GRE

RESEARCH REPORT

Effects of Preexamination Disclosure of Essay Prompts for the GRE Analytical Writing Assessment

Donald E. Powers

February 2005

GRE Board Research Report No. 01-07R
ETS RR-05-01

Effects of Preexamination Disclosure of Essay Prompts for the GRE Analytical Writing Assessment

Donald E. Powers

ETS, Princeton, NJ

GRE Board Research Report No. 01-07R

ETS RR-05-01

February 2005

The report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

Educational Testing Service, Princeton, NJ 08541

****

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board reports do no necessarily represent official Graduate Record Examinations Board position or policy.

The Graduate Record Examinations and ETS are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

****

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

Educational Testing Service
Princeton, NJ 08541

Abstract

This study examined how the practice of prepublishing prompts used on the writing section of the Graduate Record Examinations® (GRE®) General Test impacts test-preparation behavior, test performance, test validity, and examinee perceptions of the value of prompt prepublication. Researchers imposed modest experimental control over how participants used the prompts to prepare for an upcoming test. The strategy test-takers reported using most frequently was simply to "think generally about the potential topics." Slightly fewer than half of study participants wrote sample essays to prepare for the test, and very few (4%) admitted to memorizing essays that could be recalled during testing. Results provided no indication that participants benefited from encountering a prompt for which they had prepared. The vast majority of study participants, however, thought that making the GRE essay prompts available ahead of time is a good testing policy.

Key words: GRE writing assessment, prompt prepublication, test preparation, test validity

Table of Contents

List of Tables

Analytical Writing Assessment

Test score validity hinges not only on the questions that comprise a test but also on a host of procedures, circumstances, and conditions that accompany it—for example, the amount of time allowed, the quality of the testing environment, and the extent to which test-takers are motivated to give their best effort. To a considerable degree, validity also depends on what happens *before* a test is administered, especially with regard to how examinees prepare for the examination. On one hand, inappropriate pretest preparation aimed at "beating the test" or subverting the testing process may enable some test-takers to benefit from certain limitations in the testing system and, as a result, distort the intended meaning of test scores. On the other hand, some kinds of test preparation may enhance validity by *reducing* unwanted influences (such as the lack of familiarity with testing procedures), thereby decreasing the chances that some test-takers will receive test scores that are either too high or too low (see, for example, Powers, 1985).

In the interest of minimizing any effects due to insufficient familiarity with a test, many test makers now typically provide a variety of materials designed to help test-takers become intimately familiar with the tests they take. These materials often suggest strategies for approaching the various question formats, and they usually include practice exams consisting of retired test questions. The belief is that, by using these materials, test-takers will gain a thorough grounding in test-taking procedures, thereby freeing them to focus more on the *substance* of a test than on the *mechanics* of test-taking.

The Analytical Writing Section of the GRE General Test

The analytical writing section of the Graduate Record Examinations® (GRE®) General Test consists of two writing tasks, one requiring examinees to present their perspectives on an issue and the other requiring them to analyze an argument. These two tasks are designed to assess the ability to (a) discuss and critique an argument, (b) articulate and support complex ideas, and (c) sustain a focused and coherent discussion.

One testing practice that has been instituted relatively recently to help test-takers prepare for tests of writing skill (including the GRE analytical writing assessment) is to prepublish the entire pool of essay prompts from which prompts are selected for each test administration. Depending on how this practice is implemented, it has, we believe, the potential for either

enhancing or diluting the validity of test-score inferences. The impact of this relatively new practice has, however, received little study.

Rationale for Prepublishing Essay Prompts

One motive for prepublishing essay prompts is fairness—to ensure that all essay prompts are equally available to every examinee, not just the few who may obtain covert access from unethical test proctors or from fellow test-takers who tested earlier. A potential negative side effect of prepublication, however, is that some examinees may attempt to memorize exemplary essays (possibly ones written by someone else) and simply "regurgitate" these essays when testing. To minimize this prospect, some testing programs release relatively large numbers of prompts, in hopes that a sufficiently large pool will discourage undesirable test-taking behavior.

On the positive side, prepublication of a smaller, reasonably manageable pool of prompts has the potential for increasing the validity of writing test scores by providing additional time for planning—a phase of composing that most experts view as integral to the writing process. (As one graduate faculty member once informed us, most graduate student writing does not involve writing on unfamiliar topics "off the top of one's head!") In other words, greater opportunity for preexamination planning may enable test-takers to devote less test-taking time to formulating and organizing their ideas and more time to translating and communicating them (i.e., committing them to paper and improving the manner in which they are expressed). Thus, if prepublication helps examinees become more familiar with potential test topics, a writing test may be seen as more authentic—that is, less a reflection of the ability to write extemporaneously and more an indication of the kind of planful writing that is required in most academic settings. Test-takers also seem to believe that their writing skills are assessed more accurately when they are permitted to write on topics that have been considered beforehand (Powers, Fowles, & Farnum, 1993; Powers & Fowles, 1998). As a research study participant once suggested to us, prepublishing prompts should elicit writing that is "more consistent with [the kind of writing] that you would see in class."

Many cognitive psychologists seem to concur with these views. Various cognitive models of the writing process (e.g., Bereiter & Scardamalia, 1987; Collins & Gentner, 1980; Flower & Hayes, 1981; Hayes & Flower, 1980; Scardamalia & Bereiter, 1986) all emphasize the role of planning in the writing process. It is obvious, of course, that for writing assessments,

significant planning can be undertaken outside the testing session only if potential topics have been disclosed beforehand.

Besides increasing the perception of authenticity, the opportunity to plan/prepare for a writing assessment may have an additional benefit: The prepublication of essay prompts may lessen anxiety by ensuring that there are no unsettling surprises when essay topics are unveiled during the examination. As some test-takers have suggested, having at least seen, if not thought about, potential topics should relieve stress and minimize the prospects of "freezing up." One member of the GRE Technical Advisory Committee has referred to this phenomenon (i.e., drawing a topic for which no ideas come to mind) as "the blank-page problem."

Prior Research

Much has been written about test disclosure, befitting its status as one of the major education stories of the early 1980s (National Education Association, 1982). Research on its effects, however, has been relatively sparse. In addition, with the exception of three studies (Hale, Angelis, & Thibodeau, 1983; Powers et al., 1993; Powers & Fowles, 1998), apparently all of the published research has addressed the impact of releasing test items *after* a test is administered.[1] Even the exceptions that involved prepublication are less than definitive, however, because none of these studies was conducted in a high-stakes testing environment where test scores actually counted.

In one such study, Hale et al. (1983) investigated the effects of disclosing multiple-choice test items for the Test of English as a Foreign Language™ (TOEFL®) examination. The researchers found that, in general, examinees performed better on disclosed items than on undisclosed items, and that performance depended somewhat on the size of the pool of disclosed items. Specifically, disclosure had a greater effect for smaller pools, presumably because examinees could focus their study on fewer questions.

Also relevant is a study that focused on predisclosing essay prompts for a beginning teacher certification test. Before The Praxis Series™ writing assessment became operational, Powers et al. (1993) conducted a small-scale simulation to estimate the likely impact of disclosing essay topics. At four colleges, writing instructors were asked to take a small set of topics and, using any tricks they could muster, coach students to take the assessment. The subsequent difference between students' performance on disclosed and previously unseen topics

was small (an effect size of about .15), and there was no detectable effect of disclosure on test validity, as evidenced by correlation of essay scores with several other indicators of writing proficiency.

In a later study, Powers & Fowles (1998) recruited GRE General Test examinees to participate in a research administration of the (then) experimental GRE analytical writing assessment. Approximately two to three weeks before testing, study participants received test preparation suggestions and two essay topics, one of which they were later asked to write about during the subsequent research study testing; study volunteers were also told that it was *very likely* that they would be asked to write on one of the two topics they had received.

Analyses revealed a negligible effect as a result of having seen essay topics before the test was administered—an effect that was virtually the same as that noted previously (Powers et al., 1993). As the researchers pointed out, however, the consequences of test disclosure cannot be determined definitively outside the context of a fully operational testing program. Nonetheless, they speculated that, if the patterns of preparation exhibited by research study participants were indicative of what would happen under operational conditions, then test-takers (especially less proficient writers) were likely to utilize predisclosed topics as they prepared for the GRE analytical writing assessment: Even with no apparent motivation, a substantial majority (84%) of study participants reportedly spent time thinking about the prompts they had received, and a minority said they had engaged in more time-consuming activities, such as researching topics (10%) and drafting essays (10%). To reiterate, the main limitation of each of the extant studies is that they may not generalize to a high-stakes testing situation in which test-takers can be expected to be reasonably well motivated.

Objectives

A main focus of this study was the effect of different-sized pools of prepublished prompts—in particular, the tradeoff that is inherent in disclosing a very large pool versus a much smaller one. On one hand, a sizeable pool may minimize the likelihood that test-takers will memorize "canned" responses (i.e., formulaic essays designed to fit multiple prompts), thus decreasing the test's validity. On the other hand, the availability of too many prompts may dilute the (presumably) positive influence of enabling GRE examinees to engage in meaningful planning for writing. Our study was designed to address this tradeoff by identifying the impact

associated with pools of varying size. Equally important, this objective was accomplished within the framework of a larger effort that sought to provide further evidence of the validity of the GRE analytical writing assessment—the first such evidence gathered for the measure in a fully operational setting with motivated GRE General Test examinees. Specifically, the study was designed to:

1. document how test-takers prepare for the GRE analytical writing assessment and, more specifically, how test preparation behavior is influenced by the availability of essay prompts
2. estimate the effects of test preparation on test performance
3. ascertain the impact of preparation on test validity (i.e., the relationship of test scores to other indicators of writing skill)
4. establish the degree to which predisclosure may increase the prevalence of "canned" essays
5. determine examinee perceptions of the practice of prepublishing prompts

Method

Procedure

The study plan entailed overlaying an experimental design on a phenomenon that has occurred only haphazardly. That is, currently, a pool of some 240 essay prompts (about 120 each of the issue and argument types) is published on the GRE website. Prospective test-takers are free to peruse any and all (or none) of the prompts and to use them in a variety of ways to prepare for the writing assessment. Until now, there has been no attempt to document precisely how GRE test-takers use these materials.

This study did not change the current method of prompt publication. However, in order to estimate the effects of the practice, we attempted to impose a structure on the current process by contacting samples of GRE General Test registrants before they took the test and encouraging each sample to focus its preparation on a different number of prompts. Test-takers who registered to take the test during the fall of 2002, the first period in which the analytical writing assessment was administered as part of the GRE General Test, were identified from test GRE

registration files as potential participants. Subsets of prompts from the total pool were sent to these test-takers, who were strongly encouraged to think about the prompts, to develop outlines, and to compose first drafts. (A variation of this "encouragement design" had been used successfully in previous studies of test preparation for the GRE General Test. See, for example, Powers & Swinton, 1984.) To reinforce (and monitor) test preparation behavior, we asked study participants to send us copies of some of their practice essays (those of a certain minimum length). Finally, after the test, test-takers were surveyed about their preparation for the analytical writing assessment.

The study design entailed two factors:

- the number of prompts on which examinees were asked to focus during their preparation (27, 54, or 108)
- whether or not examinees eventually tested on a prompt in the pool on which they were asked to focus

When they eventually took the General Test, some of the participants were, by chance, asked to write on a prompt that was in the pool on which they had been asked to focus their preparation. The likelihood of drawing one of these prompts depended on the number of prompts that examinees were encouraged to use in their preparation (the "pool of focus"; see Table 1).

Table 2 shows the numbers of test registrants who were asked to participate in each condition. We sought to ensure that approximately 100 examinees in each condition would be tested on a prompt that was in their pool of focus; in order to produce these sample sizes, before the test we contacted the numbers of examinees shown in Table 2. Further, we assumed that, for issue prompts, only 67% of each group would choose to write on the prompt on which they focused, and that 33% would opt to write on the other prompt from the two prompts presented. Thus, in order to ensure that 100 test-takers would actually write on a issue prompt of focus, we needed to identify 150 who *encountered* a prompt of focus. Because no choice is given for argument prompts, only 100 examinees in each argument condition needed to be identified. Within each cell of Table 2, various subsets of examinees each received a different set of prompts so that all of the prompts in the pool were seen. In addition to the prompts, examinees received a set of suggestions for using the prompts.[2]

Table 1

Probability of an Examinee Getting a Prompt From a Pool of Focus

	Number of prompts in pool of focus		
Type of prompt	27	54	108
Issue	.306	.557	.890
Argument	.167	.333	.666

Note. Probabilities differ for issue and argument prompts because a choice of two prompts is presented for issue prompts, while no choice is given for argument prompts.

Table 2

Number of Test-Takers Contacted

	Number of prompts in pool of focus		
Type of prompt	27	54	108
Issue	516	270	174
Argument	600	336	150

In passing, we note that greater design efficiency would have been possible (i.e., fewer potential participants needed) by availing ourselves of "insider knowledge"—that is, information about which prompts from the total pool were in use when the study was being conducted. We preferred, however, to ignore this information and instead act as if the entire pool of prompts was being used to constitute examinees' test forms. This strategy eliminated the possibility that we might knowingly provide an advantage to some examinees.

In addition to ensuring that our methods did not inadvertently advantage some test-takers, it was critical to convince study participants (and they in turn their counterparts who were not selected for our study) that by participating in our research they were receiving no special advantage (and their counterparts no disadvantage). In particular, we needed to inform them that we the investigators, being "lowly researchers," had no more information than they did about what prompts would be administered to whom. Thus, our procedures would neither increase nor decrease the likelihood of examinees being asked to write on any particular topic. Therefore, they would fare neither better nor worse by focusing on the prompts we suggested than on some other subset.[3] We also stressed that they were of course free to use any of the prompts on which

we had not asked them to focus.

Data Collection

After the designated numbers of test registrants were contacted, test files were searched to identify those examinees who actually wrote on a prompt that they had received in a pool of focus. The GRE analytical writing scores of these test-takers were retrieved, as were the scores of test-takers who were sent prompts but who did *not* eventually test on a prompt from their pool of focus. All of these test-takers were recontacted by mail immediately after they tested and asked about how they prepared for the writing assessment. Specifically, for both the issue and the argument prompts, they were asked to indicate whether they had spent time on any of a variety of test preparation activities (e.g., reading sample essays) and, if so, approximately how much.

In order to assess the impact of preparation on test validity, study participants were also asked to provide a variety of nontest information, like that collected in previous studies of the validity of the GRE writing section's precursor, the Analytical Writing Assessment (Powers, Fowles, & Boyles, 1996; Powers, Fowles, & Welsh, 1999; Powers, Fowles, & Welsh, 2001). This information included:

- grade average in courses that required "considerable" writing
- grade on the most recent writing assignment
- grade average in courses that required "mostly reasoning and thinking"
- grade average in courses in formal logic, reasoning, or critical thinking
- grade on the most recent test or assignment that depended heavily on reasoning

Grades were recorded on a 9-point scale with "less than C" = 1, C = 2, C+ = 3, …, A+ = 9.

In addition, we asked participants to report:

- how successful they had been with various kinds of writing (personal, creative, persuasive, analytical-critical, descriptive, and applied)
- their ability with respect to the kinds of thinking skills that have been deemed by graduate faculty to be important for success in graduate education (Powers & Enright, 1987)

- the extent to which problems with writing hindered their ability to demonstrate what they had learned in college
- the degree to which they thought they had been effective in communicating their thoughts and ideas in writing while in college
- their overall impression of the GRE prompt publication policy

Finally, participants were asked to submit two samples of their course-related writing and to describe certain characteristics of each sample (e.g., the nature of the assignment that elicited it, how much time was devoted to composing the sample, whether or not it was graded, the grade it had received, and what role, if any, it played in determining a course grade). Approximately three weeks after the initial contact, nonrespondents were sent an additional copy of the questionnaire. The incentive to complete all aspects of the study was a $25 gift certificate.

Data Preparation

The course-related writing samples were evaluated by applying scoring procedures developed by four university professors, all experts in writing instruction/assessment, for a previous GRE-sponsored study (Powers et al., 1999; 2001). The scoring guide was a composite of the GRE issue and argument rubrics, expanded slightly in order to focus on the complexity of thought that was characterized by one of the previous consultants as being indicative of "scholarly habits of mind." The guide employed the same 6-point scale and labels (6 = outstanding, 5 = strong, 4 = adequate, 3 = limited, 2 = seriously flawed, and 1 = fundamentally deficient) as the issue and argument guides, and defined specific features at each score level by combining elements from both the issue and the argument guides. For example, a paper was judged "outstanding" if it displayed a cogent, well-articulated treatment of the subject/topic and demonstrated mastery of the elements of writing. At the other extreme, a paper received the lowest score (fundamentally deficient) if it displayed serious deficiencies in its treatment of the subject/topic and lacked control of the basic elements of writing (e.g., if it provided little evidence of the ability to develop and organize a coherent treatment of the subject/topic, contained severe and persistent errors in the use of language and sentence structure, or contained a pervasive pattern of errors in grammar, usage, and mechanics that resulted in incoherence).

All course-related writing samples were read by college and university faculty—all

teachers and/or experienced evaluators of writing—who were trained to apply the scoring guide. For practical reasons, the samples were read only once. GRE analytical writing assessment essays were evaluated as part of the regular operational test-scoring process. Responses to the study questionnaire were processed and analyzed as appropriate. In some cases, scales were developed from subsets of questions; in other cases, responses to individual questions served as the variable of interest.

Results

Sample

Of the test-takers whom we contacted, a total of 199 responded to our request for information about their test preparation for the GRE writing assessment. These test-takers were slightly more able than GRE test-takers in general, having somewhat higher GRE verbal and quantitative scores (*M*s = 507 and 599 on the 200-800 score scale, respectively, with *SD*s = 100 and 124) than did a reference group of 1,000 test-takers who took the exam during the same time interval (*M*s = 490 and 555, *SD*s = 106 and 134, respectively). Respondents also had slightly higher GRE analytical writing scores (*M* = 4.50 on the 1-6 score scale, *SD* = .91) than did the reference group (*M* = 4.35, *SD* = .96). Of these 199 respondents, 79 had received issue prompts and 120 had received argument prompts (see Table 3). As can be seen, because of an inexplicably low response rate[4], we were unable to meet our initial targets. Answers to each of our research questions follow.

Table 3

Number of Study Participants Responding

	Number of prompts in pool of focus		
Type of prompt	27	54	108
Issue	45	23	11
Argument	61	34	25

Research Question 1: How Do Test-Takers Prepare for the GRE Analytical Writing Assessment? Is Test Preparation Behavior Influenced by the Availability of Essay Prompts?

Table 4 shows the percentage of respondents who prepared in each of several ways for the analytical writing test, as well as the amount of time devoted to each method. The most frequently used strategy (by 82% of study participants) was to "think generally about the

potential topics." The modal time spent using this strategy was less than one hour. Slightly fewer than half of study participants wrote sample essays to prepare for the test, and very few (4%) admitted to memorizing essays that could be recalled during testing.

Table 4

Percentages of Study Participants Who Used Various Test Preparation Strategies

Method of preparation	Yes	Hours spent		
		Less than 1	1 – 4	5 or more
Thought generally about the potential topics	82	39	33	11
Read sample essays	79	36	37	6
Thought about specific points or examples to discuss	68	33	27	8
Brainstormed about ideas	62	34	20	7
Wrote sample essays	48	15	19	13
Wrote outlines for topics	40	20	14	7
Other	21	7	5	9
Did reading or research about topics	15	10	5	1
Memorized essays	4	3	1	0

Note. $N = 199$ respondents.

On average, study participants used about six-to-seven prompts of either kind in their preparation (see Table 5). These prompts may have been from either the sample that we provided or the larger pool of prompts that was available to all GRE test-takers. The number of prompts used did not vary significantly according to how many prompts we had provided (27, 54, or 108). Overall (for both issue and argument prompts), a slight majority of participants used one-to-five prompts in their preparation (Table 6).

Table 5

Mean (SD) Number of Prompts Used in Preparing for the GRE Writing Assessment by Treatment Condition

Type of prompt	Treatment groups		
	27 prompts	54 prompts	108 prompts
Issue	7.1 (9.9)	6.9 (11.7)	7.5 (10.1)
Argument	5.9 (6.2)	6.7 (7.2)	6.9 (14.8)

Table 6

Percentages of Study Participants Using Various Numbers of Prompts to Prepare

Type of prompt	Number of prompts				
	None	1-5	6-20	21-50	More than 50
Issue	22	52	20	6	1
Argument	20	53	25	1	1

Note. $N = 199$ respondents.

Research Question 2: How Does Test Preparation Affect Test Performance?

Our study was predicated on the assumption that sending varying numbers of essay prompts to study participants would result in varying levels of test-preparation effort. That is, on the basis of previous test-preparation studies, we had reason to believe that test-takers who received 104 topics would devote more time to preparing than would those who received only 27 (or 54) prompts. This turned out not to be the case, however: Participants' responses to various questions revealed that there was no significant relationship between study condition (i.e., 27, 54, or 108 prompts) and the amount or kind of test preparation in which participants engaged.

Even though our experimental manipulation proved ineffective, we carried out an analysis to compare (a) the test performance of test-takers who said they had prepared in some way for the prompt on which they were eventually tested with (b) the performance of test-takers who said they had not prepared at all for the prompt on which they were asked to write. An analysis of covariance was conducted for participants who had received issue prompts and again for those who had received argument prompts. The independent variables were (a) treatment condition (27, 54, or 108 prompts) and (b) whether the test-taker had prepared for the prompt on which he/she was eventually tested. When issue score was entered as the dependent variable, the covariates were GRE verbal ability score and score on the argument prompt. When argument scores were used as the dependent variable, issue scores were used as a covariate, again along with GRE verbal ability scores.

Table 7 shows the resulting means for these analyses. For issue prompts, the analyses revealed no significant main effect with respect to (a) treatment condition [$F(2, 71) = 0.64$], (b) preparation on prompt tested [$F(1, 71) = 0.06$], or (c) interaction of treatment condition and preparation on prompt tested [$F(2, 71) = 0.80$]. Similarly, no significant effects for argument

prompts with respect to either of the main effects [$F(2, 112) = 0.88$ and $F(1, 112) = 0.72$] or the interaction between the two [$F(2, 112) = 0.38$] were found. Thus, this analysis provided no indication that participants benefited from encountering a prompt for which they had prepared.

Table 7

Means and Standard Deviations for Issue and Argument Scores According to Whether or Not Test-Takers Prepared for the Prompt on Which They Were Tested

	Number in pool of focus			
Prepared on prompt tested?	27	54	108	Overall
Issue prompt				
Yes				
M	4.71	3.83	4.75	4.57
SD	.78	.29	1.06	.88
N	12	3	2	17
No				
M	4.64	4.55	4.39	4.51
SD	.77	.76	.82	.80
N	33	20	9	62
Overall				
M	4.66	4.46	4.45	4.53
SD	.77	.75	.82	.86
N	45	23	11	79
Argument prompt				
Yes				
M	3.77	4.38	4.50	4.14
SD	1.17	1.09	1.18	1.20
N	11	8	6	25
No				
M	4.63	4.33	4.55	4.53
SD	.86	1.14	.88	.98
N	50	26	19	95
Overall				
M	4.48	4.34	4.54	4.45
SD	.97	1.11	.93	1.01
N	61	34	25	120

Research Question 3: What Is the Impact of Preparation on Test Validity (i.e., What Is the Relationship of Test Scores to Other Indicators of Writing Skill)?

Because we found no detectable effect on test performance, we conducted no analysis of the effects of our experimental treatment on test validity. However, it is of some interest to note the correlations, across all treatment conditions, of performance on the issue and argument prompts with each of several nontest indicators of reasoning and writing ability. Thus, Table 8 provides this information.

Table 8

Correlation of Performance on GRE Issue and Argument Tasks With Other Indicators of Writing and Reasoning Skills

Indicator	Issue	Argument	Total score (issue and argument)
Self estimate of reasoning skills	.22	.21	.25
Self estimate of writing skills	.25	.18	.24
Self reported grade point average			
In "reasoning" courses	.01	.19	.12
In writing courses	.24	.25	.28
Self comparison with peers			
Reasoning	.05	.21	.16
Writing	.24	.22	.26
Self report of problems with writing	-.28	-.27	-.32
Self report of effectiveness of written communication	.27	.20	.27
Self report of success with writing	.21	.23	.26
Evaluation of writing samples			
Sample A	.29	.25	.31
Sample B	.23	.19	.24
Both A and B	.32	.30	.36

Note. Correlations of approximately .14 are significant at the .05 level, two-tailed.

n = 182 to 199.

As is clear from the table, the correlations are all modest, mainly in the .20s. For example, the correlation of each prompt type with an evaluation of two student-provided, course-related writing samples was .30-.32. (These writing samples had the following characteristics: About 86% were written outside of class, about 79% were written within the year preceding our study, about 84% were written with 9 hours or less effort, about 93% were written with little or no help from others, and about 67% had received grades of A- or better.) The correlation of a self-report index of success with various kinds of writing in college (persuasion, analysis/criticism, description, examination writing, and applied writing) was .21-.23 for the issue and the argument prompts. Responses (on a 5-point scale ranging from "hardly ever" to "almost always") to a single question, "During college, how often did problems with writing hinder your ability to show what you had learned (e.g., on tests and assignments)?" correlated -.28 and -.27 with performance on the issue and argument prompts, respectively. That is, the poorer the GRE essays, the more problems students reported in demonstrating their learning.

Research Question 4: Does Prompt Predisclosure Increase the Prevalence of "Canned" Essays?

A total of 5% of study participants who received issue prompts said they had tried to memorize essays so that they could reproduce them upon testing. None of these test-takers had attempted to memorize more than five essays. For those who received argument prompts, a total of 3% said they had memorized essays—again, none more than five essays. There was no relationship between the number of prompts received and the degree to which test-takers attempted to memorize essays.

Research Question 5: What Are Examinees' Perceptions of the Practice of Prepublishing Writing Prompts?

Study participants were asked if they thought that making the GRE essay topics available ahead of time is a good testing policy. The vast majority said either "definitely" (44%) or "probably" yes (36%), while a minority said "probably" (13%) or "definitely" not (7%). The most frequent comment from those who endorsed the practice suggested that prepublishing the topics helped to reduce pressure/anxiety by "eliminating one of the unknowns" and giving test-takers an idea of what to expect. Most often, the minority who did not favor the practice indicated that there were just too many prompts to be of use in preparing—that the task was

"overwhelming." Another relatively frequent comment from the dissenting minority was that prepublishing the prompts would diminish the test's ability to measure reasoning and organizational skills in an extemporaneous fashion.

Discussion and Implications

Because we were unable to (a) fully implement the treatment conditions as planned and (b) enlist sufficient numbers of test-takers to participate, none of the initial study objectives was fully achieved. Nonetheless, though limited, the study findings have some notable implications.

For the GRE Program

The most basic, and perhaps most important, outcome of the study is additional information about the meaning (validity) of GRE analytical writing assessment scores, as evidenced by their correlations with several nontest indicators of both reasoning and writing skills. As stated above, the correlations are best described as modest. It should be noted, however, that the correlations among the various nontest indicators are modest also, suggesting either that they reflect different facets of writing ability or that they are of modest reliability. This outcome extends previous research on the GRE writing assessment in one important way: The results are based not on experimental research administrations, but rather on fully operational administrations of the test. This information should, therefore, add to the accumulation of evidence needed to meet professional standards for educational and psychological testing.

In addition, we hoped that the study would reveal the impact on test-taking behavior of a particular GRE program practice—namely, prepublishing essay prompts. More specifically, we hoped to learn how the size of the pool might affect examinees' test preparation strategies. Unfortunately, our study sample was small and not representative of all GRE test-takers. Moreover, the study treatment was only partially implemented. However, to the extent that the results provide any indication whatsoever of other GRE test-takers' approach to testing, we can probably assume that the typical GRE test-taker will employ only a small fraction of the pool of prompts in his or her preparation for the test—on average fewer than 10% of each kind of prompt.

In addition, test-takers are very likely only to think about the topics and about possible ideas or examples about which to write: Fewer than a third of the study sample devoted more

than an hour to writing essays, and only about 1% admitted to spending more than an hour trying to commit essays to memory. Although these results do not suggest exactly how large the pool should be, they do suggest at least that, from the standpoint of minimizing inappropriate test-taking behavior, the current pool is probably sufficiently large. Finally, the results also have implications for advising test-takers about test-preparation practices—at least with regard to informing them how their fellow test-takers tend to prepare for the test. This information may provide some comfort to those test-takers who may be anxious about being less well prepared to take the GRE exam than are their fellow graduate school applicants.

For the Assessment of Writing Skill

A prevailing view among composition specialists is that writing is a *process*, one that entails complementary activities of prewriting/planning, drafting, writing, and revising. Because most tests of writing ability (like the GRE analytical writing assessment) usually allow enough time only for developing a first draft, and not for any significant planning or revision, they may not adequately elicit all of the processes that writers typically employ, and therefore may not fully represent all of the important facets of writing proficiency. In other words, the tests may suffer from a major source of invalidity—what Messick (1989) has termed "construct underrepresentation." Although the study did not allow us to assess the degree to which prepublishing prompts may have affected the validity of the GRE writing assessment, study participants were reasonably clear in their belief that prepublishing the prompts had, for several reasons, made the test a more valid indicator of their writing skills.

For Test Fairness

Some critics of standardized writing assessments apparently feel that impromptu writing measures, such as the GRE writing assessment, pose a serious threat to test fairness. Because such tests necessarily restrict access to information resources and allow little time for reflection and revision, they may penalize certain students—for example, diligent students who might perform much better when given sufficient time and adequate resources. Cultural differences may also be associated with the penchant for writing quickly and extemporaneously. We had hoped initially that the study might reveal the extent to which between-group test-score differences are reduced by allowing more time for planning, thus enhancing test fairness. However, our study sample proved too small to allow any meaningful analyses by subgroups.

For Admissions Testing

The study results provide some modest new information about the promise (and potential pitfalls) of a relatively innovative admissions testing practice. In the 1980s, complaints about the secrecy of testing agencies resulted in legislation (in New York State) that mandated the disclosure of *previously used, retired* test questions. This practice received a great deal of fanfare and, as mentioned earlier, was deemed one of the top stories about standardized testing in the 1980s. In contrast, the *pre*publication of test questions for writing assessments, a practice that seems to us to be far more noteworthy (and potentially more useful), has received far less attention and even less research. We hope that the modest information generated by the study described here will, at the least, generate interest among researchers in further studying the effects of this practice. Though it would be difficult, future researchers might attempt to focus more specifically on subgroups of test-takers thought to have the greatest motivation to prepare and to memorize prompt responses.

References

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Collins, A., & Gentner, D. (1980). A framework for a cognitive theory of writing. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 51-72). Hillsdale, NJ: Lawrence Erlbaum.

Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31-50). Hillsdale, NJ: Lawrence Erlbaum.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*, 365-387.

Fremer, J. (1981, April). *Impact of test disclosure legislation on test development*. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles.

Gilmer, J. S. (1989). The effects of test disclosure on equated scores and pass rates. *Applied Psychological Measurement*, *13*, 245-55.

Gould, J. D. (1980). Experiments on composing letters: Some facts, some myths, and some observations. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 97-128). Hillsdale, NJ: Lawrence Erlbaum.

Hale, G. A., Angelis, P. J., & Thibodeau, L. A. (1983). Effects of test disclosure on performance on the Test of English as a Foreign Language. *Language Learning*, *33,* 449-464.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum.

Lockheed, M., Holland, P., & Nemceff, W. (1982). *Student characteristics and the use of the SAT test disclosure materials* (College Board Report No. 82-3, ETS RR-82-31). New York: College Entrance Examination Board.

Marco, G. L. (1981). Equating tests in an era of test disclosure. *New Directions for Testing and Measurement*, *11*, 105-22.

Messick, S. J. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.; pp. 13-103). New York: Macmillan.

National Education Association. (1982). 10 top education stories of 1981-82. *Today's Education*, *71*, 20, 22-23.

Powers, D. E. (1985). Effects of test preparation on the validity of a graduate admissions test. *Applied Psychological Measurement*, *9*, 179-190.

Powers, D. E., & Enright, M. E. (1987). Analytical reasoning skills involved in graduate study: Perceptions of faculty in six fields. *Journal of Higher Education, 58*, 658-682.

Powers, D. E., & Fowles, M. E. (1998). Effects of preexamination disclosure of essay topics. *Applied Measurement in Education*, *11*, 139-157.

Powers, D. E., Fowles, M. E., & Boyles, K. (1996). *Validating a GRE writing test* (GRE Board Technical Report No. 93-26a, ETS RR-96-27). Princeton, NJ: ETS.

Powers, D. E., Fowles, M. E., & Farnum, M. (1993). Prepublishing the topics for a test of writing skills: A small-scale simulation. *Applied Measurement in Education*, *6*, 119-135.

Powers, D. E., Fowles, M. E., & Welsh, C. K. (1999). *Further validation of a writing assessment for graduate admissions* (GRE Board Professional Rep. No. 93-26b, ETS RR-96-27). Princeton, NJ: ETS.

Powers, D. E., Fowles, M. E., & Welsh, C. K. (2001). Relating performance on a standardized writing assessment to performance on selected academic writing activities. *Educational Assessment*, *7*, 277-303.

Powers, D. E., & Swinton, S. S. (1984). Effects of self-study for coachable test item types. *Journal of Educational Psychology*, *76*, 266-278.

Scardamalia, M., & Bereiter, C. (1986). Research on written composition. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.; pp. 778-803). New York: Macmillan.

Stricker, L. J. (1984). Test disclosure and retest performance on the SAT. *Applied Psychological Measurement*, *8*, 81-87.

Notes

[1] For instance, Lockheed, Holland, and Nemceff (1982) documented the characteristics of test-takers who requested disclosed materials; Gilmer (1989) simulated the effects of test-item disclosure on test equating; and Stricker (1984) investigated the effects of test disclosure on retest performance for the Scholastic Aptitude Test. Other researchers have considered how test disclosure (again, the post-administration release of test items) might affect both test development (Fremer, 1981) and test equating (Marco, 1981).

[2] For the issue prompt type, the following suggestions were relayed:

- Read the question. In your own words, describe the thinking and writing you will have to do for this assignment.
- Don't jump to a position on the issue. Rather, list some reasons that support one point of view and then some other reasons that support a different point of view. Which reasons are stronger? Why? What other perspectives need to be considered?
- Decide how your own position lines up with these different points of view. State your position as clearly as you can.
- As you develop your position, you might want to show your reader that you've considered various perspectives before drawing your own conclusions.
- Also, consider using concrete examples to illustrate what you mean. Your job is to impress the reader that you can think clearly and write effectively; well-chosen examples can be very persuasive.

A similar set of suggestions was developed for the argument prompt type.

[3] One of the reviewers of the report speculated that this instruction may have been misinterpreted by test-takers and thus inadvertently convinced them that using the prompts to prepare for the test was simply not a good strategy.

[4] One reviewer suggested that our request for information may have been excessive, thus discouraging participants from responding to our invitation.

GRE-ETS
PO Box 6000
Princeton, NJ 08541-6000
USA

To obtain more information about GRE programs and services, use one of the following:
Phone: 1-866-473-4373
(U.S., U.S. Territories*, and Canada)
1-609-771-7670
(all other locations)
Web site: www.gre.org

* America Samoa, Guam, Puerto Rico, and US Virgin Islands