129
Life Science	Form 1	23	-.053	-.170	.865
	Form 2	23	.114	.389	.697
	Form 3	23	.090	.304	.761
Physical Science/Chemistry	Form 1	8	.137	.983	.326
	Form 2	9	.032	.125	.901
	Form 3	9	.410	2.744	.006
Physical Science/Physics	Form 1	9	.078	.448	.654
	Form 2	9	-.361	-2.149	.032
	Form 3	8	-.024	-.125	.900

Table 20

SIBTEST Results for Subtests on the Language Arts, Reading Test

Subtest	Form	No. items	β_{UNI}	β_{UNI} Z	*p*
Poetry	Form 1	6	.076	.911	.362
	Form 2	5	.063	.565	.572
	Form 3	5	.207	1.755	.079
Fiction: Pre-1920	Form 1	11	-.018	-.116	.908
	Form 2	8	.007	.031	.975
	Form 3	-	-	-	-
Fiction: 1920-1960	Form 1	-	-	-	-
	Form 2	5	-.054	-.401	.689
	Form 3	10	.035	.224	.823
Fiction: Post-1960	Form 1	7	-.152	-1.099	.272
	Form 2	6	.072	.609	.543
	Form 3	7	-.341	-2.240	.025
General Non-fiction	Form 1	6	.004	.033	.974
	Form 2	6	.107	.804	.421
	Form 3	6	-.020	-.174	.862
Business	Form 1	4	.092	1.203	.229
	Form 2	5	-.007	-.055	.956
	Form 3	-	-	-	-
Drama	Form 1	6	.099	.648	.517
	Form 2	5	.122	.953	.340
	Form 3	6	.246	2.175	.030
Non-fiction: reference	Form 1	-	-	-	-
	Form 2	-	-	-	-
	Form 3	6	.100	.937	.349

Discussion

This study contributed to the growing literature on the effect of test accommodations. Specifically, the findings from this study supplemented research results on whether schedule or setting accommodations have an effect on test scores using a sample of adult examinees. The study examined the effect of scheduling or setting accommodations on GED Tests scores in four content areas by evaluating raw test score statistics, reliability estimates, and DIF analyses.

Examination of raw score statistics indicated that although the sample of GED candidates testing under standard administration procedures consistently (across all content area tests and test forms) achieved a higher average raw score than the sample testing under scheduling or setting accommodations, the differences in average scores were small, ranging in size from one-tenth to one-third of a standard deviation. Reliability estimates also indicated small differences in the values of K-R 20s and SEMs between the accommodated and non-accommodated samples. These small empirical differences in raw scores and reliability estimates provided evidence supporting the validity and comparability of test scores obtained under test scheduling/setting accommodations.

DIF analyses on the individual item responses of candidates testing under scheduling or setting accommodations and standard administration procedures flagged 11 items exhibiting substantial DIF. Five of the 11 flagged items were found across two of the three Language Arts, Reading Test forms, and three of the items were found across the three Language Arts, Writing Test forms. Less than one percent of any test form's items were flagged for exhibiting substantial DIF. In content areas where more than one item was flagged for DIF, the number of items favoring one group versus the other was nearly equal. Furthermore, test content specialists were sometimes able to hypothesize about item characteristics that might have advantaged or disadvantaged one group over the other. However, the small number of flagged items and the variability of the characteristics the items caution that these hypotheses require further research. Even discussion of the results of DIF analyses on content area subtests, where the Language Arts, Writing Tests' Mechanics subtest was consistently flagged for DIF in favor of candidates testing under standard administration, test content specialists were tentative about possible reasons for such differences.

Limitations of this study are connected to the definition of inclusion in the accommodated group and sample comparability. The accommodated group included candidates with either single

or multiple scheduling or setting accommodations. Further, multiple-accommodation administrations sometimes involved the use of a setting accommodation (private room) in addition to scheduling accommodation(s) (extended time or breaks). Had the data been analyzed using only single-accommodation data or using only scheduling accommodations, results may have differed. Comparability of the accommodated and non-accommodated groups is questionable because nearly half of the candidates in the accommodated group did not respond to the demographic questions of age, gender, race or ethnicity, and highest educational level achieved. Analysis of non-missing responses showed group differences in age, gender, ethnicity, and highest educational level achieved, several of which are demographic characteristics that may influence performance on an educational achievement test and, therefore, affect the results of this study. As much as previous aggregate test accommodation research results are inconsistent, this study provided partial evidence of the comparability of test scores from a set of achievement tests administered under frequently used scheduling and setting accommodations.

In conclusion, the results of this study provided support that GED Tests scores in writing, reading, social studies, and science show evidence of validity under test accommodations of (a) extended time only, (b) extended time and private room only, or (c) extended time, private room, and supervised breaks only. Further research on the validity of GED Tests scores under test accommodations is recommended and should attempt to address this study's limitations, particularly the relationship of sample demographics to performance.

References

Abedi, J., Hofstetter, C., & Lord, C. (2003). *Research on accommodations for English language learners*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, *19*(3), 16-26.

American Council on Education. (2002a). *Tests of General Educational Development*. Washington, DC: American Council on Education.

American Council on Education. (2002b). *Who took the GED? GED 2001 statistical report*. Washington, DC: American Council on Education.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Cahalan, C., Mandinach, E., & Camara, W. (2002). *Predictive validity of SAT I: Reasoning Test for test takers with learning disabilities and extended time accommodations* (College Board Research Report RR 2002-05). New York: College Board.

Camera, W., & Schneider D. (2000). *Testing with extended time on the SAT I: Effects for students with learning disabilities* (College Board Research Note RN-08). New York: College Board.

Chiu, C. W. T., & Pearson, P. D. (1999). *Synthesizing the effects of test accommodations for special education and limited English proficiency students*. Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT.

Elliott, S., Kratochwill, T., & McKevitt, B. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology*, *39*, 3-24.

Fuchs (2000). *The validity of test accommodations for students with learning disabilities: Differential item performance on reading and mathematics tests as a function of test accommodations and disability status*. Retrieved January 29, 2004, from Delaware Department of Education, Delaware Student Testing Program Research Web site: http://www.doe.state.de.us/aab/DSTP_research.html

Jöreskog, K., & Sörbom, D. (2002). *LISREL 8.52: User's reference guide*. Chicago: Scientific Software International, Inc.

Marquart, A. (2000). *The use of extended time as an accommodation on a standardized mathematics test: An investigation of effects on scores and perceived consequences for students of various skill levels*. Paper presented at the annual meeting of the Council of Chief State School Officers, Snowbird, UT.

Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika, 58,* 159-194.

Sireci, S., Li, S., & Scarpati, S. (2003). *The effects of test accommodation on test performance: A review of the literature*. University of Massachusetts, Amherst: Center for Educational Assessment Research Report no. 485.

Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (Technical Report 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved February 12, 2004, from http://education.umn.edu/NCEO/OnlinePubs/Technical34.htm

Thurlow, M. L., McGrew, K. S., Tindal, G., Thompson, S. L., Ysseldyke, J. E., & Elliott, J. L. (2000). *Assessment accommodations research: Considerations for design and analysis* (Technical Report 26). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Walz, L., Albus, D., Thompson, S., & Thurlow, M. (2000, December). *Effect of a multiple day test accommodation on the performance of special education students* (Minnesota Report 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved February 12, 2004, from http://education.umn.edu/NCEO/OnlinePubs/MnReport34.html

Wilson, D., Wood, R., & Gibbons, R. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis*. Chicago: Scientific Software, Inc.

Zuriff, G.E. (2000) Extra examination time for students with learning disabilities: An examination of the maximum potential thesis, *Applied Measurement in Education*, *13*(1), 99-117.

GED® Testing Service
One Dupont Circle NW, Suite 250
Washington, DC 20036-1163
(202) 939-9490
Fax: (202) 659-8875
www.gedtest.org