DOCUMENT RESUME

ED 432 592 TM 029 969

AUTHOR	McLeod, Lori D.; Schnipke, Deborah L.
TITLE	Detecting Items That Have Been Memorized in the Computerized Adaptive Testing Environment.
PUB DATE	1999-04-00
NOTE	23p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, Canada, April 19-23, 1999). Some figures may not reproduce clearly.
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Adaptive Testing; *Cheating; *Computer Assisted Testing; *Identification; *Item Banks; *Memorization; Test Items
IDENTIFIERS	High Stakes Tests; *Test Security

ABSTRACT

Because scores on high-stakes tests influence many decisions, tests need to be secure. Decisions based on scores affected by preknowledge of items are unacceptable. New methods are needed to detect the new cheating strategies used for computer-administered tests because item pools are typically used over time, providing the potential opportunity for test takers to share items with future test takers. Because of the serious ramifications of accusing someone as being a user of item preknowledge (or "cheater"), it may be more useful for operational computer-administered test developers to focus on item security rather than the behavior of individual test takers. This research explores the development and use of a fit index to detect items that have been memorized so that these items may be removed from the item pool, while leaving secure items in the pool. The results from this initial simulation for the developed Bayesian posterior log odds ratio index are promising. It is hoped that this work and future work will enable testing programs to determine more effectively how long to leave an item pool (or specific items) in the field. (Contains four tables, five figures, and five references.) (Author/SLD)

Detecting Items That Have Been Memorized

In The Computerized Adaptive Testing Environment

Lori D. McLeod
Deborah L. Schnipke

Law School Admission Council

Paper presented at the Annual Meeting of National Council on Measurement in Education, Montreal, CA, April 19-23, 1999.

Abstract

Because scores on high-stakes tests influence many decisions, tests need to be secure. Decisions based on scores affected by preknowledge of items are unacceptable. New methods are needed to detect the new cheating strategies used for computer-administered tests because item pools are typically used over time, providing the potential opportunity for test takers to share items with future test takers. Because of the serious ramifications of accusing someone as being a user of item preknowledge (or "cheater"), it may be more useful for operational computer-administered test developers to focus on item security rather than the behavior of individual test takers. This research explores the development and use of a fit index to detect items that have been memorized so that these items may be removed from the item pool, while leaving secure items in the pool. The results from this initial demonstration are promising. It is hoped that this work and future work will enable testing programs to more effectively determine how long to leave an item pool (or specific items) in the field.

Detecting Items That Have Been Memorized

In The Computerized Adaptive Testing Environment

Scores on high-stakes tests influence many decisions such as which universities accept a particular student and whether a candidate is certified for employment. Tests should be fair, reliable, and valid, and to maintain these characteristics, tests need to be secure. Decisions based on scores affected by cheating are unacceptable. Methods for detecting cheating in paper-and-pencil tests exist, but new methods are needed for computer-administered tests. As the use of computers to administer tests increases, the importance of developing new methods for detecting cheating in this new environment will also increase.

Two of the potential advantages for using computers to administer tests include adaptive testing and the opportunity for continuous testing, but these features also create new security risks. In computerized testing, a few quarterly test forms may be replaced by an item pool from which tailored forms are individually administered to test takers on a nearly continuous basis. Daily access to testing is made possible. However, with the daily access to the item pool, security becomes a real concern. Test takers may memorize blocks of test items and share these items with future test takers. Individuals with prior knowledge of some items may use that information to inflate their test scores. Because adaptive tests are tailored to the test taker and consistently answering items correctly increases the test taker's estimated proficiency level and the difficulty of the items administered to that person, memorizing middle- to high-difficulty items exaggerates the amount of test-score inflation. If the item pool is not very large and does not include many difficult items, test takers with prior knowledge of some of the more difficult items may have a particular advantage.

Because of the high-stakes issues of accusing someone as being a user of item preknowledge (or "cheater") it may be more useful for operational computerized adaptive test (CAT) developers to focus on item security rather than the behavior of individual test takers. Some researchers advocate frequently changing the item pools to lessen the security problem. However, just with the addition of adaptive testing, the demand on item writers has already increased. Asking item writers and test developers to provide even more items may degrade some of the measurement properties of items. There is a strong need for a quality control tool to measure the "freshness" of item pools so that compromised items may be removed more efficiently. Rather than replacing an entire item pool, the flagged items may be removed and replaced with fresh (secure) items. The new system may allow secure items to stay in the pool longer, thus decreasing the demands of maintaining a fresh item pool. This research explores the development and use of a fit index to accomplish the task of detecting items that have been memorized so that these items may be removed from the item pool, while leaving secure items in the pool. It is hoped that this work will enable testing programs to more effectively determine how long to leave an item pool (or specific items) in the field.

There are several kinds of indices that have been shown to be useful for detecting response-pattern misfit or lack of person-fit. When test takers use item preknowledge, their item responses deviate from the underlying item response theory (IRT) model, and their estimated abilities may be inflated. This deviation (or lack of fit) may be detected through the use of person-fit indices. One index that has shown power to detect item preknowledge is FLOR3, a Bayesian posterior log odds ratio index (McLeod & Lewis, 1998).

This study explores the use of a log odds ratio index for detecting items that have been memorized, rather than test takers who have memorized items. The new index, $FLOR_i$, is an extension of FLOR3 (McLeod & Lewis, 1998). In this new approach, test takers are detected as either test takers using item preknowledge (the memorizer group) or not (the null, or non-memorizer, group). Then, for each item administered, an "item-fit" index ($FLOR_i$, discussed below) is computed based on item responses (correct or incorrect) from the two groups.

Specifically, two probabilities are empirically computed for each item: the probability for a correct response for the memorizer and non-memorizer groups. Although the difference between these two probabilities may be used as a qualitative measure of difference in item performance, a more appropriate measure is the odds ratio between the two groups because this measure provides a convenient quantitative scale. For numerical and interpretive convenience, the log (base 10) of the odds ratio is used for analyses. The log odds ratio index is the item-fit index ($FLOR_i$) and is given by:

$$\mathrm{FLOR_i} = \log_{10}\left[\frac{p(u=1\mid m)/[1-p(u=1\mid m)]}{p(u=1\mid n)/[1-p(u=1\mid n)]}\right] \qquad [1]$$

where $p(u = 1 \mid m)$ is the proportion of correct responses in the memorizer group, $p(u = 1 \mid n)$ is the proportion of correct responses in the non-memorizer group, and u represents the item response (1 indicates a correct response and 0 indicates an incorrect response).

For example, a $FLOR_i$ value of 0 indicates that the odds of a correct response on the item was equal for the two groups, thus providing no evidence that the item has been compromised based on the test takers' responses. A $FLOR_i$ value of 1 implies that test takers from the memorizer group had 10 times more chance of getting this item correct than test takers from the other group. It follows that with a $FLOR_i$ value of −1, test takers from the memorizer group had 10 times less chance of getting this item correct than those test takers from the other group. In general, a large positive value for the log odds ratio provides evidence that the item has been compromised.

Method

This project used a basic CAT design. Simulated test takers were generated at 13 proficiency (θ) values, ranging from –3.0 to 3.0, in increments of 0.5. At each true θ-level, 1000 simulated test takers were generated for the non-memorizer group and 1000 simulated test takers were generated for each of the 8 memorizing-group conditions described below. For each simulated test taker, a 25-item CAT was simulated using a pool of 250 items. Items were selected based on maximum information along with conditional item exposure control (Stocking & Lewis, 1995). The maximum desired item exposure rate was 0.2. The item pool consisted of items based on the three-parameter logistic (3PL) IRT model (Birnbaum, 1968). Item discrimination, difficulty, and lower asymptote parameters were simulated for each item. Item discrimination parameters were simulated from a normal distribution with a mean of 1.0 and standard deviation of 0.2. Difficulty parameters were simulated from a normal distribution with a mean of 0.0 and a standard deviation of 1.0. Lower asymptotes were drawn from a uniform distribution ranging from 0.00 to 0.25. Proficiency estimates were obtained using Bayes modal scoring.

For the memorizing conditions, if a simulated test taker is administered one of the memorized items, a correct response is automatically given. The 3PL IRT model is used to generate a response when a non-memorized item is administered. The simulation design uses a real-world approach to simulate the "item preknowledge" process by incorporating a two-stage process much like the method used in Schnipke and Scrams (1998) and McLeod and Lewis (1998). First, for each condition, *n* "sources" (thieves) take the test and memorize their items. These *n* test takers memorize their items perfectly and combine their lists. (Some overlap is observed among the lists.) "Beneficiaries" memorize the complete list. Then, these informed test takers (the beneficiaries) are administered a 25-item test, and if they receive any of the memorized items, they answer them correctly. (Although we acknowledge that beneficiaries may not have perfect recall of the item list, the simulation is designed to produce a worst case scenario for the testing program.) By varying the number of sources and the sources' proficiency levels, we indirectly manipulate the number and difficulty of items memorized. The impact of item preknowledge on the overall testing program is also manipulated by varying the percent of beneficiaries at each condition (10 or 25%).

Table 1 shows the source-beneficiary conditions in the design. Eight memorizing conditions were formed by completely crossing the number of sources (2 or 6), the source-proficiency distribution [N(0,1) or N(1.5, .5)], and the percent of test takers who were beneficiaries (10 or 25%). The source-proficiency distributions simulated "regular" sources and "professional" sources. Regular sources had the same proficiency distribution as the other test takers (i.e., they were simulated to be regular test takers), and professional sources were more proficient on average (i.e., they were simulated to be trained professionals, specifically attempting to receive and memorize the difficult items). An additional condition was generated in which there were no sources (thus no compromised item); this condition served as a baseline or control condition. The complete design was replicated 10 times; results are averaged across the 10 replications.

Table 1: Conditions for the research design.

Number of Sources	Source Proficiency Level	Percent Beneficiaries
0	—	—
2	Regular: $N(0,1)$	10
2	Regular: $N(0,1)$	25
2	Professional: $N(1.5,0.5)$	10
2	Professional: $N(1.5,0.5)$	25
6	Regular: $N(0,1)$	10
6	Regular: $N(0,1)$	25
6	Professional: $N(1.5,0.5)$	10
6	Professional: $N(1.5,0.5)$	25

For each simulated test taker, FLOR3 (McLeod & Lewis, 1998) was calculated based on the responses to the twenty-five items administered. FLOR3 is a final log odds ratio index that measures the degree of suspicion that a test taker is using item preknowledge on a similar scale as $FLOR_i$ described previously. In general, a positive FLOR3 value provides evidence that the test taker is using item preknowledge.

The proposed $FLOR_i$ index was calculated for each item. The simulated test takers were divided into two groups for the $FLOR_i$ calculation based on their FLOR3 values. If FLOR3 was positive, the test taker was assigned to the memorizer group; otherwise the test taker was assigned to the non-memorizer group. The low cut value for the memorizer group ensured useful proportions of test takers in each group.[1] (See Table 2 for the number of test takers in each group, combined across the 10 replications.) More test takers at the higher proficiency levels were assigned to the memorizer group than at the lower proficiency levels despite the fact that the proportion of test takers who were memorizers was the same at all proficiency levels. This implies that we are more suspicious of test takers who score high on the test, in general.

Table 2: Number of simulated test takers categorized as memorizers (labeled Mem) and non-memorizers (labeled Null) by the number (2 or 6) and type [none, regular (middle proficiency), or professional (high proficiency)] of sources and by true proficiency level.

		True Proficiency												
Sources	True %	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.00	1.5	2.0	2.5	3.0
None														
Mem	0	39	62	117	226	376	350	425	691	1199	2286	4603	7220	8950
Null	100	9961	9938	9883	9774	9624	9650	9575	9309	8801	7714	5397	2780	1050
2 Regular														
Mem	10	365	375	395	446	535	438	472	695	1214	2260	4655	7170	8938
Null	90	9635	9625	9605	9554	9465	9562	9528	9305	8786	7740	5345	2830	1062
2 Regular														
Mem	25	710	714	730	727	797	719	716	848	1421	2600	4814	7236	9000
Null	75	9290	9286	9270	9273	9203	9281	9284	9152	8579	7400	5186	2764	1000
6 Regular														
Mem	10	316	342	382	444	587	543	601	850	1364	2423	4778	7260	8969
Null	90	9684	9658	9618	9556	9413	9457	9399	9150	8636	7577	5222	2740	1031
6 Regular														
Mem	25	844	847	839	876	980	827	811	992	1524	2741	4889	7387	9054
Null	75	9156	9153	9161	9124	9020	9173	9189	9008	8476	7259	5111	2613	946
2 Professional														
Mem	10	163	177	264	368	525	562	693	961	1494	2568	4835	7361	9004
Null	90	9837	9823	9736	9632	9475	9438	9307	9039	8506	7432	5165	2639	996
2 Professional														
Mem	25	390	431	491	627	770	865	996	1207	1786	2991	5047	7517	9085
Null	75	9610	9569	9509	9373	9230	9135	9004	8793	8214	7009	4953	2483	915
6 Professional														
Mem	10	633	592	684	798	1013	1116	1225	1415	1904	2897	5086	7446	9040
Null	90	9367	9408	9316	9202	8987	8884	8775	8585	8096	7103	4914	2554	960
6 Professional														
Mem	25	1369	1433	1455	1548	1893	2091	2269	2375	2829	3872	5638	7784	9189
Null	75	8631	8567	8545	8452	8107	7909	7731	7625	7171	6128	4362	2216	811

Results are summarized in terms of the

- number of compromised (stolen) items,
- amount of test-score inflation due to memorizing items,

[1] If using FLOR3 for detecting suspicious test takers, we recommend using a more conservative cut value such as 3. However, for this study FLOR3 was only used to facilitate the calculation of $FLOR_i$ and therefore a lenient cut value was used.

- distributional characteristics of $FLOR_i$,
- detection rates for the compromised items, and
- power of $FLOR_i$ to detect compromised items.

Table 3 shows the mean and standard deviation of the number of compromised items for each condition. On average two sources (either professional or regular sources) memorized 46 items. Six sources gathered about 100 items. The number of sources was more important than the proficiency of the sources for the number of items compromised.

Table 3: Mean and standard deviation of the number of items compromised by the number of sources and source proficiency level. (Test length was 25 items.)

Number of Sources	Source Proficiency Level	Mean Number of compromised items	Standard Deviation Number of compromised items
2	Regular: $N(0,1)$	45.9	2.9
6	Regular: $N(0,1)$	108.9	5.8
2	Professional: $N(1.5,0.5)$	45.8	2.0
6	Professional: $N(1.5,0.5)$	95.1	5.4

Figures 1 and 2 show the impact of using preknowledge of the compromised items on proficiency estimation. Specifically, Figure 1 shows estimated proficiency by true proficiency for memorizers only, separately for 2 or 6 regular or professional sources, averaged across all 10 replications. The no-source condition is also shown in Figure 1 to provide a baseline. As shown in Figure 1, when there were no sources, proficiency was slightly overestimated for the lowest proficiency levels and slightly underestimated for the highest proficiency levels; this is due to using Bayesian scoring. When there were 2 regular or professional sources, proficiency was overestimated for all but the highest proficiency levels. When there were 6 regular or professional sources, proficiency was greatly overestimated for test takers at the lowest proficiency levels.

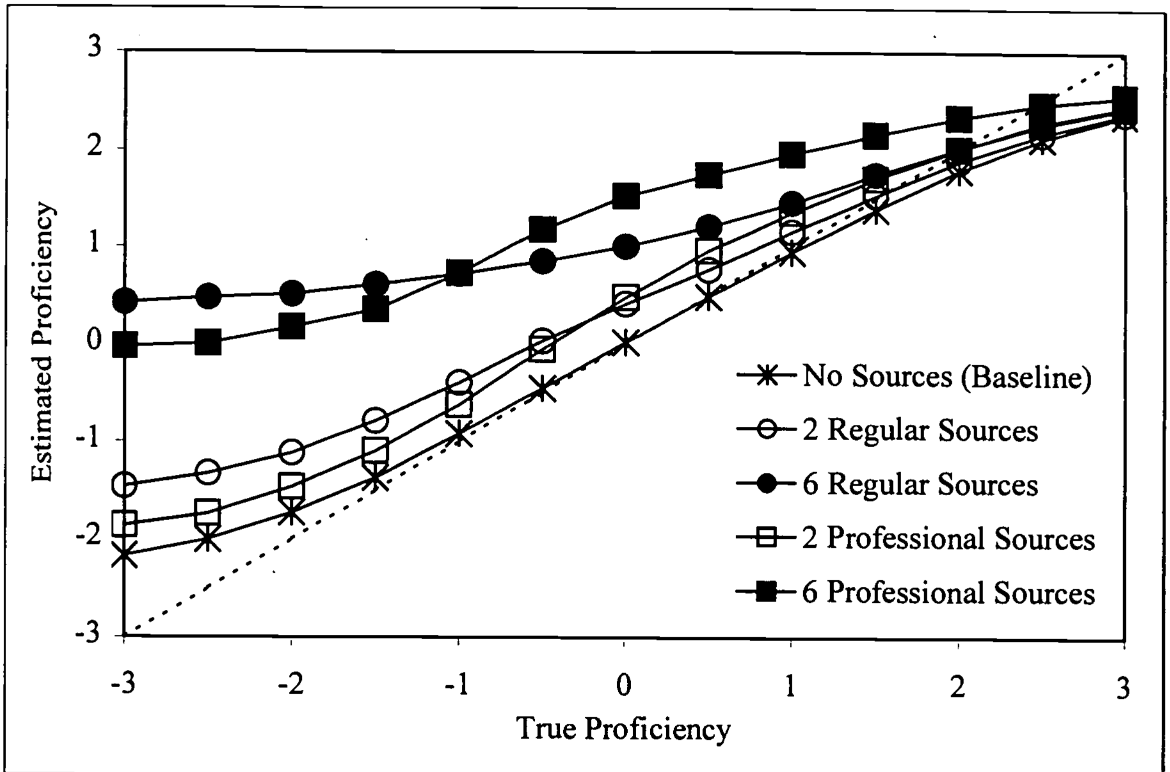

Figure 1. Estimated by true proficiency for the baseline condition (no sources, thus no compromised items) and for simulated test takers who had preknowledge of items (the beneficiaries).

Figure 2 shows the bias attributable to item preknowledge, after removing the bias attributable to Bayesian estimation. Specifically, Figure 2 shows the difference between the estimated proficiency for each condition and the baseline condition (no sources) for each true proficiency level, averaged across all 10 replications. As shown in Figure 2, more sources (and hence more compromised items) generally results in higher proficiency estimates. Preknowledge of items stolen by regular sources (mostly middle-difficulty items) most helps test takers with lower proficiency, and preknowledge of items stolen by professional sources (mostly difficult items) most helps test takers with higher proficiency.

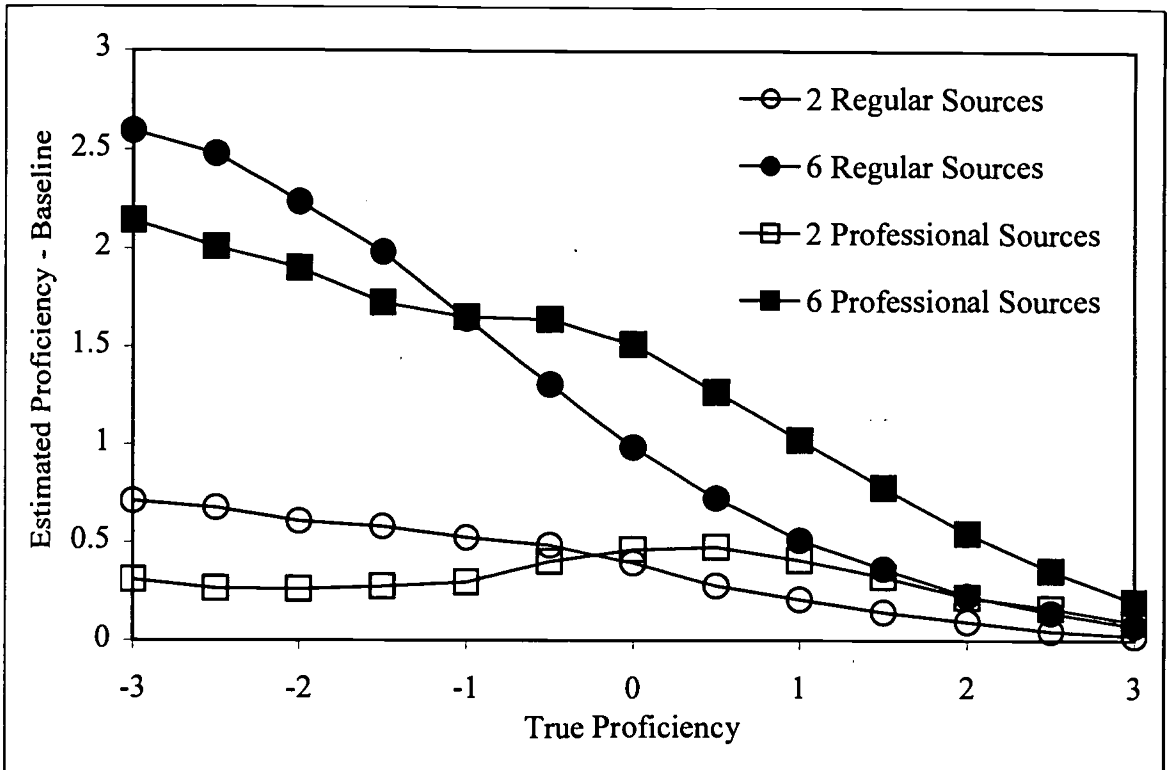

Figure 2. Bias in proficiency estimation attributable to item preknowledge, after removing bias attributable to Bayesian estimation (i.e., estimated proficiency for each condition minus estimated proficiency for the baseline condition), at each true proficiency level.

Distributional characteristics of the $FLOR_i$ index and the proportion of simulated items identified by $FLOR_i$ were studied under the various design conditions. The distributions of the $FLOR_i$ values by condition are displayed as boxplots in Figure 3. The top, bottom, and middle lines through each box correspond to the 75th percentile, 25th percentile, and the 50th percentile (or median) of each distribution, respectively. The end of the top whisker shows the 90th percentile and the end of the bottom whisker represents the 10th percentile. For each condition, the left box plot shows the distribution for the secure items; these items were not memorized. The right box plot shows the distribution for the items that were compromised (i.e., memorized by the sources and provided to the beneficiaries). The number of items in each category is shown below each box. The total number of items for each condition was 2500 (250 per replication for 10 replications).

As shown in Figure 3, $FLOR_i$ values for the secure items for each condition are more variable and lower on average than those for the compromised items. (Lower $FLOR_i$ values indicate a smaller chance that the item was compromised.) More separation is present for the professional-source conditions than for the regular-source conditions and for the 25% beneficiary conditions than for the 10% beneficiary conditions. $FLOR_i$ shows the largest separation between groups for the condition with 6 professional sources and 25% beneficiaries. The most overlap is present for the condition with 2 regular sources and 10% beneficiaries.

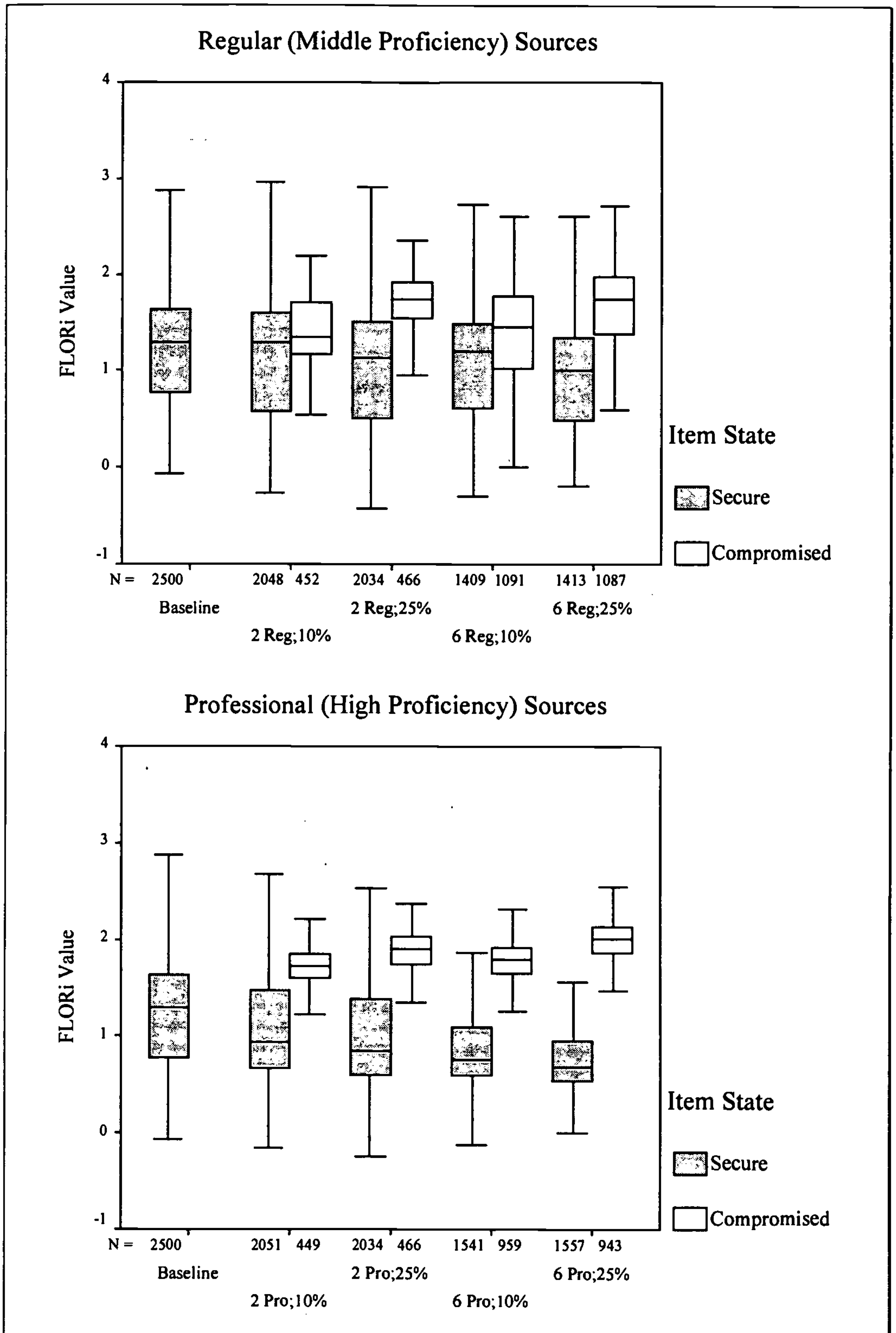

Figure 3. Distribution of $FLOR_i$ values for secure and compromised items by condition. Top panel shows baseline (no sources) and regular-source conditions. Bottom panel shows baseline and professional-source conditions.

Table 4 shows the percentage of compromised items that was detected as compromised by $FLOR_i$ (i.e., the hit rate), using either selected $FLOR_i$ cutoff values (2.5, 2.0, 1.5, or 1.0) or empirical cutoff values for $FLOR_i$ based on false-alarm rates as close to 0.1, 1.0, 5.0, or 10.0 without going over (producing cutoffs of 2.97, 2.50, 1.93, and 1.82, respectively). The false alarm rate is the percentage of secure (uncompromised) items (from the baseline condition with no sources) identified as compromised; it is also shown in Table 4 (last row). For example, using a critical value 2.0, $FLOR_i$ detected 27.9% of the compromised items in the 2-professional source/25% beneficiaries condition and 3.7% of the secure items from the baseline condition. (A value of 2.0 implies that the odds for a correct response for the memorizer group are 100 times that for the non-memorizer group.) $FLOR_i$ detected 15.0% of the compromised items at the same critical value of 2.0, again with a 3.7% false-alarm rate for the 2-regular source/25% beneficiaries condition. In general, $FLOR_i$ shows more power when the sources have higher proficiency, when the percentage of beneficiaries is higher, and when there are more sources.

Table 4. Percentage of compromised items identified as compromised by $FLOR_i$ (i.e., the hit rate of $FLOR_i$) using various cutoff values for $FLOR_i$. The false alarm (FA) rate (uncompromised items identified as compromised) is shown in the last row of the table.

Sources	% Beneficiaries	Selected $FLOR_i$ Cutoffs 2.5	2.0	1.5	1.0	$FLOR_i$ Cutoffs Based on False Alarm Rates 2.97 (FA≤0.1)	2.50 (FA≤1)	1.93 (FA≤5)	1.82 (FA≤10)
2 Reg	10	0.4	3.1	41.2	89.6	0.0	0.4	5.8	13.3
2 Reg	25	0.4	15.0	79.8	99.4	0.0	0.4	26.4	40.3
2 Prof	10	0.2	9.4	84.4	100	0.0	0.2	13.4	32.1
2 Prof	25	2.1	27.9	95.1	99.6	0.0	2.1	42.5	63.9
6 Reg	10	0.3	5.7	49.0	78.6	0.0	0.3	9.4	20.0
6 Reg	25	0.9	23.1	65.8	96.4	0.1	0.9	32.3	44.0
6 Prof	10	2.2	14.7	89.5	99.8	0.1	2.2	23.6	45.0
6 Prof	25	3.0	52.7	98.2	100	0.4	3.0	65.6	81.5
None	0	0.96	3.7	40.3	60.7	0.08	0.96	4.8	9.2

ROC curve analysis

Marginal probability receiver operating characteristic (ROC) curves (Green & Swets, 1966) offer a more detailed evaluation of the $FLOR_i$ item-fit index. The points on the ROC curve represent the hit rate for a fixed false-alarm rate; thus ROC curves provide a visual tool for assessing the power of the $FLOR_i$ index in the simulated CAT environment. Figures 4 and 5 show empirical ROC curves for the professional- and regular-source conditions, respectively. For each point on the ROC curve, the value on the horizontal axis is the proportion of secured items (falsely) identified as compromised (i.e., the false-alarm rate), and the value on the vertical axis is the proportion of compromised items identified as compromised (i.e., the hit rate). An index operating only by chance would produce a curve on the diagonal in the Figures. Indices that perform well produce curves in the upper left-hand corner.

Figure 4 shows the ROC curves for the professional-source conditions (beneficiaries received memorized items from sources with relatively high proficiency). As expected, the curves indicate that $FLOR_i$ performed best for higher percentages of beneficiaries and for larger numbers of sources (and hence compromised items).

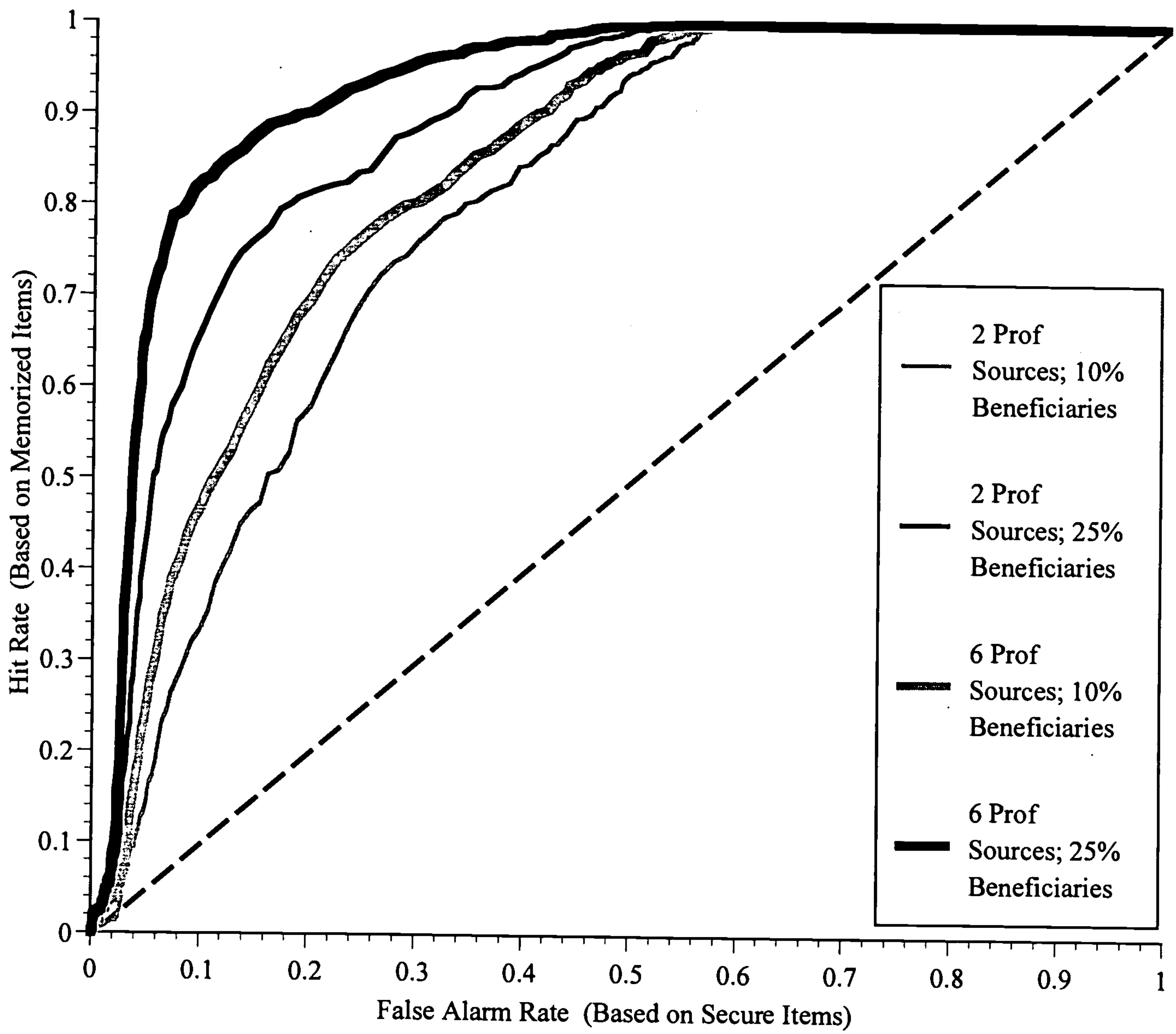

Figure 4. Receiver operating characteristic (ROC) curves: proportion of compromised items (correctly) identified as compromised (hit rate) by proportion of secure items (incorrectly) identified as compromised (false alarm rate) for professional (high proficiency) source conditions.

Figure 5 shows the ROC curves for the regular-source conditions (beneficiaries received memorized items from sources with average proficiency). The curves indicate that $FLOR_i$ did not show as much power for the regular-source conditions, especially when the percentage of beneficiaries was only 10%. In comparison to the professional-source conditions, for a 5 percent false-alarm rate, slightly over 32% of the compromised items are detected using $FLOR_i$ in the 6 regular-source, 25% beneficiary condition, as opposed to the 65% detected for the same professional-source condition. The curve for the 2 regular-source, 10% beneficiary condition even approaches chance-level detection between 40-50% false-alarm rate. Although Figure 5 shows that $FLOR_i$ has less power overall for the regular-source conditions, it is consistent with the results for the professional-source conditions in that $FLOR_i$ shows more power for the 25% beneficiaries conditions than for the 10% beneficiaries conditions.

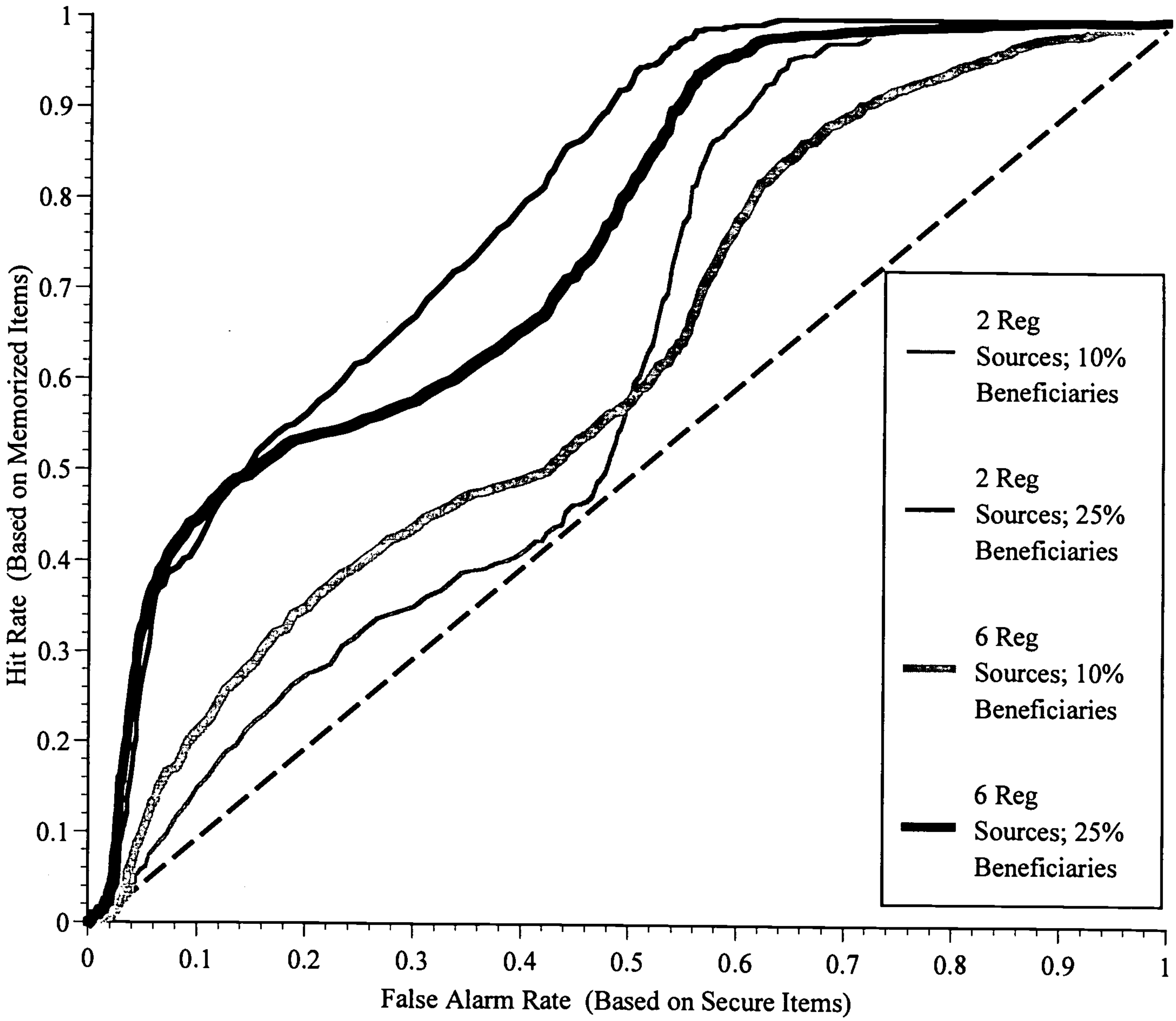

Figure 5. Receiver operating characteristic (ROC) curves: proportion of compromised items (correctly) identified as compromised (hit rate) by proportion of secure items (incorrectly) identified as compromised (false alarm rate) for regular (middle proficiency) source conditions.

Overall, the preliminary results for $FLOR_i$ show some success at detecting compromised items. It is hoped that future refinements will increase its power.

Discussion

With the increased use of computerized testing and daily access to items there is a strong need for tools that help maintain test security. There are policy issues that need to be considered when a test taker is suspected of using item preknowledge. Testing companies may want to supplement procedures for identifying text takers who may have used preknowledge with procedures that identify items that may have been compromised. Such items can be removed from the item pool so that additional test takers are not given the chance to answer them correctly with preknowledge. It is our hope to provide a quality-control tool that will enable testing programs to use their item pools more efficiently. The item-fit tool that we are proposing will flag items that are no longer secure. Flagged items may then be replaced by fresh items. Other items may remain in the item pool longer than current policies recommend, thus decreasing the demand on item writers and test developers.

The results of this study show that test takers may be very successful at score inflation when using item preknowledge gathered by six sources and somewhat successful when using item preknowledge gathered by two sources. The impact of the sources' proficiency is not as important as the number of sources (and therefore, the number of compromised items) for score inflation. These results are consistent with previous studies that investigated the use of item preknowledge (e.g., Schnipke & Scrams, 1998; McLeod & Lewis, 1998).

The main goal of this paper is to present a new index for detecting items that have been compromised. These preliminary results indicate that the new index, $FLOR_i$, does show some promise for detecting compromised items. The results also show that the proficiency of the sources and the number of sources who gathered the items affect $FLOR_i$'s power. The index shows more success at detecting items when they were gathered by higher-proficiency test takers than by average-proficiency test takers. That is, the index had more power to detect items that were administered to the higher scoring test takers (some of the more difficult items in the item pool). In addition, when more items were compromised, they were easier to detect

Based on the results from these initial simulations, $FLOR_i$ shows promise for use as a test security index in the CAT environment. Future work is needed to investigate refinements of the index that may increase its power. In particular, the current index needs a high proportion of beneficiaries to have an acceptable level of power. Refinements are planned to reduce the proportion of beneficiaries needed. In addition, another limitation of this study is that beneficiaries were assumed to perfectly memorize compromised items. Future work will investigate the power of $FLOR_i$ when other strategies are being used. It is hoped that this work will enable testing program management to more effectively decide how long to leave items in the field.

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing.

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

McLeod, L.D., & Lewis, C. (1998, April). *A Bayesian approach to detection of item preknowledge in a CAT.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA.

Schnipke, D.L., & Scrams, D.J. (1997). *Item theft in a continuous-testing environment: What is the extent of the danger?* (LSAC Computerized Testing Rep. No. 98-01). Newtown, PA: Law School Admission Council.

Stocking, M.L., & Lewis, C. (1995). *Controlling item exposure conditional on ability in computerized adaptive testing.* (ETS Research Rep. No. RR-95-24). Princeton, NJ: Educational Testing Service.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029969

REPRODUCTION RELEASE

(Specific Document)

NCME

I. DOCUMENT IDENTIFICATION:

Title: Detecting Items That Have Been Memorized In The Computerized Adaptive Testing Environment

Author(s): Lori D. McLeod and Deborah L. Schnipke

Corporate Source: (both) Law School Admission Council

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: Lori McLeod	Printed Name/Position/Title: Lori McLeod Research Scientist	
Organization/Address: Law School Admission Council 661 Penn Street Box 40 Newtown, PA 18940-0040	Telephone: 215-968-1161	FAX: 215-968-1169
	E-Mail Address: lmcleod@lsac.org	Date: 6/9/99

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com