LSAC RESEARCH REPORT SERIES

Developing an Assessment of First-Year Law Students' Critical Case Reading and Reasoning Ability: Phase 2

Dorothy H. Evensen
Penn State University

James F. Stratman
University of Colorado at Denver

Laurel C. Oates
Seattle University School of Law

Sarah Zappe
Penn State University

Law School Admission Council
Grants Report 08-02
March 2008

A publication of the Law School Admission Council

The Law School Admission Council (LSAC) is a nonprofit corporation whose members are more than 200 law schools in the United States, Canada, and Australia. Headquartered in Newtown, PA, USA, the Council was founded in 1947 to facilitate the law school admission process. The Council has grown to provide numerous products and services to law schools and to more than 85,000 law school applicants each year.

All law schools approved by the American Bar Association (ABA) are LSAC members. Canadian law schools recognized by a provincial or territorial law society or government agency are also members. Accredited law schools outside of the United States and Canada are eligible for membership at the discretion of the LSAC Board of Trustees.

This study is published and distributed by LSAC. The opinions and conclusions contained in this report are those of the author(s) and do not necessarily reflect the position or policy of LSAC.

Table of Contents

Executive Summary

The research team in this project developed a prototype, multiple-choice test to assess case reading and reasoning among law students at two points during the first year. The project was motivated by the importance of cases in legal education, a paucity of empirical evidence concerning law students' reading abilities, and a need for measurements that could be used to study pedagogical interventions and students' skill development. A primary goal of the research project was to test the theoretical argument that law students have difficulty dealing with what are called the *indeterminacies,* or discourse-specific ambiguities and vagueness of cases.

Phase 1 of this study employed a cross-sectional, matched-subjects design that produced results indicating that among the 161 first-year law students who took the prototype test version (TV1) in either the fall or spring semesters, the overall combined mean was 7.91; the most common score was 9/14 (64%) correct, and the scores ranged from 14/14 (100%) to 3/14 (21%). Individual test items showed a wide range of difficulty, with a good balance between easy, difficult, and moderately difficult items. No significant difference in test performance was detected between semesters.

Phase 2 of the study was undertaken to test the validity of the test and to see if findings could be replicated using a longitudinal design involving not only first-year, but also third-year law students. To accomplish these research goals a second, alternate version of the test (TV2) was constructed. Data were collected in the spring and fall of 2006 among the five law schools involved in Phase 1. The 146 first-year subjects who participated in a silent reading version of the study were randomly assigned to either TV1 or TV2. Sixty-three third-year students, who had participated in Phase 1, volunteered to sit for TV2. In addition, 30 students from a sixth law school responded to a truncated version of both tests under think-aloud conditions.

Reading materials for both of these multiple-choice tests consisted of three cases related to an appellate argument. The 14 questions—resulting from multiple, iterative reviews among legal experts—reflected two comprehension difficulty categories (individual cases versus cross-case questions) and two semantic difficulty categories (determinate-meaning versus indeterminate-meaning questions). All test items contained five possible answer choices, and justifications were written for each possible answer.

Statistical analyses produced results similar to the findings from Phase 1. No significant differences between first- and third-year scores were detected. The total mean scores for each test version were 8.3 for TV1 and 7.25 for TV2. Mean scores for students between semesters (two and three) or between years (first and third) were not significantly different, indicating that students' case reading and reasoning skills do not improve as a result of law school instruction. Also, consistent with Phase 1 findings, the test showed a positive but low correlation with LSAT scores and law school grade point averages.

In terms of the validity question, the means and distributions of scores between phases and administrations were highly similar. In addition, the pattern of students scoring best on single-case, determinate items and worst on cross-case, indeterminate items held, as did the decreasing compounding effect of item types. To further test the validity claim, multiple methods of establishing construct validity were undertaken: statistical analysis using a variation of differential item functioning (DIF), expert review, and process means using think-aloud data. These analyses yielded outcomes related to the equivalence of the two test versions. It was found that these methods yielded only moderate correspondence and that the process method produced the most meaningful results.

Think-aloud data were qualitatively analyzed to identify constructs, skills, and subskills believed to undergird case reading and reasoning. By comparing think-aloud findings with written item justifications, test writers were able to essentially "test their hypotheses" about case reading. These data also allowed us to revise the test on empirical, grounded bases and made manifest the students' errors.

Overall we believe that this research achieved its proposed goals. Future research in this area could focus on the uses of these revised test versions or might use them as prototypes to construct similar types of items. We see such efforts as advancing the formidable challenges of describing developmental trajectories in legal literacy.

Introduction

This project began with our first award from the Law School Admission Council (LSAC) in 2003 and serves two broad purposes, one practical and the other theoretical. The practical purpose is to develop a prototype multiple-choice instrument assessing law students' critical case reading and reasoning skills. If successfully developed, such a test would be unique and could address two important needs of legal educators. First, it could serve as a useful tool for law classroom teaching by providing formative (as opposed to summative) reading and reasoning skill assessment, so that students do not need to wait until the traditional end-of-term essay examination to receive feedback, albeit indirect, about these crucial abilities. Second, such an instrument could be used in future research, for instance, in controlled laboratory instructional intervention experiments, or in classroom-situated longitudinal studies that attempt to measure the impact of different pedagogical approaches to improving these skills.

The theoretical purpose of the project is to advance empirical efforts to describe the information-processing constructs comprising case reading and reasoning skills and to produce evidence concerning the developmental relationships among these constructs. In general, we have entered upon this line of work because it is under-researched

relative to its importance, and because critiques of legal education by legal educators themselves continue to suggest serious problems with the way the majority of law schools approach instruction developing these skills. There have been few systematic attempts to measure or describe changes in literacy skills required of legal discourse, and those that exist provide contradictory indications (American Bar Association, 1992; Feltovich, Spiro, Coulson, & Myers-Kelson, 1995; Kissam, 2000; Minnis, 1994; Munro, 2000; Schwartz, 2001; Senger, 1993). A crucial impediment has been defining the underlying constructs with theoretical clarity and consistency, such that both the hurdles students face and the factors contributing to overcoming them can be reliably described and measured (Blasi, 1995; Gross, 1973; Lustbader, 1997; McKinney, 2005; Palasota, 1991; Schlag, 1989). Quantitatively scored tests have the potential to address this difficulty. Given the success of such tests in identifying underlying constructs involved in comparably challenging nonlegal reading and discourse analysis tasks, their potential contribution to understanding the nature and development of law students' case reading and reasoning skills warrants attention (e.g., Bryden, 1984; Hofer, 1987).

Nearly all of the previous scholarly efforts to describe the structure and development of case reading and reasoning skills proceed from classroom teachers' perceptions and tacit theories about the origins of students' difficulties. While teachers' perceptions are certainly valuable, they are often conflicting. For example, some teachers may see processes A, B, and C as completely accounting for a competent case reading, while others, discussing these processes, may see B and C as necessary, process A as largely irrelevant, and process D as crucial but often missing. Similarly, in terms of longitudinal development, one teacher may perceive a certain component process to be an essential "building block" that needs to be learned early in a case-reading instructional program, while another teacher may perceive this same component process to be more "advanced." Developing and field-testing a multiple-choice instrument assessing case reading and reasoning skills can help refine these tacit theories and address conflicts arising between them.

In particular, our test development effort builds upon previous research findings and theoretical insights across four relevant lines of inquiry:

1. Studies of law classroom and law student study group discourse (Evensen, 2004; Mertz, 1998)

2. Law teachers' analyses of the origins of students' case reading and reasoning difficulties (Fajans & Falk, 1993; Gross, 1973; Minnis, 1994; Schwartz, 2001, 2003)

3. Real-time think-aloud studies of law students' case reading and legal problem-solving processes (Christensen, 2007; Lundeberg, 1987; Oates, 1997; Senger, 1993; Stratman, 2002, 2004; Weinstein, 1998)

4. Non-law educational studies investigating the relationships between students' question-asking activity during reading, their purposes for reading, and their reading comprehension (Deegan, 1995; Graesser, Langston, & Baggett, 1993; Graesser & Person, 1994; Otero & Graesser, 2001; Rosenshine, Meister, & Chapman, 1996).

Collectively, these previous studies have led us to five theoretical assumptions that we judge to be critical in designing our prototype instruments.

First, we posit that a valid test of case reading and reasoning skill must assess students' understanding of the *relationship between* their specific purpose(s) for reading cases and the information in them that is more versus less relevant to that purpose. In real-world law practice, no one reads a case just to read it, but normally does so in response to some larger legal problem or exigency. Consequently, a test of law students' case reading and reasoning skill should not simply be a test of the students' retention of the "content" of a case *apart from* any larger purpose or exigency, but also a test of students' ability to focus upon content appropriate for a given purpose. The conventional term for students' understanding of the purpose for which they read is their "task representation," and thus in developing our prototype we assume that a competent case reading must be one that is *responsive to* this representation. Experimental research investigating the relationships between task representation and text comprehension shows fairly consistently that readers' text comprehension can be improved or degraded depending on the nature of the task representation that such readers are assigned (Mills, Diehl, Birkmire, & Mou, 1995; Otero, 1998; Pressley & Afflerbach, 1995; Stratman, 2002; Wiley & Voss, 1999). For this reason, in our prototype we ask students to assume a purpose for reading sample cases involving a legal exigency. This purpose should equip competent readers with specific goals and constraints that shape their information-processing activities during reading.

Second, because we posit that assessing competency in case reading and reasoning must include an assessment of students' understanding of the *relationship between* their specific purpose(s) for reading cases and the relevance of particular information contained in them, we also posit that competency requires readers to be able to *generate and distinguish more from less purpose-relevant questions to ask.* In other words, a shift in one's underlying purpose for reading a case often means one must shift the questions one might raise about its contents. Indeed, from our point of view, being able to identify *purpose-relevant questions* to pursue during or after a case reading *is as important as* gaining a clear understanding of a case's issue(s), facts, jurisdiction, reasoning, holding, policy, and dicta. The more complete the readers' understanding of the purpose for which they are reading, the more likely it is they will ask purpose-relevant questions as part of their text processing.

Although the relationship between readers' understanding of purpose and the quality of questions they ask during reading has only begun to be empirically investigated in law schools (Deegan,1995; Stratman, 2002, 2004; Weinstein, 1998), extensive nonlegal educational research investigating factors that promote relevant question-asking behavior during reading strongly supports this assumption (Graesser, Langston, & Baggett, 1993; Graesser & Olde, 2003; Graesser & Person, 1994; Hacker, Dunlosky, & Graesser, 1998). A common observation in this line of work is that question-asking during reading is highly beneficial because it helps students distinguish portions of the text they do and do not understand. However, students often fail to ask questions because they fail to monitor their comprehension as they read. Consequently they often overestimate their understanding. For example, in a large meta-analytic review of experimental studies investigating students' ability to forecast how much they will understand from a text, Maki (1998) found only a small correlation (r = .27) between students' scores on objective comprehension tests and students' predictions, offered before or during reading, concerning how well they would understand the material. Similarly, students also retrospectively overestimate both the number and quality of questions they actually ask during reading.

Third, this view of the centrality of purpose in stimulating question-asking behavior also leads us to assume that for any given purpose for reading a case, some parts of the case discourse will manifest *indeterminacies of interpretation while other parts will not.* When we speak of such indeterminacies, we are speaking of more than the obviously important uncertainty involved in assessing whether or not a court correctly applies a rule or common-law precedent to a novel set of case facts. We are also speaking of all the rhetorical ways that statements offered by courts may be open to interpretation, such that there may be no way to tell precisely what a court means or precisely how it is reaching its conclusion. For example, a court's statement of legal issues to be decided may seem inconsistent relative to what the facts in the case might suggest or in relation to what one or the other of the advocates actually pled; the facts as described may seem to be confounded or selectively focused; and a court's explanation of the legal principles, rules, or precedents it is applying may seem logically inconsistent, arbitrarily chosen, truncated, or simply absent. Faced with linguistic and rhetorical indeterminacies in case discourse beginning with their first days in law school, law students need to develop the ability to decide when the meaning of such discourse and reasoning is more questionable and open to interpretation and when it is less so (Fajans & Falk, 1993; Hoeflich, 1991; Moore, 1981; Wetlaufer, 1990).

What we wish to stress is that, in competent case reading, the purpose for which students read should lead them to detect certain indeterminacies of interpretation in the case discourse that have strategic value given their particular legal exigency. That is, just as we posit that a shift in readers' underlying purpose should stimulate a shift in the questions readers ask, so a shift in this purpose should also stimulate a different appraisal concerning what indeterminacies are present in a court's discourse and, among these, which ones are more versus less relevant. For example, when lawyers are researching and reading cases in anticipation of advocacy on behalf of a client, their reading is likely to be sensitive to indeterminacies that are potentially problematic or helpful, either for their contentions or for those of their opponents. These indeterminacies may turn out to be ones that help direct further legal research, that suggest strategically valuable concessions to be made in argument, or that provide viable counterarguments to positions taken by their opponents or by the judiciary in court. Of course, competent case readers remain aware of the fact that they may need to adjust their purpose in view of the indeterminacies they discover. Less competent readers, in contrast, may fail to detect purpose-relevant indeterminacies or, if they do detect them, may fail to properly assess their degree of relevance to the purpose and legal exigency for which they are reading. Empirical research investigating how both professional attorneys and students detect and evaluate indeterminacies in case discourse is limited, but the available studies consistently underscore the importance of assessing this ability (Hofer, 1987; Stratman, 1994, 2002, 2004).

Fourth, there are some additional dynamic relationships among these assumptions, which we can set forth with the help of an illustration.

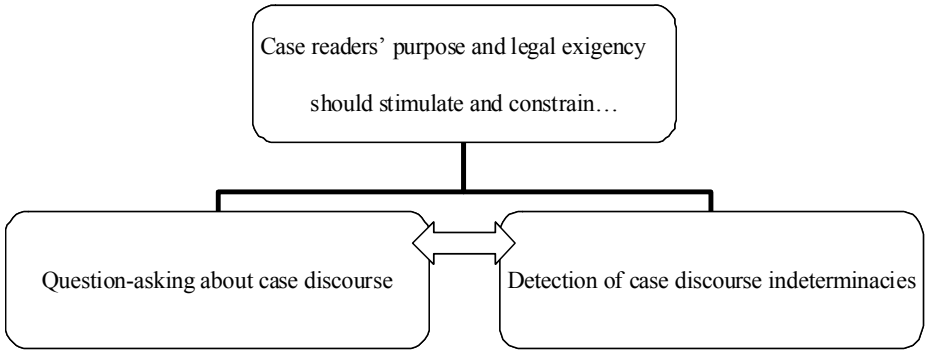


FIGURE 1. *Components of competent case reading*

As Figure 1 shows, and as explained above, in competent case reading and reasoning the readers' purposes and legal exigencies should constrain both the indeterminacies of interpretation readers detect and also the questions they raise about them. We assume that the process by which readers' initial perception of purpose stimulates their question-asking behavior and their detection of relevant indeterminacies in the court's discourse is a primary process (Stratman, 1990). However, as a secondary process, the detection of purpose-relevant indeterminacies should itself lead readers to generate still more purpose-relevant questions. Raising purpose-relevant questions *while reading* should facilitate the detection of still more purpose-relevant indeterminacies in the case discourse. The importance of this secondary process as a component of competent case reading finds support in both legal and nonlegal empirical studies of students' information-processing during text comprehension (Chinn & Brewer, 1993; Deegan, 1995; Otero & Kintsch, 1992; Stratman, 2002, 2004). Conversely, one way that a case reading becomes *less* competent is when readers perceive only purpose-irrelevant indeterminacies of interpretation, or when they largely ask purpose-irrelevant questions, or both. As Figure 1 should help to suggest, then, we view both processes as essential to competent case reading, and we have attempted to elicit and measure both processes in the design of our pilot tests.

Fifth, and finally, we posit that a valid test of case reading and reasoning skill should assess students' ability not only to read individual cases, but to read multiple cases. Specifically, students need to be able (a) to construct accurate representations of multiple, closely related cases, (b) to detect indeterminacies of interpretation arising between or among them, and (c) to distinguish more from less purpose-relevant questions about their relationships to each other. The legal education literature is replete with anecdotal illustrations of the difficulties that students encounter when transitioning from analyzing single cases to analyzing multiple cases (Fajans & Falk, 1993; Lustbader, 1997; Schwartz, 2001, 2003). Like these educators, we do not think that deploying single-case reading skills to multiple cases is simple and straightforward (i.e., as if the extension of single-case skills to multiple related cases can be achieved through mere replication of the same process). Cognitively speaking, we think the question-asking and detection of indeterminacies involved in multiple-case critical analysis is more complex than is involved in single-case analysis, and that the difference is not simply quantitative. In particular, complex indeterminacies of interpretation not visible in the critical analysis of a single case suddenly emerge for students when reading multiple related cases. Attempting to read "across" multiple cases by simply replicating what one does when reading a single case is a "method" that is likely, more often than not, to break down. For this reason we theorize that comparing students' single and multiple (or, "cross-case") reading abilities should be a goal in designing a test of case-reading competence.

Phase 1

Summary of Findings

In the first phase of our research, we designed a prototype test version (TV1) in which we attempted to operationalize the five foregoing theoretical assumptions. We also field-tested this prototype with 161 first-year law students sampled from five different law schools. Because we detailed our rationale for the cases we used and our content-validation processes (using a panel of legal experts) for TV1 in our first technical report (Stratman, Evensen, & Oates, 2005), we will not provide that information again here. Instead, we will confine ourselves to summarizing the following:

1. The basic task demands presented by the test and the cases focused in it
2. The 2 × 2 grid that provided the conceptual basis for constructing and differentiating test items
3. Our hypotheses and key findings about students' field-test performance in relation to this grid
4. The key conclusions and questions arising from these findings that we targeted in our second study

Task Demands

TV1 required students to read three court opinions, from three different cases, that have been used in previous empirical studies of law students' case reading (Stratman, 2002, 2004).[1] These cases all tested the limits of the same Pennsylvania procedural rule affecting appeals from arbitration and thus were closely related to each other in terms of legal issues. Two of the cases referred explicitly to each other, and significant indeterminacies of interpretation arose both within and among all three. Before students began to read these cases, they were instructed to assume the role of an advocate (attorney) representing the defendant who lost in one of the cases. As a specific purpose and legal exigency for reading the cases, students were instructed that they were being asked to plan an appeal of the adverse decision on behalf of this defendant. The directions stressed that students needed to bear this purpose in mind while they read and analyzed

[1] A more detailed discussion of these cases and their legal issues can be found in our first technical report (Stratman, Evensen, & Oates, 2005).

the cases. However, students were also told that they could read the cases in any order or manner they chose, and were free to take notes (on the cases themselves or on a legal pad) as they worked. They were also free to use highlighter pens and had access to a legal dictionary.

Matrix Underlying TV1 Item Design

In keeping with the theoretical assumptions outlined above, we created four kinds of multiple-choice test items based on a 2 × 2 matrix (Table 1). One dimension in this matrix contrasts students' ability to read and analyze *single* cases with their ability to read and analyze *multiple* cases (single-case vs. cross-case items). The second dimension contrasts students' ability to *accurately recognize case content regardless of their purpose* with their ability to *identify purpose-relevant questions about indeterminacies of interpretation* (determinate vs. indeterminate items). The entire test contained 14 items, in a distribution also shown in Table 1. To help illustrate the distinctions between the four item types, each cell in Table 1 shows an actual sample test question, plus two (of the five) possible answers (i.e., the correct answer, and one incorrect answer).

TABLE 1
Initial test design, with sample question types

	Single-Case Items[a]	Cross-Case Items[b]
Determinate Items (Possible answers to these items are *declarative statements* whose accuracy students must judge.)	Which one of the following statements best summarizes the *reasoning* in the minority opinion (dissent) in *Meta*? d) (Incorrect) Citing the *Smith* case, the minority dissents because it sees no reason to question legislative acts that have remained unchanged for over 100 years. e) (Correct) The minority dissents for the following reasons. It argues that, in order to vacate the statute, the majority should have made (but failed to make) an argument that the record cost statute is unreasonable, and it argues that the defendant had a responsibility to learn the correct amount of costs. (4 items)	Which one of the following statements best summarizes the similarities and differences between the courts' reasoning in *Meta* and the court's reasoning in *Black and Brown*? a) (Incorrect) The reasoning used by the two courts is the same: both courts agree that the record cost prepayment requirement serves no real function as a condition of appeal and that record costs in comparison with dollars in controversy make the latter costs *de minimis*. b) (Correct) The reasoning used by the two courts differs: although both courts refer to the principle of the sufficiency of substantial compliance in their decisions, the *Meta* court further reasons that the record cost requirement is a matter *de minimis* because in comparison with the amount of money in controversy in arbitration decisions, record costs are usually quite small. The *Black and Brown* court does not discuss this disparity in amounts involved. (4 items)
Indeterminate Items (Possible answers to these items are *posed as questions.* Students must choose the question most relevant to their legal purpose and exigency).	All court opinions are interpretations, and interpretations invite questioning. Assuming that you decide to appeal the *Mackey* decision to Superior Court, which one of the following questions presented by the *Mackey* opinion is most relevant to this task? d) (Incorrect) Why does the court rely on its own "inspection" of the pencil note rather than calling for or inviting an expert witness to offer testimony concerning its legibility? e) (Correct) Why does the court note each of Mackey's claims about his attempt to contact the Prothonotary's office yet offer no reasons for concluding that Mackey did not make either an "honest effort" or "a valid attempt to make … full payment"? (3 items)	Because the *Mackey* case was decided in a lower level court (Court of Common Pleas), an appellate court judge reviewing *Mackey* would need to examine decisions in higher appellate courts to learn the applicable law. Given this circumstance, which one of the following questions about the relationship between the *Mackey* decision and the other two Superior Court decisions (*Meta* and *Black and Brown*) is most important for you to think about? a) (Correct) Can the *Mackey* court, without qualification, conclude from the decisions in either *Meta* or *Black and Brown* that record costs must be paid in twenty days? b) (Incorrect) Can the *Mackey* court properly conclude that "it is within the court's discretion to decide what reasonable conditions can be imposed" upon the right of appeal? (3 items)

[a]Question targets only one of the three cases.
[b]Question targets two or three of the three cases.

TV1 Hypotheses and Key Findings

The hypotheses we tested using TV1 fell into three categories: (a) those concerning the performance of students taking the test in their first semester (fall 2003, $n = 81$) of law school versus those taking it in their second semester (spring 2004, $n = 80$); (b) those concerning students' comparative performance on the four different item types shown in Table 1; and (c) hypotheses concerning students' performance on this test in relation to other measures of achievement, including LSAT, and first-year law school grade point averages (GPAs).

Because to our knowledge no similar test of students' legal case reading and reasoning skills exists, we could not form hypotheses about how well students might perform overall. Nevertheless, we report key overall findings first because they will help contextualize findings specifically related to our other hypotheses. The mean score across all test participants ($n = 161$) was 7.91/14 answers correct, or 56.5% (SD = 2.08; SE = .16). One-way analysis of variance confirmed that there was no statistically significant difference among the mean scores of the five participating law schools. In fact, different schools' mean scores fell in a very narrow range, from a high of 7.96 to a low of 7.81. Total individual scores ranged from a perfect score of 14/14 (one student) to a low score of 3/14 (two students). On the most difficult item on the test, only 36% of the students got the correct answer; on the easiest item, 92% of students got the correct answer (thus, no item was answered correctly by 100% of the students). Additionally, as shown in Figure 2, the combined scores for the spring and fall semesters were fairly normally distributed.

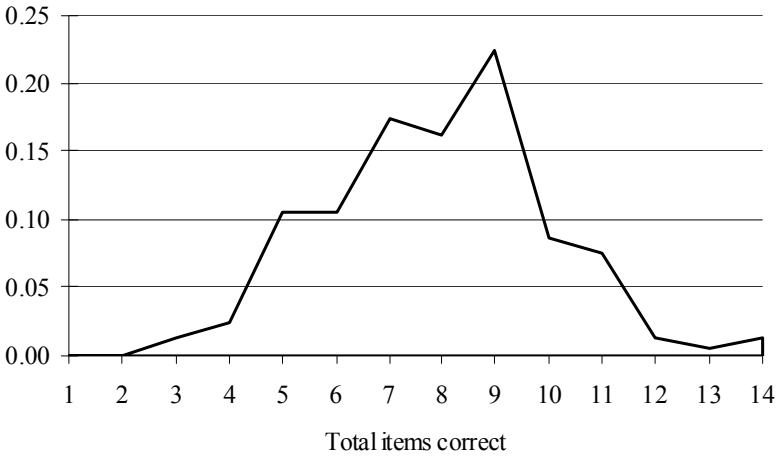


FIGURE 2. *Distribution of total test scores*

Developmental Hypothesis

We had hypothesized that the mean score of students taking the test in their second semester would be significantly higher than that of students taking the test in their first semester, owing to increased classroom practice in answering determinate-type questions about case content. To investigate this hypothesis, we tested fall and spring students' samples matched on the basis of LSAT scores, rather than a conventional within-subjects, pre-test/post-test design. Our results, however, failed to support the expectation of improved scores. As Table 2 shows, fall and spring students' mean test scores scarcely differed (fall M = 7.96; spring M = 7.86), while the mean LSAT score of students for both semesters was identical: 156. Table 2 also shows that there was no significant difference between fall and spring students across different test item types. Because of the obvious limitations in our research design, however, these results cannot be taken as conclusive that students' case reading and reasoning skills do not improve between first and second semesters in law school. Indeed, as we discuss later in this report, we designed Phase 2 of our study to provide more valid data indicating whether or not students' case-reading skills improve between second and third semesters of law school, and also whether they improve over all 3 years.

TABLE 2
Comparison of fall and spring semester students' mean scores by item type

	Single-Case/ Determinate (n = 4)	Cross-Case/ Determinate (n = 4)	Single-Case/ Indeterminate (n = 3)	Cross-Case/ Indeterminate (n = 3)	All Items (n = 14)	Mean LSAT Score
Fall (n = 81)	2.84	2.26	1.57	1.30	7.96	156
Spring (n = 80)	2.77	2.07	1.46	1.57*	7.86	156

*p = .07

Hypotheses About Relative Difficulty of Item Types

We hypothesized that cross-case, indeterminate test items would prove most difficult, while single-case determinate questions would prove least difficult. The data support this hypothesis. Substantively, students' average proficiency in getting correct answers was progressively reduced as they shifted from dealing with determinate items to dealing with indeterminate items, and also as they shifted from dealing with single-case items to multiple, cross-case items. As Table 3 shows, the mean percentage of students getting correct answers on each type of item progressively decreased. For example, for the single-case/determinate items, we examined the percentage of students who answered each of these four items correctly (i.e., out of 161), then calculated the mean of these four percentages. Similarly, for the single-case/indeterminate items, we examined the percentage of students who answered each of these three items correctly, then calculated the mean of these four percentages. Table 3 shows the negative, compounding effect of combining these underlying dimensions on students' performance.

TABLE 3
Mean percentage of students with correct answer for each test item type

Single-Case/ Determinate	Cross-Case/ Determinate	Single-Case/ Indeterminate	Cross-Case/ Indeterminate
.70	.54	.50	.47

Importantly, as we had also hypothesized, the indeterminate variable had a more pronounced negative effect on students' overall scores than the cross-case variable. As Table 4 shows, the biggest drop in the mean percentage of students getting correct answers occurred between the determinate and the indeterminate dimension (.14). In contrast, the drop between the single-case and cross-case dimension was .10.

TABLE 4
Impact of main item dimensions on mean percentage of students with correct answer

Determinate Items	Single-Case Items	Cross-Case Items	Indeterminate Items
.62	.60	.50	.48

We also obtained somewhat counterintuitive results when we examined the pair-wise correlations between total scores on the two main variables. First, students' total scores on single-case items (i.e., when both determinate and indeterminate items were included) showed only a very small positive correlation with their scores on the cross-case items, r = .15 (p < .05). We had predicted that this correlation would be positive because it seemed reasonable that skill in single-case analysis would be a necessary but not sufficient skill for dealing with multiple related cases. But because of significant cognitive differences in these two skills, plus a lack of instruction in cross-case analysis, we had hypothesized that the correlation between these scores would fall below r = .50. This result thus lends support to the theory that moving from single- to cross-case analysis not only constitutes an important developmental threshold for law students, but also supports Fajans and Falk's (1993) view that improvements in single-case analytic skill do not necessarily lead to corresponding improvements in cross-case analytic skill. The lack of connection between single- and cross-case reading skill is further reinforced by the item-type intercorrelations shown in Table 5. There is only a small correlation between students' scores on single-case/determinate items and their scores on cross-case/determinate items (r = .26), and even a faintly negative correlation between their scores on single-case/indeterminate items and their scores on cross-case/indeterminate items (r = −.01).

TABLE 5
Correlation matrix for four item type scores and total test score

	Total Score	Single-Case/ Determinate	Cross-Case/ Determinate	Single-Case/ Indeterminate	Cross-Case/ Indeterminate
Total Score	1.00				
Single-Case/ Determinate	.56	1.00			
Cross-Case/ Determinate	.69	.26	1.00		
Single-Case/ Indeterminate	.43	.03	.02	1.00	
Cross-Case/ Indeterminate	.56	.06	.16	– .01	1.00

Second, and also somewhat counterintuitively, students' total scores on determinate items (i.e., when both single- and cross-case items are included) showed only a very small positive correlation with their total scores on indeterminate items, $r = .13$ ($p < .10$). We had predicted that this correlation would be positive, but less than .50, because we believe that skill in identifying task-relevant indeterminacies of interpretation is partly dependent on skill in accurately retrieving and representing case information. This very low correlation, however, lends support to the theory that moving from the ability to accurately retrieve and represent canonical components of case discourse to the ability to identify task-relevant questions about indeterminacies of interpretation constitutes a quite challenging threshold for law students. More significantly, this correlation supports the idea that, beyond a certain point, increases in the ability to tackle determinate test items are unlikely to lead to corresponding increases in tackling indeterminate items. This apparent disconnection between the two skills is reinforced by the inter-item correlations shown in Table 5. There is almost no correlation between students' scores on single-case/determinate items and single-case/indeterminate items ($r = .03$); similarly, there is almost no correlation between students' scores on cross-case/determinate items and their scores on cross-case/ indeterminate items ($r = .16$)

Hypotheses Concerning Other Measures of Achievement

We hypothesized that students' overall scores on our test would correlate positively with their LSAT scores and also with their first-year GPAs. However, we expected both of these correlations to be low (well below .50), for two reasons. First, we expected the correlation with LGPA to be low due to the limited nature of instruction in case reading and case analysis at most law schools. Second, we expected the correlation with LSAT to be low because of the task-embedded nature of our test and the novel underlying construct for our test items, particularly the inclusion of indeterminate items as we have defined them.

The results in Table 6 generally support these hypotheses. Students' overall test scores did not correlate strongly with their first-year LGPA and most recent LSAT scores, $r = .22$ and $r = .22$, respectively. At the same time, as shown in Table 6, when we examine the correlations between the four different test item types and students' first-year LGPA, we again generally see very low correlations. The one possible exception is the correlation between students' LSAT scores and their scores on the single-case/indeterminate items.

TABLE 6
Correlation matrix for total test score, item-type score, LSAT, and LS-GPA

	Total Score	Single-Case/ Determinate	Cross-Case/ Determinate	Single-Case/ Indeterminate	Cross-Case/ Indeterminate
LSAT	0.22	0.11	0.05	0.29	0.07
LGPA	0.22	0.08	0.19	0.15	0.07

Conclusions and Questions Arising from Phase 1 Findings

Overall, the results of this first study provide tentative support for several conclusions. First, the study findings support the feasibility of designing a test assessing not only law students' ability to read for accurate understanding of the canonical components of case discourse (e.g., issue, holding, reasoning, rule, policy, dicta), but also assessing their ability to identify purpose-relevant questions about indeterminacies of interpretation surfacing in that discourse. In this respect, the findings comport well with the working assumptions about the nature of critical case reading that we outlined at the start of this report. The one assumption we did not test regarding this competency is whether—and if so, how—a shift in the reader's purpose results in a shift in the indeterminacies of interpretation that the reader should

query. In this pilot, we did not manipulate readers' purposes and legal exigency so as to observe the effect of such manipulations upon performance (Stratman, 2002), but instead used only one purpose and legal exigency.

Second, legal educators' anecdotal observations of students' difficulties in transitioning between two kinds of reading (reading for accurate understanding versus reading to generate strategically useful questions) are supported by the difference in students' performance on the determinate versus the indeterminate test items (Fajans & Falk, 1993). Similarly, legal educators' classroom observations concerning students' difficulty in transitioning from critically reading single cases to critically reading multiple, closely related cases find support in the difference between students' performance on single-case test items in contrast with their performance on cross-case items. The results thus reinforce recent recommendations in the legal education literature that more systematic instruction targeting both transitions should be provided in law schools (Lustbader, 1997; McKinney, 2005; Schwartz, 2001).

Third, the study findings suggest that, relative to standard measures of achievement such as the LSAT and law school GPAs, the test we developed appears to be tapping certain skills only weakly reflected in these measures. Some reviewers of our first research report expressed the concern that, for a test of critical case-reading skills to have external validity, it should correlate highly with LSAT scores. However, we would suggest that such a view merely presumes that the LSAT is already the "gold standard" for measuring such skills, whereas our goal in developing the test lies exactly in the opposite direction. Specifically, we sought to develop a test that would target at least some abilities *quite different* from those targeted by the LSAT but that would nevertheless reflect the constraints and challenges characteristic of "real-world" case reading and case reasoning.

We refer to all of the conclusions above as tentative because both our field-test findings and our experience attempting to develop the instrument have raised numerous questions. First, determining whether a test of critical case-reading skills is feasible requires determining whether the item typology can be replicated using a different set of cases than we used. For instance, we do not know if the apparent success we had in generating four distinct kinds of test items in TV1 is merely an artifact of the particular cases we used and their particular relationships to one another, or whether the typology is also usable in designing a test with very different cases. At the same time, the conclusions drawn above concerning the relative difficulty of the different skill constructs underlying the typology also call for an attempt at replication with a different set of cases. Students' relatively weaker performance in dealing with cross-case items and with indeterminate items may be due more to the cases used in TV1 than reflective of an actual developmental threshold between these skills. Perhaps most importantly, we do not know whether or not the test items developed for each category in the typology using different cases involve significantly different information-processing demands, as might be demonstrated, for example, in think-aloud data collected from students taking the test. For instance, while we might design cross-case/determinate items for a new set of cases that share the same surface characteristics as the items we created for this category in TV1, the new items might nevertheless require certain knowledge and cognitive or metacognitive behaviors of students *not* required by the "same" item type in TV1. Effectively, with only one test version focused on one set of cases, the construct validity of that single test remains unknown.

Additionally, while the data from our field test with TV1 suggest that students' critical case reading and case reasoning ability does not improve as a result of their first two semesters in law school, the research design makes the soundness of this conclusion subject to doubt. The chief reason is that we do not know whether or not the second-semester students would have performed the same as the first-semester students had we also tested the second-semester students in their first semester. In other words, we did not use a within-subjects,"pre/post" design. Since all students were volunteers, neither the fall semester nor the spring semester students constituted a random sample. Therefore it is possible that the students we tested in the second semester were simply a less proficient group to begin with. This possibility cannot be ruled out especially in view of the low overall correlation between students' LSAT scores and their scores on our test. LSAT scores were our only means of assuring some equivalence between these groups. Conversely, since we did not have the opportunity to retest the same first-semester students when they reached their second semester, we cannot safely conclude that they would have scored no better at that later time. To obtain more valid evidence concerning whether students' critical case reading and reasoning skills improve during law school, it is clearly desirable to develop a second, equivalent test that would permit testing the same students at two different points in time.

It was with these challenges before us that we designed and embarked upon Phase 2, described in the next section.

Phase 2

Designing an Alternate Form of the Case Reading and Reasoning Test

In response to concerns about both the internal and external validity of our first test, we designed a study that would involve the construction of an alternate form. Our goal was to design a second test version (TV2) whose items and their underlying skill constructs would closely parallel those in the first, piloted version (TV1). To achieve this parallel structure, the task necessarily required (a) three cases that approximated both in length and complexity the three cases used in TV1; and (b) an equal number of items that corresponded to two comprehension levels (single- and cross-case) and two semantic levels (determinate and indeterminate). In addition, as was the case in the development of TV1, these cases, tasks, and items were to be evaluated by expert reviewers and subjected to pilot testing.

Selecting and Editing the Three Cases

As with the cases chosen for TV1, we looked for three cases for TV2 that presented the kind of indeterminacies of interpretation that legal educators have complained students are poor at analyzing or even recognizing (Fajans & Falk, 1993). Also as with TV1, we made some alterations to these cases, both in order to make their relationships to each other roughly parallel to the relationships among the TV1 cases and to bring them closer to TV1 cases in their overall length. Before we provide summaries of each of the TV2 cases, we first describe some general similarities and differences between TV1 and TV2 cases.

Similarities between TV1 and TV2 cases. First, just as the TV1 cases include two decisions from the state appellate court level and one from a lower court, so in TV2 the cases include two federal circuit court decisions and one lower, district court opinion. Second, as in TV1, students in TV2 are informed that their purpose for reading the cases is to plan an appeal of the lower (district) court decision on behalf of the plaintiff who lost that decision. Specifically, the three TV1 cases deal with whether defendants satisfied the terms of a state procedural statute for perfecting appeals from arbitration; the three TV2 cases deal with plaintiffs who claim to have had their employment welfare benefits unfairly denied or reduced by their employers under the Employment Retirement Investment Security Act, or ERISA. Third, just as a key problem with the lower court decision in TV1 involved problematic reasoning and an arguable failure to fairly analyze controlling precedent (i.e., one of the other two cases), so also in TV2 the lower (district) court arguably fails to fairly use persuasive precedent (i.e., one of the two circuit court opinions). Fourth, one of the TV1 cases overrules and reverses another TV1 case; somewhat similarly, in TV2 one of the circuit cases is a reversal of a lower court opinion (although the latter opinion is not one the students read). Finally, just as we did with the TV1 cases, we manipulated the TV2 case dates so that the lower district court case occurred chronologically *after* the two circuit court decisions.

Differences between TV1 and TV2 cases. Having noted these broad parallels, however, we do not claim that the two sets of cases are isomorphic: There are a number of potentially important differences. Indeed, in order to conduct a varied replication of the item typology used to design TV1, the TV2 cases *should* differ. First, the TV1 cases combined are shorter in total length than the TV2 cases: TV1 cases total 4,391 words, while TV2 cases total 5,970 words (an increase of 26.5%). Second, substantively, whereas the lower court decision in TV1 explicitly cited the two previous appellate court decisions that students also read, the district court decision in TV2 cited only one of the two related circuit court decisions that students also read. Third, whereas the lower court decision in TV1 narrowly focused on the interpretation of a state procedural statute containing two distinct elements, in TV2 the district court case disposed of two distinct issues arising under different sections of ERISA. Fourth, and finally, while one of the TV1 cases included an important dissenting opinion, none of the TV2 cases did so.

The Three New Cases

The following summaries of each of the TV2 cases and their relationship to each other will help clarify their similarities and differences in comparison to the TV1 cases. As part of these summaries, we will also point out some of the significant indeterminacies of interpretation that arise within and between the cases, which in turn became the focus of our indeterminate TV2 test items.

Bridell v. Ribier Clothiers, Inc. (1991)

This is the district court case whose plaintiff is identified in TV2 test directions as the students' client (Joan Bridell). As their purpose for reading, students were instructed to represent Ms. Bridell as their client and to consider appealing her loss in the U.S. Southern District of Texas (Houston Division) court to the U.S. Fifth Circuit. Notably, this case is actually a controversial 1990 ERISA decision, *McGann v. H & H Music Co.* (742 F. Supp. 392, 1990), whose parties we renamed and whose date we changed.[2] As described below, we also slightly altered the facts and reasoning to bring this decision into greater potential conflict with the two other earlier circuit-level cases that TV2 students read, *Hamilton v. Air Jamaica Ltd.* (1991), and *Alexander v. Primerica* (1991). In *Bridell*, the plaintiff had worked for Ribier Clothiers for about 5 years when, in 1987, she contracted AIDS. She notified her employer of her illness and began collecting company-provided medical insurance. At the time she notified her employer of her illness, the lifetime total medical coverage offered to AIDS patients was $1,000,000. However, in July 1988, the company notified all employees that it was reducing AIDS medical benefits to a lifetime total of $5,000; at the same time, it also notified employees that it would no longer provide medical coverage for chemical dependency. In response, Bridell filed suit in which she made two complaints: first, that under ERISA the company had no right to reduce the lifetime amount it had previously

[2]See, for example, Dirrim, C. (1993). Unpopular but not unfair: The Fifth Circuit considers the terms but ignores the endearment in *McGann v. H & H Music Co*. *Nebraska Law Review*, *72*, 862–886.

indicated it would provide in its health benefit Summary Plan Document (SPD); and second, that in acting to reduce its lifetime medical benefit for employees with AIDS, the company engaged in illegal discrimination under ERISA.

On summary judgment, the court decided in favor of the defendant Ribier Clothing, indicating that it could find no genuine issue of material fact upon which either of Bridell's claims could be based. Specifically, the court pointed to a reservation of rights disclaimer in the company's SPD. Under ERISA, if companies provide health care or other "welfare" benefits to their employees (e.g., severance or disability pay), they must describe these benefits and any limitations upon them in SPDs. The disclaimer in Ribier's SPD was presented in a footnote and stated that "The Plan Sponsor [Ribier Clothing] may terminate or amend the Plan or terminate any benefit under the Plan *at any time* pursuant to relevant changes in federal legislation." The court reasoned that this disclaimer permitted the company to reduce its medical coverage for AIDS at its discretion, and that the disclaimer's inclusion in the SPD meant that Bridell was never "promised" a lifetime benefit. Without analyzing the language, placement, or understandability of the disclaimer itself, the court nevertheless declared that the disclaimer satisfied ERISA's "plain language" requirement, that is, that SPDs "must be written in a manner calculated to be understood by the average plan participant (29 U.S.C. § 1022(a)(1))." The absence of such analysis constitutes one indeterminacy of interpretation that students sensitive to their purpose for reading the cases need to ask questions about. The court also reasoned that "Bridell has not and could not reasonably contend that the SPD's … reference to 'lifetime medical benefits of up to $1,000,000' for eligible employees and their spouses somehow renders their reservation of rights in the disclaimer ambiguous."

In support of this reasoning, the court cited a third circuit decision, *Hamilton v. Air Jamaica, Ltd.,* 945 F. 2d 74 (3d Cir. 1991), one of the two other cases that TV2 students read. *Hamilton* also involved a contested reservation of rights disclaimer, and the *Hamilton* court also upheld the defendant company's use of the disclaimer as a basis for reducing benefits described in an employee handbook. Notably, however, we altered the disclaimer in Ribier's SPD (i.e., in *Bridell*) so that it included the phrase "pursuant to relevant changes in federal legislation." This phrase was *not* contained in Air Jamaica's otherwise similar disclaimer, yet the court in *Bridell* makes no mention of this difference between the two cases even though it discusses *Hamilton* as supporting its decision. This lack of mention of the disclaimer's difference in *Hamilton* is one of the problematic indeterminacies of interpretation arising between the *Bridell* and *Hamilton* decisions that TV2 students, in their role as advocates for Joan Bridell, need to ask questions about. Finally, the *Bridell* court also rejected Bridell's discrimination claim, reasoning that Bridell failed all three parts of a three-pronged discrimination test set forth in an earlier ERISA case (*Gavalik v. Continental Can Co*., 812 F.2d 834 [3rd Cir. 1987]).

Hamilton v. Air Jamaica (1991)

This case is an actual third-circuit ERISA decision that we edited, primarily for length (945 F.2d 74, 3rd Cir. 1991). The plaintiff had worked for Air Jamaica for about 13 years when his position was terminated in May, 1989. An employee handbook that he had been given in March 1989 indicated that, as an exempt employee, Hamilton would receive 4 weeks of severance pay for each year he had worked for the company. However, 1 week before he was terminated on May 31, 1989, Hamilton and other employees received a company letter indicating that at termination they would receive severance only in the amount of 2 weeks for every year of employment. After his petition to the company to provide the larger severance amount indicated in the employee handbook was denied, Hamilton sued in district court for recovery of the larger severance pay under ERISA. The district court awarded judgment in favor of Hamilton, but on appeal to the third circuit, Hamilton lost and consequently only received the reduced severance pay amount indicated in the May 1989 letter.

In reversing the district court decision, the third circuit reexamined two arguments that Air Jamaica had made. First, although the handbook Hamilton was originally given did indicate that his severance benefit as an exempt employee would amount to 4 weeks pay for each year worked, the company argued that the handbook was not its ERISA plan. Air Jamaica argued that the severance amount printed in the handbook was an editorial error, not its actual (if unwritten) policy. In support of these claims the company further argued that, in practice, it had not previously paid 4 weeks severance to any terminated exempt employees. Second, the company argued—as was the case in *Bridell*—that the handbook contained a disclaimer stating that it reserved the right "to alter, reduce or eliminate any pay practice, policy or benefit, in whole or in part, without notice." The company also argued that the handbook contained a "statement of understanding," which employees were instructed to date, sign, and return to their supervisors, indicating that they understood this disclaimer as well as other handbook contents. Importantly, while the original, unedited third-circuit opinion in *Hamilton* did not make it clear whether or not Hamilton signed this statement, we altered the original facts in our test version of this case to state definitively that he did *not* sign it and did *not* return it.

Responding to the first argument, the circuit court agreed with the district court's reasoning from the precedent *Hozier* decision. *Hozier* stated that, to be enforceable, ERISA plans must be written, and therefore Air Jamaica could not claim that its "actual" severance practice, as an "oral plan," could take precedence over what its handbook explicitly stated in print. However, the circuit court, in reversing the district court, accepted Air Jamaica's second argument. The circuit court reasoned that Air Jamaica's disclaimer "is part of the [company's] written promise [about benefits] and a limitation upon it." The circuit court rejected the district court's contention that the disclaimer inherently violated

ERISA's goal of providing "security in benefit packages." Significantly, and as was the case in *Bridell*, the *Hamilton* court also does not analyze the language, placement, or understandability of the handbook disclaimer. In particular, the *Hamilton* court does not discuss the fact that Hamilton did not sign the "statement of understanding" regarding the disclaimer, nor does it discuss what Hamilton's missing signature might or might not mean in relation to ERISA's requirement that SPDs "be written in a manner calculated to be understood by the average plan participant (29 U.S.C. § 1022(a)(1))." Thus, in addition to the difference in the wording of the disclaimers contested in *Bridell* and in *Hamilton*, this lack of discussion of Hamilton's missing signature constitutes another problematic indeterminacy of interpretation that TV2 students, in their role as advocates for Joan Bridell, may need to ask questions about.

Alexander v. Primerica Holdings Inc. (1991)

Like *Hamilton*, this case is also an actual third-circuit ERISA decision that we edited, primarily for length (967 F.2d 90, 3rd Cir. 1992). As with the *Bridell* and *Hamilton* cases above, this case also involves the plaintiffs' claim that their employer violated ERISA by reducing benefits that were allegedly promised in the company's summary plan documents. In this case, the plaintiffs were retirees whose health insurance premiums were increased significantly by the company. The plaintiffs argued that their lower, existing premiums were a lifetime benefit that could not be increased, whereas Primerica argued—just as Ribier and Air Jamaica had argued—that the summary plan documents contained a disclaimer giving the company an absolute right to raise premiums or reduce other features of its health benefit plan. Unlike the *Bridell* and *Hamilton* cases, however, the interpretation of disclaimer language and its meaning was the only issue examined in our shortened version of the *Alexander* opinion.

Specifically, the court analyzed several versions of a reservation of rights disclaimer that Primerica used in its summary plan documents. The plaintiffs argued that the disclaimer language did not give the company the unfettered right to increase premiums or reduce benefits; the defendant argued that it did. After carefully weighing the arguments offered by both parties, the court determined all of Primerica's disclaimers to be ambiguous, and remanded the case back to the district court. In doing so, the court indicated that the district court must examine whether any evidence—including Primerica's previous behavior in operating its health plan—clarified how the disputed disclaimer should be interpreted. Importantly for students taking TV2, Primerica's disclaimer is almost identical to the disclaimer in *Bridell*. Note particularly the italicized portion of the first sentence:

> The Company expects to continue this Plan indefinitely, but necessarily reserves the right to amend, modify, or discontinue the Plan *in the future in conformity with applicable legislation* [italics added]. The Plan does not provide for benefit payments in any case or under any condition not identified and provided for in this booklet. The Group Insurance Policy and the certificates thereunder issued by the Insurance Company are consistent with the terms and conditions outlined in this booklet.

Just as the disclaimer in *Bridell* contained the phrase, "pursuant to relevant changes in federal legislation," here the Primerica disclaimer contains the phrase, "in conformity with applicable legislation." This similarity is of potential strategic importance to the students taking TV2, since they are directed to consider an appeal on behalf of Joan Bridell, and the *Bridell* court failed to cite *Alexander*—despite the fact that the *Bridell* court did cite *Hamilton,* also a third-circuit decision*,* with its rather different disclaimer. Indeed, the *Alexander* court points out how different the Primerica disclaimer language is compared to the language of the disclaimers evaluated in other precedents, including the disclaimer in *Hamilton*.

To summarize, then, just as we attempted to do in TV1, in TV2 we attempted to provide students with cases that manifest indeterminacies of interpretation, both individually and in their discursive relationships to each other. As we discussed in setting out our theoretical assumptions about competent case reading and reasoning, some of these indeterminacies of interpretation should become more identifiable to student readers the more seriously and carefully they take their assigned purpose and exigency. At the same time, we chose cases for TV2 whose issues and law would likely be unfamiliar to most, if not all, students. We altered the names of parties in the *McGann v. H & H Music Co.*, case out of a concern that, given the attention paid to it by some ERISA scholars, some third-year students would have encountered it previously.

Methods of Test Construction

Writing the Items

We implied in the previous section that the editing of the cases and the writing of items was an iterative process. The description that follows, focused as it is on the item-writing process, loses the dynamic nature of the process, but presents the central foci of each level of writing, reviewing, and revising.

Writing a draft version of the test (Summer 2005). A three-person group (Evensen, Stratman, and Oates) comprised the item-writing team. The process used in constructing these items was necessarily constrained by the theoretically derived and empirically tested structure that governed the creation of TV1, and a careful parsing of the three cases outlined above. Items in specific categories were originally drafted by one team member (Stratman, determinate items; Evensen, indeterminate items) and then sent to the other person for response, discussion, and revision. These items were then sent to the third team member (Oates), who reviewed the items particularly for their legal content. This process resulted in the *alpha* version of TV2 that consisted of 20 items distributed as shown in Table 7.

TABLE 7
Alpha version of TV2

	Determinate Items	Indeterminate Items
Single-case items	5	4
Cross-case items	5	6

These items consisted of a stem, which specified the reading/reasoning task, and five possible choices, one of which was keyed as "correct." At this point, no formal justifications were written for correct versus incorrect choices.

Expert review of alpha version (September 2005).[3] The first level of expert review involved eight readers, all law professors. Each received the alpha version of TV2 (3 cases, directions regarding the overall task, and 20 multiple-choice items) by mail. These reviewers had agreed to reserve up to 2 hours to complete the task of "taking the test." They were not given any keys to the test. Indeed, because this was an early version of TV2, they were told that although we deemed one answer to each item as "best," we also realized that there could be good arguments against our preferences.

Specifically, we asked each reviewer to indicate his or her first and second choices for each item, to mark choices believed to be "nonsensical," and to jot down general reactions to the test. These reviewers subsequently participated in a telephone interview with one of two lead item writers (Evensen or Stratman) to share their responses and comments. Each received a $125 honorarium for participation.

The outcome of this review was generally positive and yielded much information to guide the revision process. Sixteen items survived the review. Items were revised and justifications for each correct and incorrect choice were crafted. This process resulted in a beta version of TV2 that was prepared for a second level of expert review.

Expert review of beta version (December 2005). This second level of review involved three legal experts: one a director of a legal writing program, another a director of academic support and a reading specialist, and the third a lawyer and Ph.D. candidate in educational psychology. This group was given the beta version of the test along with keyed responses and justifications. The reviewers were asked, however, to attempt to "take the test" before looking at the keyed responses and justifications. They were asked to critique all materials in their final evaluations.

This second group of experts met with one item writer (Evensen) for a half day to discuss their responses. They were given an $800 honorarium for participating. Fourteen items with justifications survived this review, and the resulting item types were distributed in the same pattern as in TV1: 4 single-case/determinate; 4 cross-case/determinate; 3 single-case/indeterminate; and 3 cross-case/indeterminate. Revisions to items retained from the alpha review were negotiated; in a few cases substantial changes were made, but in many cases the changes were quite subtle. The beta review resulted in a final version of the test that was prepared for field-testing.

Field-Testing Test Version 2

Designing the Study

The field-testing of TV2 was to occur concurrently with further testing of TV1. Data collection was scheduled during the spring 2006 and fall 2006 law school semesters.[4] The following three research questions were pursued:

[3] *A note on expert reviews in TV1 versus TV2.* The reviewing process adopted for TV2 differed in some respects from the one used in the development of TV1. In retrospect, we found that the reviewing process used in TV1 (i.e., setting up three independent panels of experts meeting face-to-face with one team member) produced highly redundant information. In addition, because in the development of TV1 only a single level of expert review occurred, we were concerned that some items might not have been subjected to an adequate level of scrutiny before pilot testing. Thus, we revised the expert reviewing process to include two levels.

[4] Our original proposal had scheduled data collection during the fall 2005 and spring 2006 semesters, but institutional constraints forced us to delay our data collection and shift comparisons from within the first year to between the second semester of the first year and the first semester of the second year.

1. *Do students' case reading and reasoning skills improve between their first and second years in law school?*

   To get a clearer answer to this question, we proposed collecting data from approximately 150 first-year students using TV1 and TV2 in a counterbalanced design. Specifically, we planned to administer TV1 to about 75 first-year students, and TV2 to another 75, in the spring 2006 semester; students were to be randomly assigned to one of the two test versions. Students were to be recruited from the same five schools that we sampled in our first study. Members of this cohort of 150 would be approached for retesting with either TV1 or TV2 in fall 2006 (i.e., whichever version students did not take in spring 2006). This design additionally would allow us to rule out any confounders due to test version order.

2. *Do students' case reading and reasoning skills improve between their first and third years in law school?*

   This comparison involved administering TV2 in spring 2006 to a subgroup of the same students who took TV1 (n = 161) during fall 2003 or spring 2004. We aimed at getting about 20 volunteers from each of the five participating schools or a total of about 100 students.

3. *Can the underlying 2 × 2 question typology developed for TV1 be replicated in TV2, thus advancing the construct validity of this typology?*

   We planned to approach this question in the following ways:

   a. By comparing statistically the same students' results on both versions of the test to see if their mean scores and score distributions were similar or different.

   b. By determining whether both test versions resulted in the same compounding effect produced by combined question types that we observed in our first study and reported earlier in this report. This compounding effect demonstrated that items became increasingly difficult for students as they moved from single-case/determinate items to cross-case/indeterminate items.

   c. By testing statistically for construct validity. Analyses were performed to determine whether the probability of correctly responding to each theoretically parallel item was comparable between forms. Although logistic regression procedures associated with differential item functioning (DIF) were originally proposed to undertake this step, the nature of the dataset collected and limitations of the sample size prohibited this type of analysis. Rather, a combination of several established DIF methods as well as other statistical methods that have not yet been previously used for DIF were used in this study to examine the equivalence of paired items on the test.

   d. By examining students' score results to see to what extent these scores correlated with students' LSAT, law school GPA, and some individual course grades (e.g., grades from first-year lawyering or legal writing). As in our first phase, we again expected positive, but somewhat low correlations (i.e., <.50).

   e. By examining concurrent think-aloud data collected from students taking both test versions to determine whether students' use the same cognitive, metacognitive, and behavioral processes when responding to the same types of questions and whether they use the same processes when taking both versions of the tests. Approximately 32 students for the think-aloud task were to be solicited from a sixth law school and asked to respond to eight items, two of each of the four item types, on either TV1 or TV2. We adopted this strategy because of the fatigue factor that accompanies the think-aloud task, and because we only collected six usable think-aloud tasks in Phase 1 of this project. In addition, during this administration, a researcher was present to prompt the reader to talk continuously through the task. The procedure was to be practiced with participants before the session began. The data generated from this part of the study was deemed key to substantiating validity claims and would allow us to find evidence of constructs and subskills foundational to the design of the tests.

Participants and Procedures

A total of 239 students participated in this second phase of the study. In spring 2006, 146 second-semester, first-year students took either TV1 or TV2; 83 students from this group were retested with the alternate form of the test early in the first semester of their second year of law school (fall 2006). In spring 2006, which for all was their final semester in law school, 63 third-year students who had taken TV1 in either fall 2003 or spring 2004 volunteered to take TV2. Students

were about equally representative of each of the five participating law schools. The ratio of female to male participants was about 4:3.

As was our practice in Phase 1, these students received $50 to compensate for the 1–2 hours necessary for the task. All sessions were proctored. Students were given a packet of materials and took the test under standard silent reading conditions. They were urged to perform as best they could, being reminded of the importance of the research for legal education and the fact that they were receiving fair compensation for their time. The average time for completion was 1 hour, 10 minutes (range, 45 minutes to 2 hours).

In addition, 30 first-year students participated in the think-aloud portion of the study in spring 2006. As mentioned earlier in the report, these students were solicited from among the first-year class at a sixth law school. Students met in a private room with one of three test administrators trained in think-aloud methods. After practicing the think-aloud procedure with the test administrator and ascertaining that each was comfortable with the procedure, students were given the three cases from either TV1 or TV2 to read silently. They were also given the purpose of the reading task—to represent the losing party who wishes to petition a trial-level court to reconsider its decision or to appeal the decision to the next level of the judicial system. Students were encouraged to mark the copies or take notes as they normally would when reading cases and were given access to a legal dictionary. Once students believed that they had satisfactorily read the cases, they were given a set of eight multiple-choice items that represented subtests of either TV1 or TV2. Each set contained two each of the four item types associated with this test: single-case/determinate, single-case/indeterminate, cross-case/determinate, and cross-case/indeterminate. Students were asked to talk out loud about what they were thinking as they decided on best choices and to read and reason out loud as they selectively returned to the texts of the cases. Periods of silence lasting 30 seconds were followed by prompts. Times for the think-aloud portion of the activity averaged 33 minutes and ranged from 15 to 60 minutes. Sessions ended after students completed a short interview concerning their perception of the task and their confidence in their performance.

Results

Quantitative Findings

The Improvement Questions

Of particular interest in this study is whether students' case reading and reasoning skills improve as they progress through law school. Phase 1 of the study concluded that the students did not improve on their total test scores between their first and second semesters. Of course, in that study we had only one version of the test available (TV1) and because of that we were able only to compare matched groups.

In Phase 2 we wanted to examine this question further to see if the findings would be replicated when an alternate test (TV2) was used and a longitudinal design using the same cohort of test takers was possible. Thus, we were able to examine whether case reading and reasoning skills improved between semesters two and three, and between years 1 and 3.

Analyses of Test Scores From Second to Third Semesters

A total of 83 students took one form of the test during their second semester of law school and the second form of the test during their third semester. The order in which each test form was administered was counterbalanced and randomly assigned to the students. A total of 49 students received the TV1 test form first. The scores are described in Table 8. A total of 34 students received the TV2 test first.

First, the results show that these students' case reading and reasoning skills did not improve between their second and third semesters. As shown in Table 8, the students' mean test score in the spring was 7.88, and the students' mean test score in the fall was 7.67. Although the fall semester test scores were slightly lower, the difference was not significant. Moreover, TV2 was more difficult for students than TV1, regardless of when TV2 was taken. In both semesters, the students scored higher on TV1.

TABLE 8
Mean scores for TV1 and TV2 among second- and third-semester students (n = 83)

	Spring 06	Fall 06	Mean Total Score
TV1	8.41	8.18	8.3
TV2	7.35	7.16	7.25
Mean total score	7.88	7.67	

Note: When test scores were analyzed by law school, there were no significant differences between schools. However, the difference between TV1 and TV2 mean scores was significant only for the students who received TV1 first: TV1 (spring 06) $M = 8.41$ vs. TV2 (fall 06) $M = 7.16$ ($t = 3.457$, $p = .001$). The difference between TV1 and TV2 mean scores was not significant for the students who received TV2 first ($t = 1.601$, $p = .119$). Although we counterbalanced test-version order and randomized assignment as is methodologically appropriate given two test versions that are only theoretically equivalent, statistical tests for a test-order effect were nevertheless conducted and no significant effect was found.

Analyses of Test Scores From First To Third Year

A total of 63 students took the TV1 form during their first year in law school (during Phase 1 of the study) and the TV2 form during their third year in law school. A paired *t* test failed to find a significant difference between the TV1 and TV2 total scores: TV1 (first year) $M = 7.7$ vs. TV2 (third year) $M = 7.48$, ($t = 0.564$, $p = .575$). At the same time, there was also no significant difference on any of the four subscales.

One hypothesis that requires consideration given these findings is that TV2 is simply harder than TV1. Assuming that is the case, scores from first and third years were combined and subjected to mean equating in order to put the two forms of the test on the same scale. A constant of 0.80 was added to the TV2 mean, and longitudinal comparisons were reexecuted. Even with scaling, no significant difference was detected between test administrations for students who took the tests in their first and third years. Thus, as was found in the first study, it does not seem that students improve in their case reading and reasoning abilities, at least according to these particular instruments.

It must be mentioned at this point, however, that our instruments, TV1 and TV2, are very much under development and require additional scrutiny before any conclusions about law students' case-reading performance can be drawn. That issue is taken up in the next section of this report.

The Replication Questions

1. *Is the pattern of results obtained with TV1 in the Phase 1 study (2003–2004) replicated in the data for TV1 in the Phase 2 study (2006)?*

First, the Phase 1 study total score distribution shown earlier in this report (see Figure 2) is very similar to the total score distribution we obtained with TV1 in the Phase 2 study; the latter distribution is shown in Figure 3. Second, as shown in Table 9, the pattern of results across the four item types that we obtained in the Phase 1 study is replicated in the Phase 2 results. Indeed, the results are strikingly similar across three of the four item types. In particular, as occurred in the Phase 1 study, we see that the difference in the average percentage of students getting the correct answer on the determinate items compared with the indeterminate items is exactly the same (.14).

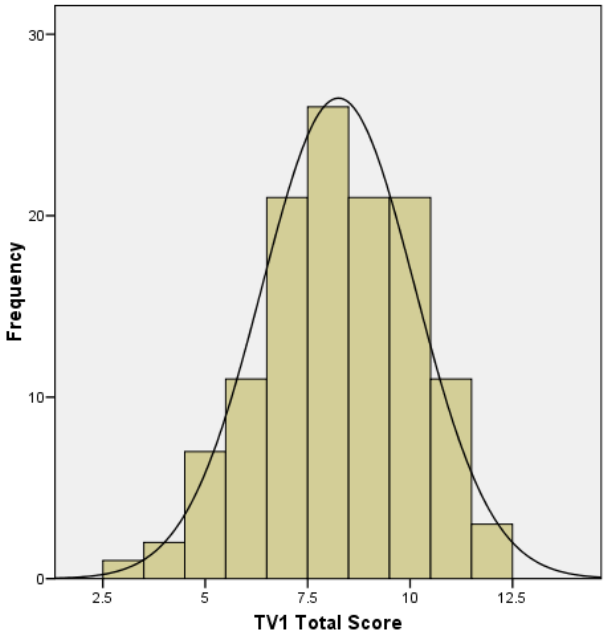
TABLE 9
Average difficulty levels for students taking TV1 in the Phase 1 and Phase 2 studies, across the four item types

	Single-Case	Cross-Case	Determinate	Indeterminate
TV1 Phase 1 study (2003–2004)	.60	.50	.62	.48
TV1 Phase 2 study (2006)	.65	.50	.64	.50

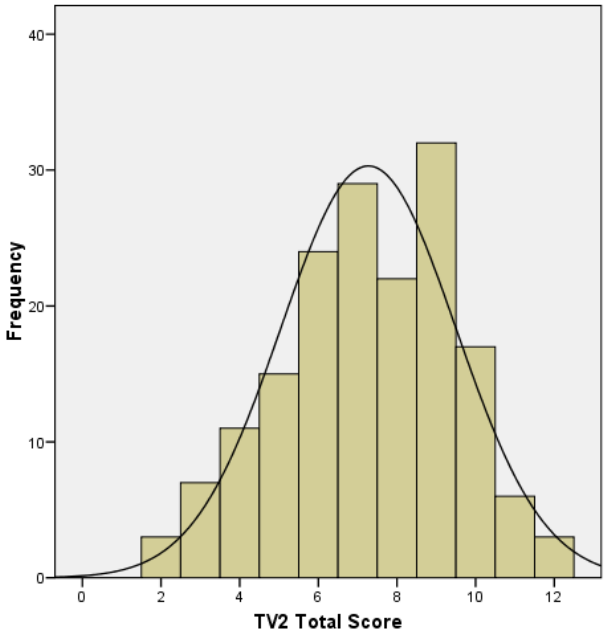
Third, just as we observed in our Phase 1 study with TV1, the correlation between students' scores on the single- and cross-case items in our Phase 2 study was positive but very small, $r = .15$; indeed, this is the exact same correlation that we observed in the Phase 1 study. Similarly, as in the Phase 1 study, the correlation between students' scores on the determinate and indeterminate items again was positive but very small ($r = .18$). In the Phase 1 study results for TV1, this correlation was $r = .13$. Collectively, then, we take these results to indicate a good replication of our results for TV1 in Phase 2 of our study.

2. *Is the pattern of results obtained with TV1 in the Phase 2 study replicated in the data for TV2?*

We address this question with five separate comparisons. First, we compare the total score distributions on the two different instruments. As shown in the two histograms in Figure 3, both distributions approximate a normal distribution and are also similar to each other. However, the mathematical mean for TV2 total score is lower than that for TV1, 7.27 vs. 8.25, respectively (each test has 14 items). Second, comparison of the average percentage of students getting correct answers on each of the four test dimensions across TV1 and TV2 yielded similar results. As shown in Figure 4, the average percentage of students who correctly responded to the determinate items was quite similar between the forms (64% for TV1 and 61% for TV2). The indeterminate items in TV2 seemed to be slightly more difficult than those in TV1 (40% versus 50%, respectively).



Mean = 8.25 (SD = 1.868)
N = 124



Mean = 7.27 (SD = 2.225)
N = 169

FIGURE 3. *Distribution of total test scores on TV1 and TV2 in Phase 2 study*

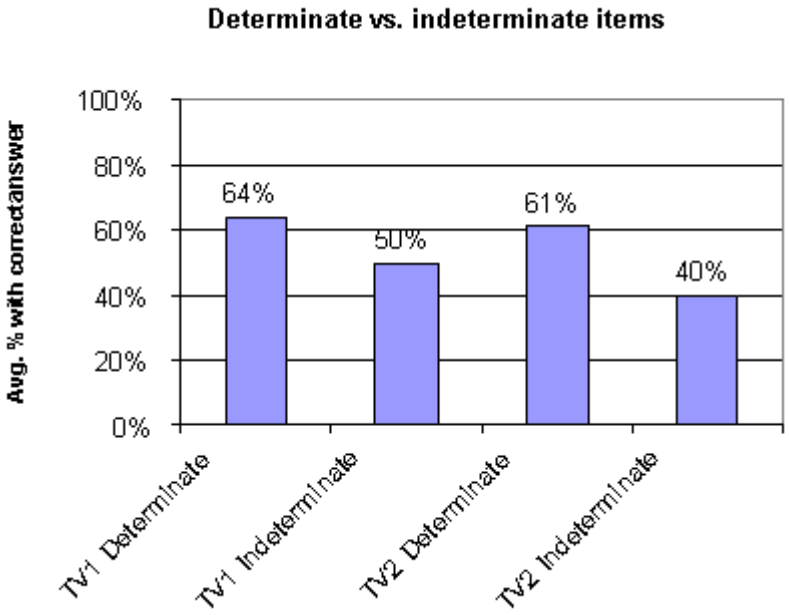


FIGURE 4. *Average difficulty levels for determinate vs. indeterminate items on TV1 and TV2*

Similar distributions between versions were also found based on the case dimension, as illustrated in Figure 5. Again, the TV2 form seemed to be slightly more difficult for both the single-case and the cross-case items.

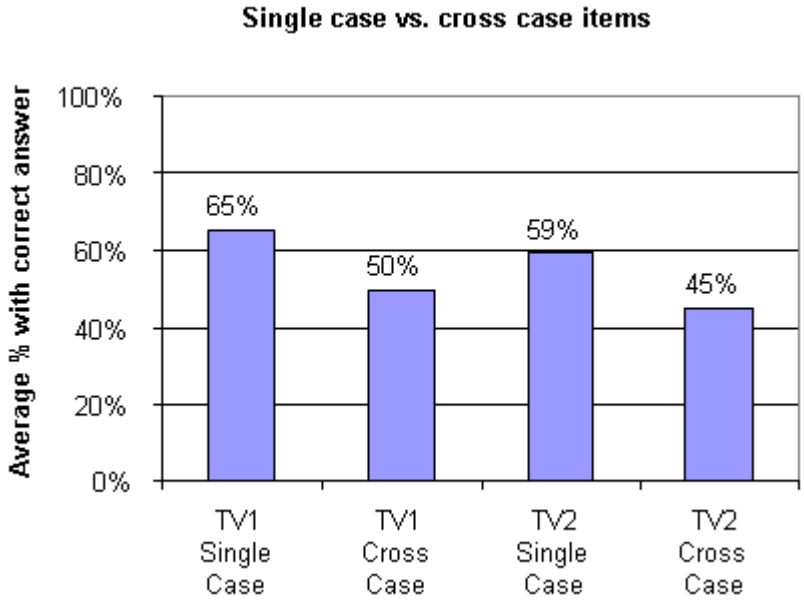


FIGURE 5. *Average difficulty levels for single-case vs. cross-case items on TV1 and TV2*

Third, when the four categories of item types were examined together, as illustrated in Figure 6, the average difficulty levels also followed a similar pattern between TV1 and TV2. However, with the exception of the cross-case determinate category, the items on TV2 were found to produce lower scores, supporting the hypothesis that they may be more difficult. This hypothesis is also supported by Figures 7 and 8, which illustrate the range of difficulty levels for the items in the TV1 and TV2 forms, arranged from most difficult to easiest.

Fourth, we observed both a similarity and a difference in the inter-term type correlations between TV1 and TV2. In TV1, students' scores on single-case items correlated poorly with their scores on cross-case items (r = .15); similarly, their scores on determinate items correlated poorly with their scores on indeterminate items (r = .18). On TV2, however, while we observed the same poor correlation between determinate and indeterminate items (r = .19), there was a stronger correlation between single-case and cross-case items (r = .40). The latter result suggests that TV2 single-case items may have been more closely related to the cross-case items than they were in TV1.

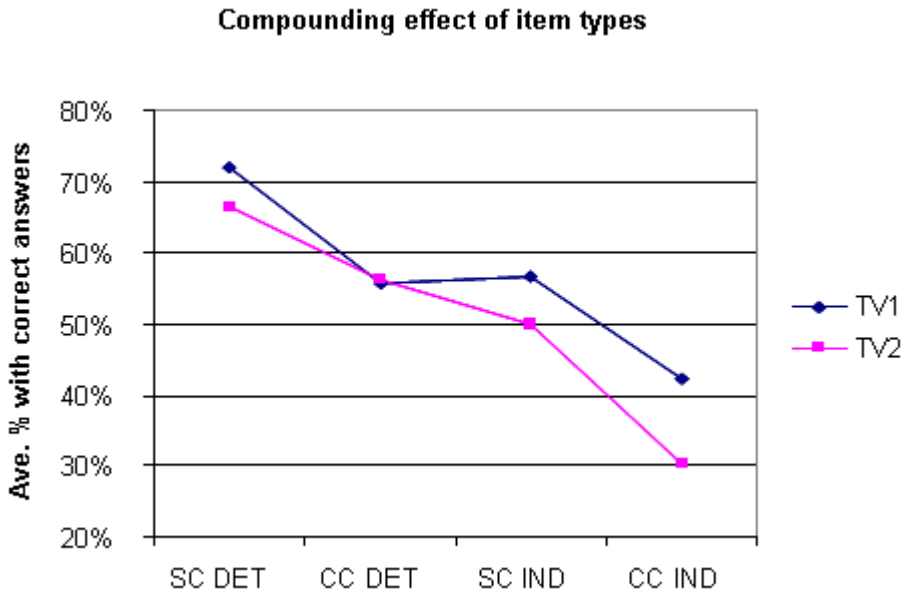


FIGURE 6. *Average difficulty levels for each item type for TV1 and TV2*

Fifth, and finally, given the apparent difference in difficulty level between TV1 and TV2, another psychometric statistic was pursued. Analyses revealed that many TV2 items had higher discrimination levels than paired items from TV1. A discrimination level is determined by correlating the item score with the total score of the test. All items in each version correlated positively with total test scores, but many were very low. This issue will be taken up in the section that follows dealing with the complexity of constructs that inhere in each item. Nonetheless, TV1 item discrimination levels ranged from .01 to .19. For TV2, these levels ranged from .02 to .31. In short, this analysis may indicate that although TV2 appears to be more difficult on average, it discriminates better among students on the bottom or top of the distribution.

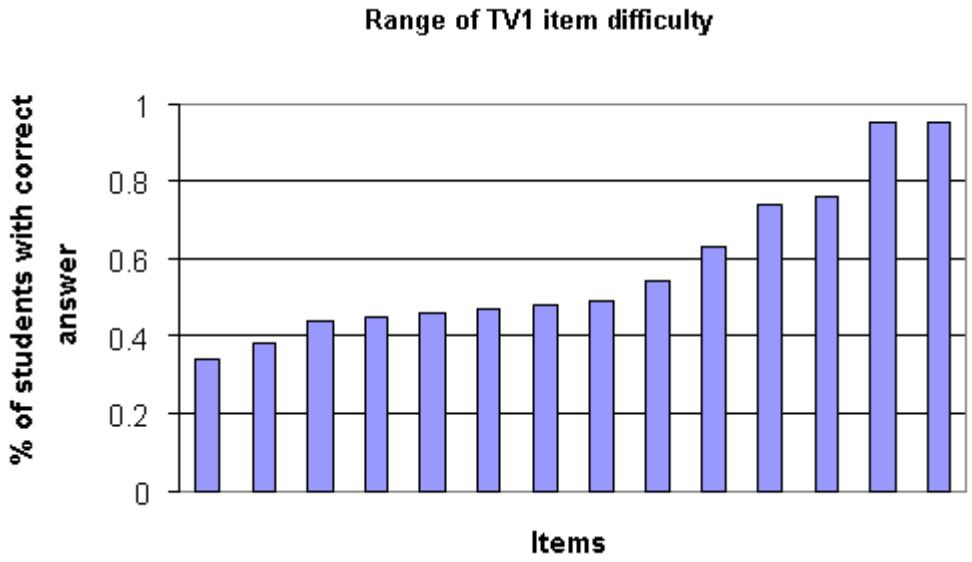


FIGURE 7. *Range of TV1 item difficulty ordered from hardest to easiest*

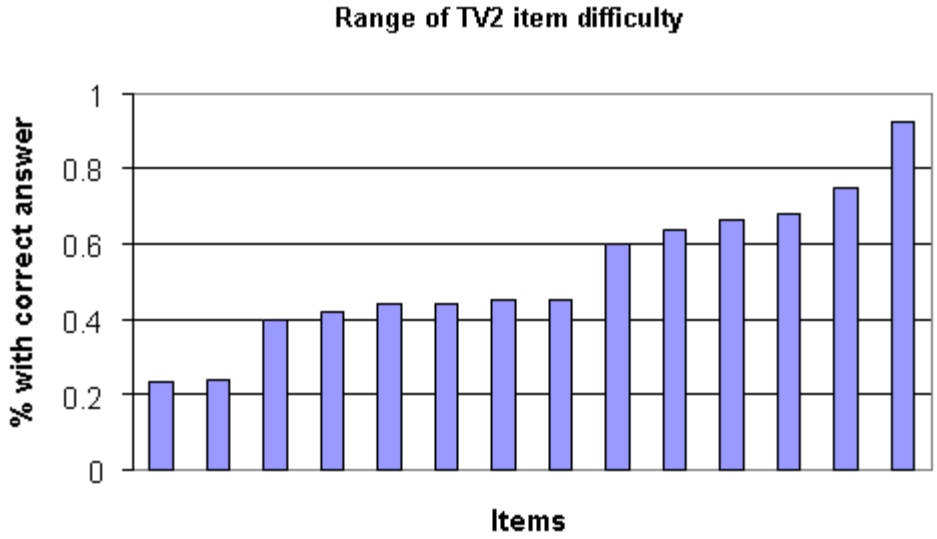


FIGURE 8. *Range of TV2 item difficulty ordered from hardest to easiest*

Validity

Multiple Methods of Establishing Construct Validity

To further examine whether the 2 × 2 question typology could be replicated in TV2, three different methods were used to examine the equivalence of the two test forms: statistical analysis using DIF techniques, expert review, and analysis of test readers' item-response process through the window provided by student think-aloud protocols. These three methods specifically examined whether items that were considered to be of the same item type could be found to be equivalent. Each of these methods is described further below.

Statistical Analysis Using Differential Item Functioning

To determine whether the items from the TV1 form were equivalent to the items from the TV2 form, the 14 items from each test version were paired and analyzed using a variation of conventional DIF techniques. Typically in DIF, the person belonging to a particular group (i.e., reference or focal) is the unit of investigation. In these analyses, the item pair, consisting of theoretically equivalent items from each test form (TV1 and TV2), was considered the unit of investigation. Due to the nature of the dataset, a combination of several established DIF methods as well as other statistical methods that have not yet been previously used for DIF were used in this study to examine the equivalence of paired items on the test.[5] The analyses included the delta plot method, the signed proportion difference index (SPD-X), the Mantel–Haenszel log odds ratio, and a modified version of the McNemar test. Of the 14 item pairs, 4 were identified as potentially problematic using all of the above indices and were thus flagged as being "nonequivalent."

Expert Review

In the fall of 2006, seven law professionals, each a law professor, were asked to serve as "legal experts" in determining whether the two forms of the test were equivalent. First the law professionals were asked to skim over each test form (which included the directions to students, the cases, and the items) to get a general idea of what the task entailed. Next, the professors were asked to complete a set of rating sheets consisting of a series of 25 item pairs. Because of the length of the items, only the stems were presented to them in the rating sheets. The keyed choice and distractors, or multiple-choices, were not included in the rating sheets, although the experts were exposed to these during their initial scanning of the tests and could look back if they wished. Fourteen of the pairs coincided with the theoretically equivalent item pairs from the TV1 and TV2 test forms. The remaining item pairs were randomly put together so as to reduce potential bias in experts' ratings. The experts were asked to rate each in terms of the similarity of each item pair on a scale from 1 to 10 with 1 being "Very Similar" and 10 being "Very Different." Item pairs with an average above 5.5 were flagged as being perceived as dissimilar among the experts and therefore potentially problematic. Items with an average rating below 5.5 were identified as being perceived as equivalent among the experts. This activity determined 3 of the 14 items to be nonequivalent. None of the randomly constituted pairs was assessed as equivalent. Only one of the pairs flagged by DIF was deemed problematic by the experts.

Analysis of Process Through Think-Aloud Protocols

A total of 30 students were asked to think out loud while completing portions of the TV1 and TV2 tests. As mentioned earlier, because the test was quite lengthy, students completed the verbal protocol for only a select number of items in order to reduce fatigue and secure more reliable results. Therefore, the two main test versions were each divided into subforms of eight items each, and each subform contained a similar distribution of item types. Students were randomly assigned to subforms and took only one subform. In all, there were four subforms: two for TV1 and two for TV2; since each full test had 14 items, and each subform sampled 8 items (drawing two from each item type), two items overlapped between subforms.

Following a short practice exercise, students were handed the task directions and the three cases associated with their assigned test form, which they were asked to read silently. Students were given access to a legal dictionary. When they finished reading, the students notified the proctor who then administered the eight-item test. Students were asked to read each item out loud and verbalize what they were thinking as they pursued the best response. In order not to influence or bias responses, very few prompts were used during the think-aloud period. Rather, only basic prompts such as "Keep talking" or "What are you thinking now?" were used if the student did not verbalize his or her thoughts after a 30-second period of silence. Following the verbal protocols, short interviews of the students were conducted in order to collect some retrospective information about the strategies used in answering the multiple-choice questions and to elicit general ideas about case reading and reasoning. The practice think-aloud task, the actual think-aloud task, and the interviews were audio-recorded and later transcribed.

Conducted by two members of the project (Zappe and Evensen), the data analysis occurred in three phases: *open coding*, *discourse analysis*, and *paired comparisons*. Because the item was the unit of analysis, all student transcripts for an individual item were coded together. One individual item was analyzed initially to develop a general coding scheme. Analysis of the protocols was facilitated by the use of the N6 software program (QSR International, 2002). Figure 9 provides a graphic display of the nodes used during the initial open coding stage of analysis.

The coding strategy developed was based on both the analysis of an individual item and a theory of the strategies that students would use when responding to the items. The researchers theorized that students would primarily depend on references to the case text, either by directly referring to the case or depending on their memory of the case. The task would require some basic knowledge of the law, such as definitions of "remand" and the concept of the appellate system.

[5]For a complete report on these methods see Zappe, S. E. (2007). *Response process validation of equivalent test forms: How qualitative data can support the construct validity of multiple test forms*. Unpublished doctoral dissertation. The Pennsylvania State University.

However, this requisite knowledge would be very basic, primarily including concepts that students learn during their first few weeks of their first semester in law school. All test questions were designed so that no student would be able to correctly respond to an item based solely on prior legal or doctrinal knowledge. Instead they were designed to induce students to rely on references to the cases as well as case reasoning skills to respond correctly.

In addition, the researchers were concerned as to whether construct-irrelevant influences would lead students to the correct response. Because multiple-choice tests constitute a genre of discourse, it was expected that students would employ some test-taking strategies to respond to items. However, of particular concern were potential instances of a student correctly responding to an item using *solely* test-based strategies.

Essentially, this open-coding step provided a means to first "cut" the data and sort it in a way that would facilitate the next level of analysis. Responses were first sorted into one of two nodes: correct or incorrect. Next, the responses were sorted into nodes indicating whether the student used text-based, knowledge-based, or test-based strategies to either accept or eliminate the option.

In the second phase of data analysis, the responses were analyzed in accordance with a set of guiding questions. This phase of the analysis depended on the item writers' justifications for each item, which described reasons why the key was correct and why each distractor was considered incorrect. First, the student responses for the keyed (correct) response were analyzed to determine if an ideal student response existed in the data. In other words, we asked, "What evidence supports that students were able to select the key for reasons that the item writers had intended?" Next, the distractors were each individually analyzed to determine if there was evidence to support that students were able to eliminate these for the reasons intended by the item writers. These responses were also matched against written justifications. The final guiding questions concerned only the students who incorrectly responded to the item and investigated the reasons why they eliminated the key and the reasons why they selected their incorrect option. Based on the answers to the guiding questions, a tentative conclusion was drawn as to whether or not each item was measuring the intended construct. The results to these guiding questions were written up in a summary document for each item pair.

In the final stage of the think-aloud data analysis, termed *paired comparisons*, the summary documents compiled during the prior stage were reviewed. The pairs were rated in terms of whether or not an ideal response was evident, whether or not each distractor was eliminated by at least one student for the reasons intended by the item writers, whether or not the types of errors that occurred in each item were generally similar, and whether or not the items were indeed equivalent in terms of not only the foci, but also the types of skills and reasoning strategies employed. The raters first independently selected a dimension for each pair and then discussed the ratings until a consensus was achieved for each pair. The dimension assigned to the pair depended upon two factors:

1. A perceived sense of equivalence between the items in terms of whether they matched the justification and the strategies and reasoning invoked in responding

2. Whether or not revisions were suggested for one or both of the items in the pair.

Item pairs were assigned to one of four dimensions:

1. equivalent with no suggested revisions

2. equivalent with revisions suggested for one item in the pair

3. equivalent with revisions suggested for both items in the pair

4. not equivalent.

As a result of this form of analysis, 5 of the 14 pairs were judged to be nonequivalent.

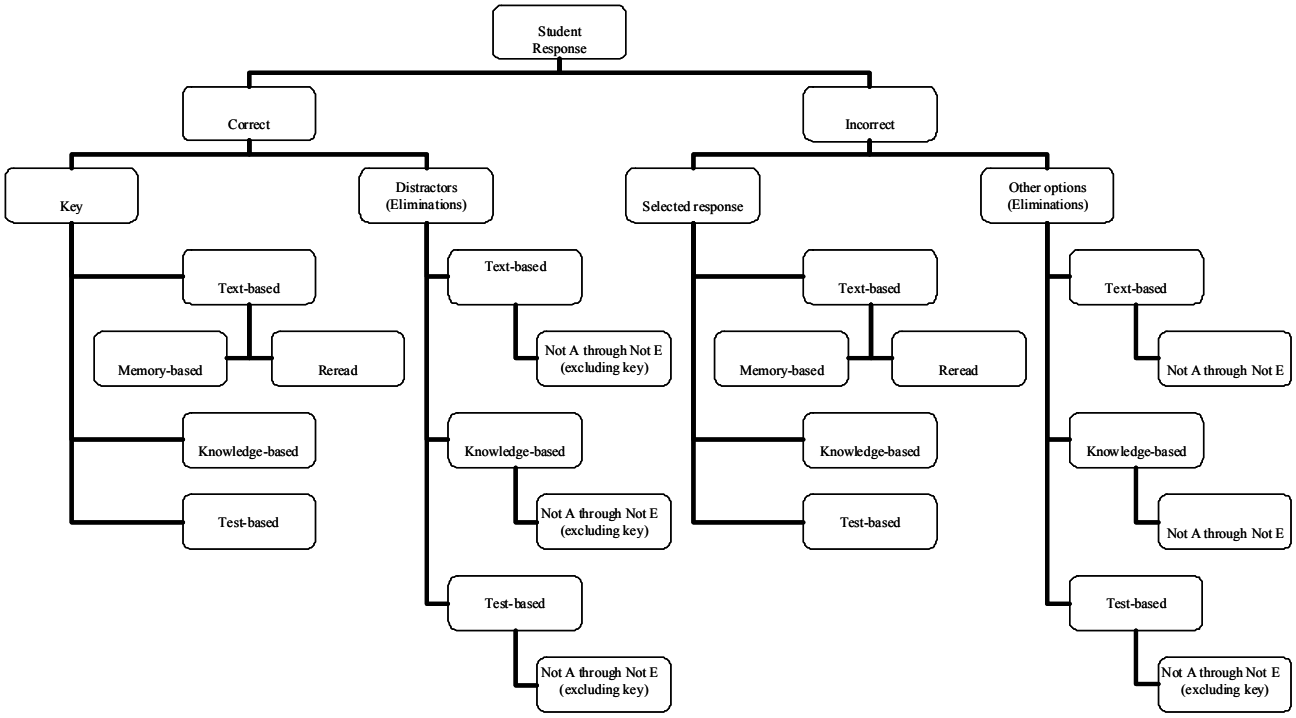


FIGURE 9: *Nodes used in open coding of item-level transcripts*

Summary of Different Methods of Comparison

Following the analysis of all three methods, the items were compared in terms of equivalence/nonequivalence, as displayed in Table 10. The results of the study showed points of disagreement on the equivalency of the items within a pair. We were then left with the decision as to how to use this information. Consistent with the assumptions of these different methods, each provided information about necessary changes to certain items to make them more equivalent across the pairs. In particular, the information collected through the think-aloud methodology yielded rich, complex information about necessary revisions to the item pairs. The findings from the think-alouds and how they influenced the test development process are discussed in a later section of this report.

TABLE 10
Summary table for comparison of methods

Pair	Think-Aloud	Expert Judgment	Statistical DIF Methods
1	**Nonequivalent**	Equivalent	**Nonequivalent**
2	Equivalent	Equivalent	Equivalent
3	Equivalent	Equivalent	Equivalent
4	Equivalent	Equivalent	Equivalent
5	Equivalent	Equivalent	Equivalent
6	**Nonequivalent**	Equivalent	Equivalent
7	Equivalent	Equivalent	**Nonequivalent**
8	**Nonequivalent**	Equivalent	**Nonequivalent**
9	Equivalent	Equivalent	Equivalent
10	Equivalent	Equivalent	Equivalent
11	**Nonequivalent**	**Nonequivalent**	Equivalent
12	Equivalent	Equivalent	Equivalent
13	Equivalent	**Nonequivalent**	Equivalent
14	**Nonequivalent**	**Nonequivalent**	**Nonequivalent**

Comparative Analyses

The Correlation Questions

Just as we did in our Phase 1 study, we also assessed the correlations between TV1 and TV2 test scores and other conventional measures of achievement, specifically LSAT and LGPA. These overall correlations are shown in Table 11, below (Pearson's *r*). As this table shows, the LSAT correlations with test scores are nearly identical. Similarly, cumulative LGPAs are also very similar, across test versions and across Phase 1 and 2 field tests. One explanation for the low LSAT correlations is that the range is restricted by the participating law schools' admission criteria; the GPA range at these schools is also restricted. However, these low correlations do support the notion that the skills being tapped in our instruments are somewhat distinct from those captured by LSAT and LGPA.

TABLE 11
Pearson correlation of total test scores with LSAT and cumulative LS-GPA

	LSAT	LGPA
Phase 1 study, TV1	.22	.22 (1L)
Phase 2 study, TV1	.20	.18 (1L + 3L) (1L = .17; 3L = .20)
Phase 2 study, TV2	.21	.15 (1L = .19; 3L = .11)

Importantly, we think the fact that these LSAT correlations are consistent and positive across both versions of the test provides some confidence that the tests do relate to other skills involved in LSAT performance. That this confidence is warranted can be seen when we segment students' test performance into three groups—high, medium, and low—as shown in Figure 10.[6] This figure shows that students with higher LSAT scores definitely perform better than students with low LSAT scores, and we suspect that the common skill tapped by the two instruments is manifested in the single-case/determinate kinds of test items found in both instruments.

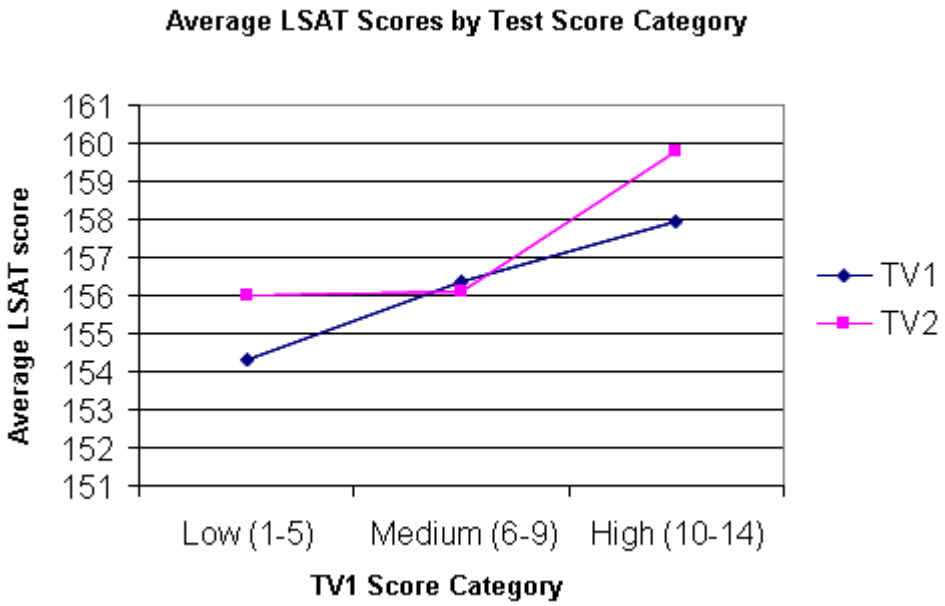


FIGURE 10. *Relationship between LSAT scores and segmented scores on TV1 and TV2*

Similarly, we think the fact that LGPAs are consistent and positive across both versions of the test provides some confidence that the tests do relate to skills involved in law school course examinations, although we would also note that the relationship between 1L GPAs and test scores versus 3L GPAs and test scores differs. As above, these differences can be seen when we segment students' test performance into three groups: high, medium, and low. As shown in Figure 11, 1L students with high GPAs definitely perform better on both TV1 and TV2 than students with low GPAs. As with

[6]Using Tukey's honestly significant differences (HSD) test, the mean LSAT score difference between the three groups is significant at $p < .05$.

the LSAT, we suspect that the common skill tapped by our tests and 1L course-based examinations is manifested in single-case/determinate test items. The same relationship between high test score and high GPA is seen in the data for the 3L students. However, unlike the scores of the 1L students and as Figure 12 shows, low and medium test scores of 3L students do not show a consistent relationship with their GPA. While a number of explanations for the difference in the 1L and 3L GPA curves are possible, we suspect that the kinds of assessments that occur later in law school may focus on a wider range of skill sets than are reflected in our tests.

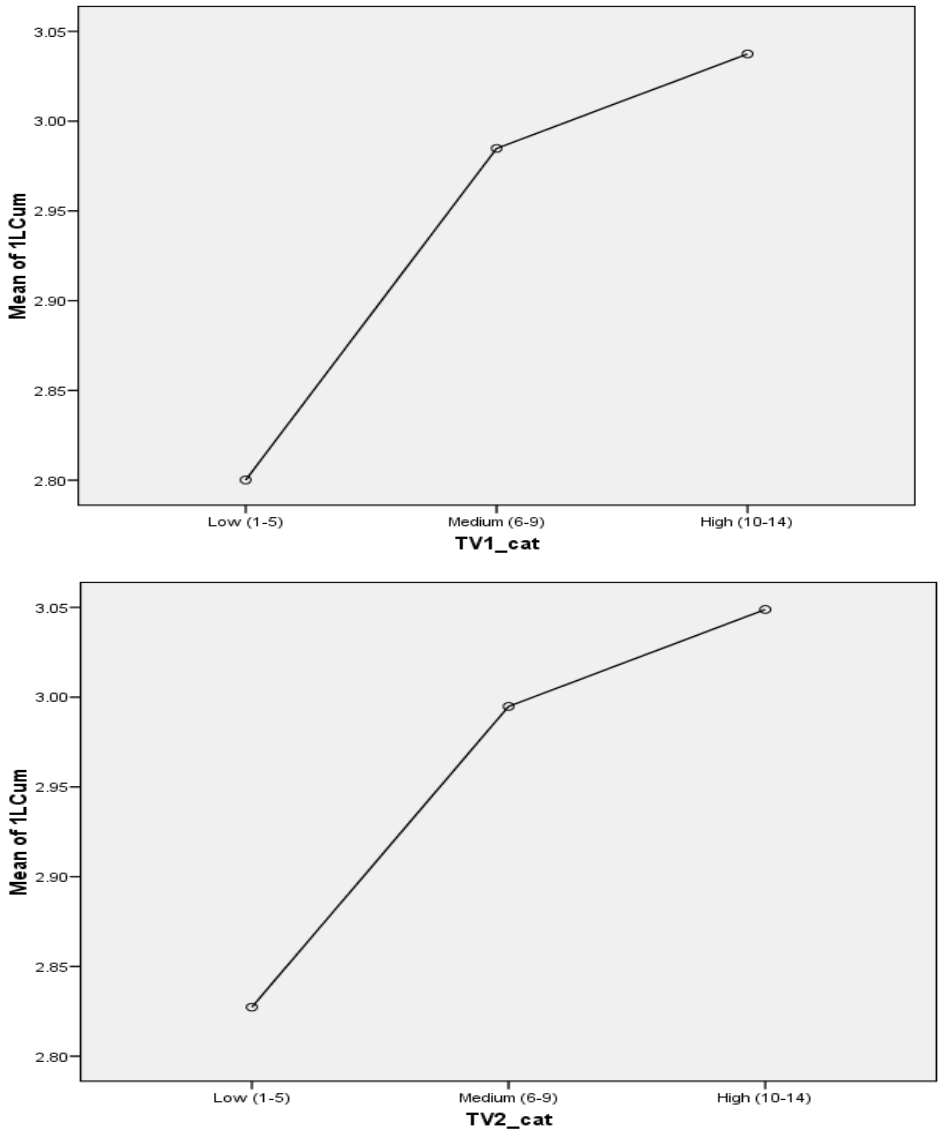


FIGURE 11. *Relationship of mean 1L GPA to TV1 and TV2 test scores*

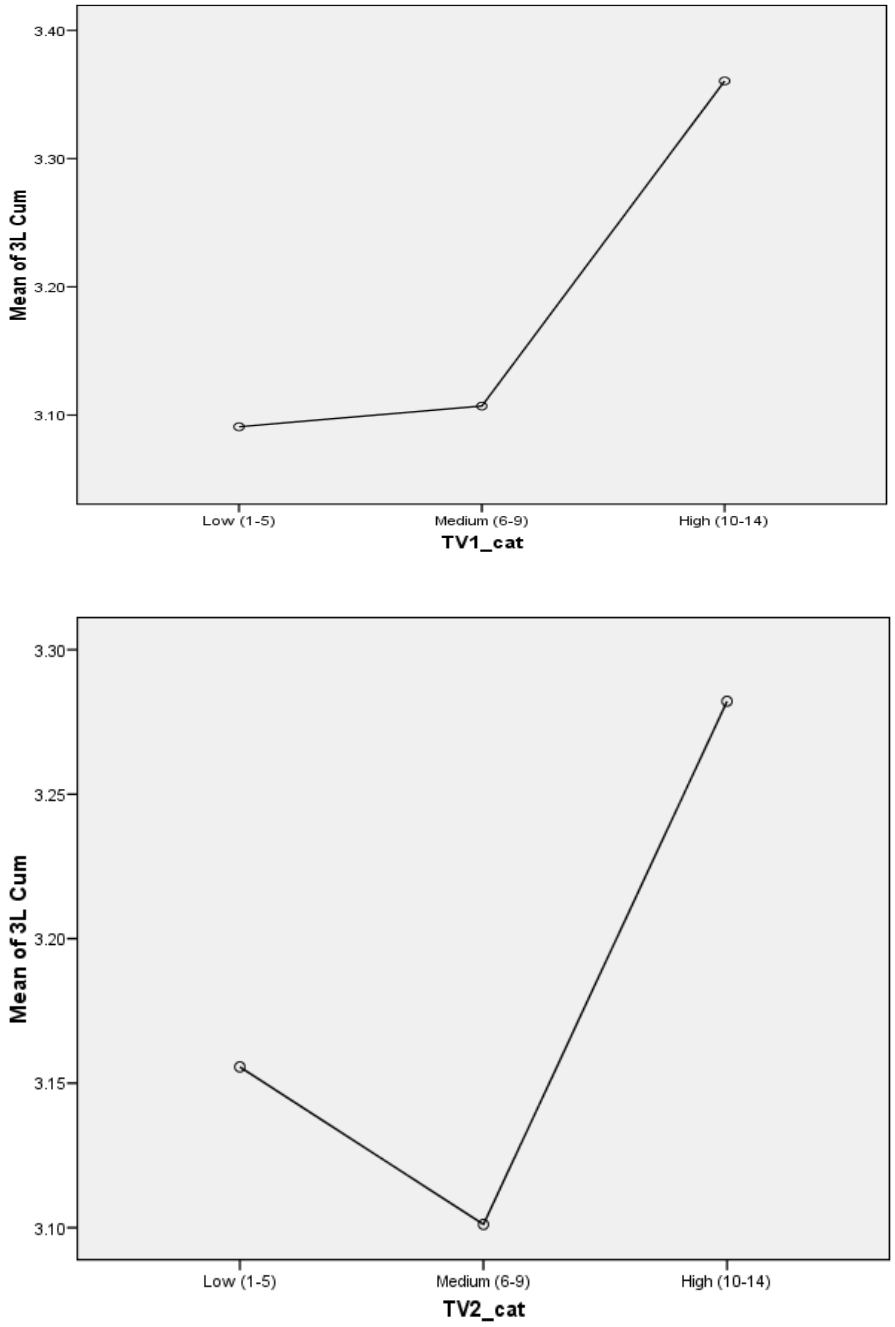


FIGURE 12. *Relationship of mean 3L GPA to TV1 & TV2 test scores*

Summary of Field Test Results

The purpose of the foregoing section was to address the following three questions:

1. Do students' case reading and reasoning skills improve between their first and second years in law school?
2. Do students' case reading and reasoning skills improve between their first and third years in law school?
3. Can the underlying 2 × 2 question typology developed for TV1 be replicated in TV2, thus advancing the construct validity of this typology?

We would like to summarize here our findings as they relate to each of these questions.

1. Overall, the results reported above concerning the question of law student improvement between second and third semesters suggest that students' case reading and reasoning skills do not improve during this period. Because we used two forms of our test in a within-subject, fully counterbalanced and randomized design, we have a stronger basis for drawing conclusions about students' improvement than we had in our Phase 1 study, where we used only one test form and a between-subject, matched LSAT design. Indeed, our Phase 2 study also uses a better design for investigating improvement than Bryden's (1984) study, in which he also used only a single test form in a between-subjects design (see Stratman, 1990, for a full discussion of the strengths and weaknesses of Bryden's design and results). Our quantitative analyses did consistently indicate that TV2 is a more difficult test than TV1. However, even after statistically adjusting TV2 scores to compensate for this greater difficulty, students' scores between the second and third semester did not significantly improve.

2. The results reported above concerning the question of law student improvement between first and third years also suggest that students' case reading and reasoning skills do not improve during this period. However, the design of this particular study, while better than the design of our Phase 1 study, is not as strong as the design used to investigate student improvement between the second and third semesters. While our study of first to third year improvement used a within-subject design with two test forms, unlike our study of second to third semester improvement it was not counterbalanced for test version and test order. We cannot therefore be as confident about our finding of no improvement, even though we did again statistically adjust TV2 scores for difficulty.

3. Overall, the results above indicate that we were able to replicate in TV2 the underlying 2 × 2 question typology developed for TV1, and thus we were able to advance the construct validity of this typology. Having made this assertion, however, it is important to recognize that replicating the typology does not mean the test forms are of equal difficulty; as noted, our data clearly indicate that they are not equally difficult. Both tests might equally reflect the constructs associated with the typology without necessarily being tests on which students can achieve the same scores. The primary indicator that we were able to successfully replicate the typology is the similarity in the pattern of scores across the four item types. As in our Phase 1 study, the results in our Phase 2 study show the same decrease in performance as test items shift from single-case to cross-case items, and from determinate to indeterminate items. The replication of typology also is supported by the similar score distributions obtained with each form, as well as by the replication of correlation patterns with students' LSAT scores and LGPA. At the same time, we do not claim that TV2, in the version we used in our field test, perfectly replicates the typology as it was applied in TV1 and that it needs no adjustment. As we discuss below, the field test, especially the think-aloud protocol results, has proved valuable in indicating ways to modify both tests so that they are both more similar to one another in terms of difficulty and more closely aligned in terms of the precise skill constructs they assess.

Further Analysis of Think-Aloud Data

Identifying Constructs, Skills, and Subskills

In designing the two forms of this test of case reading and reasoning, item writers and expert reviewers of the draft versions essentially were hypothesizing about what skill-based constructs become operational when readers encounter a legal case for a specific purpose. When such skills are embedded into a test genre, in particular a multiple-choice test, the task on the reader's part involves an activation of the cognitive processes and a determination as to whether and how such processes are appropriate. Test writers, in a sense, position themselves to anticipate and stimulate this activation and evaluation activity.

While standard test results offer indirect ways of confirming or disconfirming test makers' hypotheses, think-aloud data allow us to investigate hypotheses more directly. As illustrated in the previous section, the information gleaned from the think-alouds could be reduced in a way that spoke to the validity of alternate forms of the test and in ways missed by more conventional measures (i.e., statistical analyses and expert review). At a more essential level, the 30 think-alouds provided windows into the cognitive processes of the test takers and allowed us to witness effective reading and reasoning. Conversely, they also allowed us to find concrete examples of what might be going wrong when students misread or engage in faulty reasoning. This close, or microanalysis of the think-aloud protocols enabled us to further specify what originally had been conceptualized as general constructs related to the reading/reasoning task. In other words, the outcome of the analysis is a specification of what might be called overarching skills and subskills associated with case-reading activity. Finally, the think-aloud data provided us with evidence that the fault, when detected, might reside in the items rather than the test takers, and such information led us to consider revisions for the tests. In the following sections, we provide illustrations of each of these affordances of the think-aloud data.

Verifying hypotheses and specifying constructs. Our test was constructed along two broad dimensions: (a) single- and cross-case items; and (b) determinate and indeterminate items. Within those dimensions we composed items that reflected discourse-specific aspects relevant to the comprehension and use of legal cases. For example, we tested whether or not students were able to identify the *legal issue* taken up in a single case. In published legal cases, canonical parts are not typically labeled as such, and often when looking for the issue addressed in a case, a reader must contend with linguistic constructions used within a case that might be mistaken for the issue. In light of this rhetorical characteristic of legal discourse, students must learn specific, often subtle, strategies to discern appropriate articulations of canonical parts of cases. The think-aloud data provided evidence concerning how students went about doing this.

Item 4 in TV1, which is presented below, focused on the *issues* of single cases.

ITEM 4: SINGLE-CASE, DETERMINATE

Which one of the following questions best expresses the *legal issue* that the *Mackey* court sees itself as addressing?

a) Is it within the court's discretion to decide when the difference between stated requirements for making an appeal from arbitration and a defendant's partial compliance with those requirements can be disregarded a mere technicality?

b) Under Pennsylvania rules, what reasonable conditions may a court impose upon the right of appeal from arbitration?

c) ***Should a defendant's nonpayment of record costs, coupled with an attempt to pay them two days late, prevent a defendant from making an appeal from an adverse arbitration decision under Pennsylvania rules?

d) Under Pennsylvania rules, may defendants who find themselves in the following factual situation be allowed to appeal an adverse arbitration decision:

   – Received an unclear communication from the plaintiff concerning the amount owed and the due date for paying record costs; and
   – Tried but failed to reach the Prothonotary's office to learn the exact amount of record costs owed; and
   – Paid $237 of the $327 due for record costs two days late; and
   – Met all other procedural and monetary requirements for appeal?

e) Should a defendant's partial payment of record costs be considered more or less important than payment of other costs, such as fees for the arbitrators and security?

Choice C, the keyed response, very closely paraphrases language used by the trial court, and one student recognized this: "It's almost the exact restatement of the issue that the court presents." The skill operating in this instance might be described as **recognizing an appropriate paraphrase of a given statement of an issue**. In spite of the ease with which this one student deployed this skill, the same could not be said of a large number of other test-takers. Indeed, this proved to be a difficult item, with only 54% of the silent reading sample[7] getting it correct and only three out of seven think-aloud respondents getting it correct. Other options made available through the distractors, or incorrect choices, for this item needed to be considered, and for some students, these proved attractive. Still, some readers eliminated the distractors for reasons anticipated by the item writers. For example, one student eliminated Choice E because "a balancing test was never undertaken by the court." We could say that a subskill of discerning the issue would be **recognizing when a particular issue is absent from a case**. Another student eliminated Choice A because "Obviously they think it is within their discretion to decide because they <u>do</u>[8] decide." This subskill might be described as **recognizing that a question about subject matter jurisdiction is settled, and therefore not an issue**. A standard reason that a statement of an issue may not be appropriate (not unlike its counterpart of discerning the "main idea" in nonlegal texts) is that the statement may be too broad or too narrow. Students who eliminated Choice B recognized these types of distractors, commenting that "*Mackey's* definitely not going as far as B would go" (in other words, finding the statement overly broad). Students who eliminated Choice D also recognized these distractors, saying, "It's so factually specific … you're not going to be able to abstract any sort of legal concept from that. So you have to make it a little more general." This last comment presented a skill not anticipated by the test writers, and one that is very much tied to the kind of knowledge-based reasoning that is activated when cases are read. This might be articulated as **understanding how issues are tied to legal rules**. Another student who correctly responded to this item demonstrated a combination of

[7]Because statistical analyses found no significant differences among any groups involved in either phases of the study, data were aggregated in these analyses. Hence the *N*s reported in this section are 283 for TV1 and 172 for TV2.

[8]Underlining is used to represent the stress put on this word or phrase by a test taker.

recognizing the paraphrase and being sensitive to the need to avoid too general or too specific a construction. In deciding on Choice C, she said, "That's the basic legal issue with a little bit of the facts put in there."

What the above analysis demonstrates is that the skill of articulating the issue in a case is not as straightforward as it may seem, and we think that our research has not only specified some theory-based constructs, but has provided further evidence that competent legal readers operate within the parameters of such constructs while reading. Yet, our data also revealed that other readers were confounded by the construct-relevant requirements of an item and fell into comprehension pitfalls of actually selecting the **too general, too specific**, or **settled** statement of the issue. Furthermore, and most likely peculiar to this particular case, some students failed on this item due to reading problems of a more general nature. Two of the think-aloud test takers rejected the correct response on this item because they failed to appreciate the legal versus the everyday meaning of the terms "nonpayment" and "attempt to pay" despite the fact that these were terms explicitly used by the court in the body of the opinion. One student dismissed the choice by saying, "It isn't nonpayment." The other erroneously reasoned, "And it wasn't an attempt to pay them two days late—It was successfully paying them two days late." Of course, if the court considered the late payment "successful" it would not have found against the defendant, Mackey.

Matching constructs and understanding differences between test versions. When we embarked on TV2, we were sensitive to the item types we were constructing, focusing particularly on the hypothesized constructs inherent in the choices. We were also attempting to keep items as parallel as possible. Still, we recognized that we were constrained in this goal by the authenticity of the texts and tasks employed in our venture. A close look at the *issues* item from TV2 (Item 3), below, which was paired with TV1 (Item 4), above, illustrates this challenge.

ITEM 3: SINGLE-CASE, DETERMINATE

Which one of the following statements best expresses the issue(s) which the District Court of New Jersey, on remand, will have to decide in view of the *Alexander* court's analysis and decision?

a) The district court must decide whether or not American Can's previous changes to its welfare benefit plan were consistent with ERISA's goal of preserving the solvency of its plan and hence preserving the continued availability of benefits already promised to employees.

b) The district court must decide whether the benefit plan documents provided by American Can unambiguously reserve the right to alter participants' benefits or whether these documents unambiguously promise lifetime benefits that cannot be altered.

c) The district court must decide whether American Can's contested disclaimer phrase, "in conformity with applicable legislation," satisfies ERISA's requirement that summary plan descriptions adequately warn employees of adversity and conditions under which their benefits may be cancelled or denied.

d) The district court must decide whether, given the arguments of both the plaintiffs and the defendants, the benefit plan documents provided by American Can to retired employees have only one or multiple meanings.

e) ***The district court must decide whether American Can's prior behavior in raising the cost of plaintiffs' medical insurance premiums supports the company's claim that it unambiguously reserved the right to reduce plan benefits and never promised lifetime, fixed-cost benefits in its benefit plan documents.

From the outset, and simply by comparing the two stems, one can see that these items differ in a fundamental way: the TV1 item asks the reader to identify the issue of a given case; the TV2 item requires the reader to discern an issue that is framed by an appellate court for a trial court to consider on remand. Despite this difference, we observed through the think-aloud data students who correctly responded to the item in ways evidenced in its paired, TV1 item. The strategies that students used to do this, however, differed. One student prefaced her response by returning to the text and locating the section in the *Alexander* case where the court speaks of the remand. She then combined reading with reasoning, commenting, "That's a question of fact for the district court. If it should [student reading] 'reasonably infer that, because benefits had been reduced in the past…" Another student skipped the rereading part and relied on memory to conclude, "A prior record of benefit reductions would show action consistent with the defendant's position." Again, employing a relatively straightforward skill, **recognizing an appropriate paraphrase of a given statement of an issue,** might have produced a low level of difficulty on this item. Yet, similar to those taking TV1 (Item 4), only 42% of the larger sample correctly responded to this item, and only four out of eight think-aloud subjects chose the keyed response. Two of the distractors on this TV2 item reflected skills also necessary for TV1 (Item 4), specifically, **recognizing when an issue has already been settled** (Choices B and D) and **recognizing when a particular issue is absent from a case** (Choice A). Choice C required a skill not assessed through TV1 (Item 4), **recognizing when a statement of an issue is**

only partial. One student in the think-aloud condition recognized this partiality, indicating that the appellate court is saying that the lower court must "look at the whole document" (i.e., not just one phrase) "and must look at the behavior" (reference to behavior is missing from choice A). Also missing from this item in the pair is the more generic skill of recognizing when a statement of the issue is too broad or too narrow; this skill is not tested in this item. The decision not to craft distractors that followed precisely the same pattern was based on the fact that the issue was framed by one court for another, and because reference to the remand appeared twice in the decision and was therefore retrievable provided readers combined those two references. On the other hand, the issue related to the *Mackey* case in TV1 could have been articulated based on its reference in either part of the case.

What the above analysis of the two paired items illustrates is that one test cannot be a clone of another. At least test makers who use actual rather than hypothetical cases strive for their tests to be isomorphic to a degree that is determined by the characteristics of the cases used. Locating those aspects of the cases that require the deployment of certain skills (and underlying constructs) is the primary task for item writers; seeing how they can be replicated in an alternate form is important, but secondary. Table 15 lists the constructs, overarching skills, subskills, and alignments between single-case/determinate and cross-case/determinate items for both versions of the test. The table also illustrates the points of equivalence among items by pairing items or leaving them unpaired.

TABLE 12

Comparisons of TV1 and TV2 determinate item types (both single- and cross-case) based on common constructs, overarching skills, and subskills

Construct	Overarching Skills and *Subskills*	TV1	TV2	Pair
Legal issue	Identifying the ISSUE when it is articulated by the court in a single case	4.C*	3.E*	3
	Recognizing when a statement of the issue is too broad	4.B		
	Recognizing when a statement of the issue is too narrow	4.D		
	Understanding that an issue statement must include some relevant facts	4.C		
	Recognizing when a statement of the issue reflects information absent from the case	4.E	3.A 3.B,C	
	Recognizing when a statement of the issue is only partial	4.A	3.D	
	Recognizing that a given issue is settled		3.A,E	
	Understanding the distinctive function of levels of courts			
Legal reasoning	Identifying the REASONING relative to a single case	1.D* 3.C*	1.D* 4.C*	1 2
	Recognizing that a statement of reasoning must be tied to the holding	1.D*		
	Recognizing that a statement of reasoning must be tied to the issue	1.A,B	4.B,E	
	Identifying reasoning related to policy		1.C,4.B,E	
	Recognizing when a statement of the reasoning is too broad	3.E		
	Recognizing when a statement of the reasoning is too narrow	3.A,B	4.D	
	Recognizing when a statement of the reasoning is inaccurate	1.C	1.A,B,E	
	Recognizing when only part of the reasoning is given	1.E,3.D	4.A	
Legal issue	Identifying the ISSUE common to multiple courts by synthesizing across cases	6.C*	6.A*	4
	Recognizing that a synthesis statement leaves out a crucial piece of information	6.D	6.C	
	Recognizing that a synthesis statement is too narrowly pitched	6.A,B,E	6.B,D	
	Recognizing when a synthesis statement reflects none of the courts involved		6.E	
Fact-to-fact comparison	Recognizing the similarities and/or differences between FACT patterns of two cases	5.E*	7.E*	5
	Recognizing similar fact pattern useful to case	5.A,C	7.E	
	Recognizing that given facts do not pertain to both cases	5.B	7.B,D	
	Recognizing that factual comparisons must take place at an appropriate level of specificity		7.A	
	Recognizing when a fact, although true, is not relevant	5.D	7.C	
Legal reasoning	Understanding similarities and/or differences in the REASONING used by two courts	7.C*	5.B*	6
	Recognizing when a statement of reasoning contradicts that presented in a given case		5.A	
	Recognizing that the reasoning must be tied to the issues and holding of a case		5.C,D	
	Recognizing subarguments within the reasoning of a case		5.E	
	Recognizing when the reasoning statement is incomplete	7.B,D		
	Differentiating between reasoning and dicta	7.A		
	Differentiating between reasoning tied to policy	7.E		

* Indicates keyed response.

Specifying indeterminacies. Perhaps the main challenge of this overall project was whether or not we could devise items that captured what scholars refer to as the indeterminacies of legal texts. On a theoretical basis we recognized the relationship between a legal reader's ability to both recognize and pragmatically use these features of legal discourse. We are also aware of and impressed by scholarship undertaken in cognitive psychology that has empirically tested the advantages of questioning strategies when readers are faced with the sophisticated and complex discourse of specific disciplines (see references to Graesser et al. [1983; 2003] in an earlier section of this report). Thus, we constructed items on our indeterminate dimension wherein each *answer* was a *question*. Consistent with the assumptions introduced at the start of this report, the role of the reader was theorized as one in which he or she would consider the question posed in terms of its *relevance* to the purpose stimulated by the item's stem and its *utility* relative to other putatively relevant (although incorrect) questions posed. TV1 (Item 10), which follows, demonstrates both the success and the shortcoming of our efforts.

TV1—ITEM 10
SINGLE-CASE, INDETERMINATE

All opinions are interpretations, and interpretations invite questioning. Assuming that you decide to appeal the *Mackey* decision to Superior Court, which one of the following questions presented by the *Mackey* opinion is most relevant to this task?

a) Why does the court state that the issue in the case is whether an appeal may be quashed for "nonpayment" when the facts show that Mr. Mackey partially paid the amount owed, albeit two days late?

b) Why does the *Mackey* court say that Mackey's shortfall in payment constitutes "more than a technicality"?

c) Why does the court rely on its own "inspection" of the pencil note rather than calling for or inviting an expert witness to offer testimony concerning its legibility?

d) Why doesn't the *Mackey* court reprimand the plaintiff for communicating the wrong due date for the record costs?

e) ***Why does the court note each of Mackey's claims about his attempt to contact the Prothonotary's office yet offer no reasons for concluding that Mr. Mackey did not make either an "honest effort" or "a valid attempt to make … full payment"?

Two of the think-aloud subjects reasoned through this item to land on the keyed response. One stated upon reaching Choice E, "This would be very relevant because he did try and it was a substantial effort to find out what the full payment was as well as the time, and he could not get through. And as we saw in *Black and Brown*, the court sees substantial effort as being good, so the court should not invalidate the appeal." Another student was more succinct in concluding, "I think if I was taking this appeal to the Superior Court, I think this would be the most important issue because it seems like they accept the facts, but they don't give them much weight." Likewise incorrect choices were discarded for anticipated reasons. In rejecting Choice A, one student recognized that "nonpayment is just a broader term … it's just considered nonpayment in the eyes of the law because it's an all or nothing standard." This student (unlike the student presented in the previous section whose commonplace usage of the legal term detracted from his reading ability) was able to **distinguish between the legal and lay usage of a term**. Another subskill was evidenced in one student's dismissal of Choice C. He concluded, "I don't think they would be concerned with that at the Superior Court level." In this case the student's strategy was tied with his knowledge of the courts and provided a defensible rationale, although it did not match that provided in the justification for the choice, which stressed the erroneous proposition implied in the choice that somehow the court and not the defendant was to blame. So in reviewing the data for this item, we found it to be functioning appropriately. Yet, further scrutiny seemed necessary to discern exactly how correct responders were getting at the best choice.

As mentioned above, we realized that especially for the indeterminate items, readers would be required to work their way through each of the choices. The difficulty level of .74 found through the statistical analysis of the larger sample of test takers seemed to point to this as an item of moderate difficulty, but the discrimination level of 0.001 flagged the item as problematic. In short, getting this item right or wrong had almost no correlation with total test scores. In looking at the think-aloud data, we found that four out of seven readers responded correctly to the item but that no one eliminated Choices B or D for the reasons given in the justification. In addition, closer inspection of the choices revealed that they were not syntactically balanced in the same way as the other choices and led to the conclusion that test takers may have eliminated this choice for test-based, rather than text-based reasons.

More disconcerting was the response from a student who tossed out the keyed Choice E for a very good reason. She said, "The court <u>did</u> offer reasons for concluding that Mr. Mackey did not make an honest effort. He relied upon the

miscalculation of the plaintiff as to the date when record costs were due." One might defend this item by pointing out that the overall skill being scrutinized is one's ability to **evaluate the reasoning of a given case in light of facts presented**, and that this reader failed to consider the part of the choice that asked her to focus on the relation between Mackey's attempt to contact the appropriate office and an interpretation of that attempt. However, the term "no reasons" in the choice makes the probability of rejecting this choice unfairly high. For this reason the item was revised as follows:

TV1—REVISED ITEM 10
SINGLE-CASE, INDETERMINATE

All opinions are interpretations, and interpretations invite questioning. Assuming that you decide to appeal the *Mackey* decision to Superior Court, which one of the following questions presented by the *Mackey* opinion is most relevant to this task?

a) Why does the court state that the issue in the case is whether an appeal may be quashed for "nonpayment" when the facts show that Mr. Mackey partially paid the amount owed, albeit two days late?

b) How can the *Mackey* court opine that Mackey's shortfall in payment constitutes "more than a technicality" while accepting that the shortfall amounted to only a small fraction of the arbitrator's award of damages?

c) Why does the court rely on its own "inspection" of the pencil note rather than calling for or inviting an expert witness to offer testimony concerning its legibility?

d) Why does the *Mackey* court go into great detail about the confusion surrounding the communication between the plaintiff and the defendant, but only holds Mr. Mackey culpable for not acting within the parameters set out by the statute?

e) ***Why does the court note each of Mackey's claims about his attempt to contact the Prothonotary's office yet not link these to its reasons for concluding that Mr. Mackey made neither an "honest effort" nor "a valid attempt to make … full payment"?

One more aspect of this item warrants attention: Its pair in TV2 (Item 11), although deemed "equivalent" through the statistical analysis, was flagged as "nonequivalent" by both the experts and the think-aloud analysts. The stem for TV2 (Item 11) reads:

Although you did not represent Bridell at the District Court level, you have agreed to take Bridell's case to the Fifth Circuit. In your research you find that Section 510 of ERISA states: "It shall be unlawful for any person to discharge, fine, suspend, expel, discipline, or discriminate against a participant or beneficiary for exercising any right to which [s]he is entitled under the provisions of an employee benefit plan …" 29 U.S.C. § 1140. In light of this provision, which one of the following questions, if found in the affirmative, would be most likely to thwart Ribier's position in a new case?

Even at face value, this item seems nonequivalent to TV1 (Item10) and, indeed, that is probably why the expert reviewers saw it as such. During the validity analysis we recognized that the two items had been paired solely on the basis of the shared larger construct of *indeterminacy* and the fact that it was otherwise unpaired.

This recognition raised a few issues. First, it pointed out the need to combine methods (statistical, expert reviewer, process-based) to establish the equivalence of items among alternate forms of tests. Second, it indicated to us that indeterminate items may be less susceptible to pairing because they are so closely tied to the somewhat idiosyncratic rhetorical moves that contribute to the recognized interpretative uncertainty that inheres in legal opinions. Third, and this may seem somewhat circular, it likewise pointed to the need for multiple ways of attesting to the validity and compatibility (if not equivalence) of alternate forms of a test. In other words, it may not be possible to produce perfectly isomorphic items in terms of the equivalence of constructs within all choices (as illustrated in the **recognize the issue** example above). Further, it also may not be possible to produce items that are equivalent except at the broadest level of construct (e.g., as *indeterminacies*). With reference to the cases used in TV2, it did not seem possible to produce an item built around the construct of **evaluating the reasoning of a given case in light of the facts noted within that case**. As we pointed out in an earlier section where the cases were summarized, there were differences between the sets of cases and individual cases that could not be ignored when writing items. Rather than forcing a specific item type and distorting the meaning of the case, we opted to allow the item to stand alone as does TV1.10) in Table 13. As with Table 12, we list in Table 13 the constructs, overarching skills, subskills, alignments, and pairings between single-case/indeterminate and cross-case/indeterminate items for both versions of the test.

TABLE 13

Comparisons of TV1 and TV2 indeterminate item types (both single- and cross-case) based on common constructs, overarching skills, and subskills

Construct	Overarching Skills and *Subskills*	TV1	TV2	Pair
Legal ambiguity	Recognizing an AMBIGUITY inherent in a case	9.D*	9.B*	7
	Recognizing that a question has been answered in a case, thus presenting no ambiguity	9.A,C,E	9.A,C,D,E	
	Recognizing that a distinction being posed is insignificant	9.B		
	Differentiating between holding and dicta			
Hypothetical fact	Considering the relevance and usefulness of a HYPOTHETICAL FACT to an instant case	11.E*	10.C*,11.A*	8
	Recognizing that an intentionally misleading or illegal act could interfere with a legal process	11.E*	10.C*	
	Using criteria established in a case as way of judging legality of an action		10.D,11.C	
	Weighing the legal significance of a hypothetical fact	11.A,D		
	Recognizing that a fact is not "unknown" in relation to a case	11.B	10.B	
	Weighing whether previous behavior might inform a requirement of a case	11.C	10.E	
	Judging whether a hypothetical fact might prove relevant and, therefore, dispositive		10.A,11.B,D	
	Understanding burden of proof among parties in a case		12.E	
Legal reasoning	Evaluating the REASONING of a case in light of the facts noted within the case in order to advance your own case	10.E*		No pair
	Distinguishing between the legal and lay usage of a term	10.A		
	Recognizing a distortion in reasoning	10.B		
	Understanding function of different courts	10.C		
	Understanding burden of responsibility in a case	10.D		
Analogical reasoning	Evaluating the analogical reasoning of a case in light of the way the opinion drew upon precedent, in order to advance your own case	13.B*	13.A*	9
	Weighing the usefulness of one argument over another	13.A,C		
	Understanding the conventions of citation in cases	13.D		
	Understanding issue of precedent case (basis for analogy)	13.E	13.C	
	Weighing the equivalence of terms		13.B,E	
	Understanding contextual appropriateness of analogy		13.D	
Multiple perspectives	Understanding multiple perspectives in order to anticipate arguments opposing counsel might make	14.B*	12.E*	10
	Understanding the role of controlling case	14.A, C		
	Recognizing an argument that would help your case	14.A		
	Recognizing a theory that would jeopardize opposing counsel's case	14.D	14.B,C,D	
	Understanding the strength of arguments directly related to facts of case	14.E		
	Recognizing a distinction that is not significant		14.A	
Theory of case	Recognizing a question that would lead to a feasible implied theory of the case	12.D*		No pair
	Recognizing a "far out" theory	12.A		
	Recognizing an incorrect inference concerning a possible rule from related cases	12.B		
	Distinguishing a premise from a rule	12.C,E		
Theory of case	Recognizing a strategic move linked to a theory of a case		14.A*	No pair
	Recognizing a potentially unfruitful strategic move based on an inappropriate theory of a case		14.B,C,D,E	

* Indicates keyed response.

In summary, the think-aloud protocols allowed us to probe deeply into the processes students used as they engaged with the purpose set out in the directions for the test. In recent years psychometricians have been stressing the importance of this method for the development of reliable and valid tests. The terms *process validity* or *cognitive validity* (Karabenick, Kempler, & Kelly, 2007; Woolley et al., 2004) are used to indicate this method, which posits *individual items* rather than an entire scale as the unit of analysis. Cognitive validity is "determined according to whether respondents' reports conform to theoretically defined validity criteria" (Karabenick et al., 2007, p. 142). In our tests,

these criteria were contained in the written justifications. The results of our analysis demonstrated congruence among many of the student think-alouds and our justifications. They allowed us to observe how constructs were operationalized and how skills were deployed. If such congruence was missing, it signaled to us that the fault most likely existed with the item itself or with the students' reading and/or reasoning activities. We focus on these areas in the next sections.

Revising the Test[9]

One of the chief goals of this phase of the study was to get information that would allow us to revise the test. Phase 1 results indicated that we could construct a test, but the Phase 2 study was designed to allow us to improve both tests based on empirical data. In some ways test making is more like engineering than science. So, like engineers, we scrutinize the results, indicating weaknesses in our design, and begin the process of redesigning and reevaluating.

The statistics for TV2, Item 12 indicated that only 23% of the students in the silent reading condition chose the keyed response. That cross-case/indeterminate item read as follows:

> ITEM 12
> CROSS-CASE, INDETERMINATE
>
> Assume that your petition to appeal the *Bridell* case has been successful. As you prepare your case you will be thinking about how opposing counsel might look to other cases to effectively develop its case. In light of this task, which one of the following questions regarding opposing counsel's argument strategy would concern you the most?
>
> a) To what extent will opposing counsel try to establish that just as medical benefits could be reduced for retirees as they were in *Alexander,* so can they be reduced for active employees like Bridell?
>
> b) How might opposing counsel compare the disclaimer evidenced in Bridell's plan with the disclaimer found in Hamilton's plan?
>
> c) To what extent might opposing counsel draw upon both *Hamilton* and *Alexander* to synthesize a rule concerning promises that inhere in benefit plan documents?
>
> d) How might opposing counsel rely on the *Alexander* court's hypotheticals about possible federal legislation affecting the provision of private health care plans?
>
> e) ***To what extent will opposing counsel attempt to show that the case by case justification put forward in *Hamilton* rightfully extends to Bridell's case?

Of the two students in the think-aloud condition who selected the keyed response E, one clearly demonstrated how he reasoned through the choice:

> I'm not really seeing how case by case justification has really any application in the *Bridell* case. There was a blanket policy regarding AIDS patients and it wasn't claimed to have been made on a specific case basis. But if it had, that might have been disallowed if it were discriminating vis-à-vis USC code. I guess you could potentially say a case by case being lowering AIDS while not lowering benefits for other groups as case by case, which goes to the discrimination aspect.

The justification for this keyed choice read:

> CORRECT. This move, on the part of opposing counsel, might serve to preclude an argument about individual discrimination. The language offered by the *Bridell* court left somewhat ambiguous what was meant by the statement that ERISA does not proscribe discrimination. It goes on to specify that this discrimination is in relation to "the provision of employee benefits." This might mean either that companies can discriminate when it comes to certain benefits OR it might mean that it can discriminate among individuals with regard to certain benefits. This court cites *Shaw* to support this claim, but opposing counsel might strengthen an argument by citing to *Hamilton*, where the court referred to the *Hlinka* case where that court concluded that "ERISA permits an employer to reserve the right to make early retirement benefit determinations (a welfare benefit) *on an individual basis*" (emphasis added). You, therefore, should be aware of this possibility.

[9]The revised form of both TV1 and TV2, including justifications for each item, can be found at http://www.personal.psu.edu/dhd2/casereadingtests/index.html

Although the student's reasoning through the item came close to that hypothesized by the test makers, two things became evident. First, the student did not qualify discrimination as permissible under ERISA (in other words, he held to a definition of an unlawful act under United States Code [USC]). Second, he translated "case" into a collective term in order to positively decide on the choice. This form of reasoning seemed overly complex and convoluted and certainly did not align well with the target skills hypothesized for the item. In addition, when looking at the think-aloud protocols for those who eliminated the keyed response, we found that "case by case" signaled its inappropriateness for almost every student. Still, going back to the *Hamilton* case, we thought we could substantiate the basis upon which the choice was supported, while conceding that the item appeared to be unduly difficult. Hence, choice E was revised to read:

e) ***Will opposing counsel attempt to deflect any showing of discrimination against Bridell by citing to the case by case justification put forward in *Hamilton*?

By doing this, we intend to cue the reader to the issue of discrimination and also signal that it would be good to go back to scan the *Hamilton* case for references to "case by case." The reader should also attend to the motive supplied in the stem (i.e., to locate a possibility of "concern") and perhaps realize that Bridell's lawyer might have an argument (i.e., USC) that would thwart opposing counsel's strategy. Of course, this revised item would require additional field-testing to determine whether or not the revision achieves its goal.

Our analysis of the think-aloud data resulted in dropping three items. TV1 (Item 2) asked students to identify an appropriate statement of the reasoning in the *Mackey* case. Although over 90% of students in the silent reading condition and 100% of the students in the think-aloud condition selected the keyed response, the think-alouds revealed that readers, with one exception, were ignoring the second clause in the choice, which the item writers considered vital to understanding the reasoning. The item had also been flagged during the validity investigation by those who were reviewing the think-aloud transcripts. They determined that the second clause served to *evaluate* rather than *summarize* the court's reasoning, and that the underlying construct governing this item was *indeterminate* rather than *determinate.* Additional analysis revealed that the item was redundant with Items 10 and 13. On these bases, the item was dropped.

One item on each test version that aimed at querying the legality of certain actions referenced in the cases was also dropped. For TV1 (Item 8), there simply was not enough reference to the constitutionality issue to construct reasonable distractors or to mark the key as less than obvious. TV2 (Item 8) raised the issue of the legality of disclaimers within welfare benefit plans, and the keyed response justifiably captured a synthesis among cases. As was the case with the TV1 (Item 8) item, however, writing reasonable distractors was problematic. This item, although dropped, might be revisited at a later date.

Tables 14 and 15 summarize the decisions regarding whether to retain, revise, or drop items on both versions of the test. The final versions of TV1 and TV2 will contain 12 items each, 3 in each of the four targeted categories. One additional TV2 single-case/determinate item survived reviews and field-testing, and that item can serve as an extra item to be field-tested.

TABLE 14
Final decisions on items (to retain, revise, or drop) for determinate *items*

	TV1			TV2	
Item	Decision				
1	retain		1	retain	
2	drop	Does not meet SC/DET criteria. EVALUATES rather than articulates reasoning. Reasoning very complex and not easily summarized.	2	revise	Remove redundancy from choices and refer to remand in stem.
3	revise	B (incorrect). No one matched justification. Need to overstate dissent's reasoning more.	3	retain	
4	revise	Present as propositions not questions.	4	revise	Choices are overly wordy. Delete redundancies.
5	revise	A: Delete "identical"—flags as incorrect. B: Delete second sentence—does not present a fact inherent in the case, but the disposition of the case. E (key): Unnecessarily refers to dollar amount leading reader to inappropriately focus on that.	5	retain	
6	revise	C (key): Change "in what ways" to "under what conditions"—readers attributing motive to former.	6	revise	Use varied synonyms for "situations"—"circumstances."
7	revise	Stem misleading. Change to differences only. C (key): Final sentence is not supportable—delete. Too much overlap among choices—eliminated on test-based reasons.	7	revise	C (incorrect): Simplify.
8	drop	Constitutionality issue not significant in any cases.	8	drop	A (key): Two choices offered are not rhetorically parallel. D (incorrect): Internally contradictory.

TABLE 15
Final decisions on items (to retain, revise, or drop) for indeterminate *items*

	TV1			TV2	
Item	Decision				
9	revise	Choices too lengthy. Fatigue during think-aloud. A-C-E test-based elimination.	9	retain	
10	revise	B & D (incorrect): Do not provide secondary clause proposing alternative action. No one matched justification for B. E (key) incorrect: Court DID offer reasons—delete "no reasons."	10	retain	
11	revise	Stem must allow for possibility that some choices are not unknowns. C (incorrect): Could be reasoned correctly. Make "wronger."	11	revise	C (incorrect): Defensible as is. Include specific dates to make incorrect.
12	revise	Delete reference to cases in stem. Add reference to case in E to make parallel with other choices.	12	revise	E (key): Cue discrimination claim.
13	revise	In stem, make clear that the relevance refers to readers' goal of challenging opinion.	13	retain	
14	revise	Choices differ in terms of level of abstraction. Add specificity to A and B (key).	14	retain	

When Reading and Reasoning Break Down

Finally, the think-aloud data provided concrete evidence of the inadequate reading and faulty reasoning strategies employed by some law students. The most frequent reason that students missed an item was their over-reliance on memory and a failure to return to the available texts to reread. For example, in TV2 (Item10), the keyed response clearly signaled that Ribier employees were having difficulties understanding their SPD. Yet, only one think-aloud student returned to the *Bridell* case to verify what was said about ERISA requirements, specifically that an "SPD must be written in a manner calculated to be understood by the average plan participant." Another reading problem emerged, one that is often discussed by academic support personnel in law schools, who refer to it as a failure to "read closely." This problem was illustrated by one student who, in considering TV1 (Item 13), returned to the *Black and Brown* case to find the language that filled a reference where ellipses were used in a possible response. When finding that portion of text and rejecting the keyed response he said, "So they left out the part about the timely payment. So I don't think that's relevant either." What this reader failed to do was to contextualize the excerpt by reading one or two sentences around it. If he had, he would have seen that also left out was the idea of "substantial compliance," which was key to getting the correct answer. A lack of careful reading was evident among many of the students thinking out loud about the wrong case or making assertions such as "they [the court] already settled the prior behavior thing" (when they did not) or "only the one case dealt with welfare benefits" (when, in TV2 they all did). Interestingly, one student caught herself about to gloss by and reject a choice when she stopped herself and said, "If that is true, my reading comprehension is really disappointing in that that would be a great thing for my case." She then went back to the text, found out that the statement was indeed true, and got the item correct.

Another frequent error, and this was reflected in the more extensive think-alouds collected during the 2003–2004 field-testing, was the practice of selecting an option based on its "truth" rather than its being the "best" choice given the task requirements of the stem. This was exemplified by a student who eliminated the keyed response for TV2 (Item 6) because, as she said, "*Bridell* was not really focused on that [ambiguity], but whether they have the right to change anything in the plan." The stem had asked the reader to think more broadly by including what they had learned from the two other cases in which the option to change the plan had to be undertaken in a certain way (*Hamilton*) and had to be couched in unambiguous language (*Alexander*).

In an earlier discussion, we spoke about the errors committed by students who drew upon common rather than legal usages. The same type of error was detected in relation to reasoning. In thinking about TV2 (Item13), Choice D, one student reasoned about the difference between severance benefits (noted in *Hamilton*) and medical benefits related to AIDS (noted in *Bridell*). He said as he selected this incorrect choice, "Severance benefit, you could always go get another job and get more money. Whereas you exhaust your limit on your insurance—you may not be able to pay for your medication. I don't think that's a defensible analogy." This student was basing his decision on a common-sense view of justice, ignoring how ERISA would categorize each example similarly as "welfare benefits." This tendency to reason from common sense was often signaled by words such as "to me" or "it seems"—dangerous terms to use when making legal arguments and clearly at odds with the convention of legal citation and the "close reading" it demands.

Some students superseded lawyerly roles and adopted postures more like judges. In considering Mackey's (TV1) efforts one student stated, "Mackey only contacted [the Prothonotary's office] once and got put on hold, and then he hung up. Yeah, he could have tried a bit harder. That doesn't seem to be that honest an effort." Others adopted the role of

policy maker as illustrated by a student who reasoned, "Maybe you could make an argument to the appellate court of policy concerns related to a catastrophic disease and that that's different and is calling for a different law other than ERISA."

Some students were lured by what the reading literature calls "seductive details" (Wade, Schraw, Buxton, & Hayes, 1993). The circumstances of the *Bridell* case understandably created a pathos for the client as an AIDS sufferer. On legal grounds, however, given the circumstances and parameters of ERISA legislation, it would have been difficult to support a theory of discrimination. On the other hand, the case did present an opening for a lawyer to question on legal grounds the comprehensibility requirements of the statute. A number of students in the think-aloud condition demonstrated a tendency to gravitate to the former and remain unaware of the latter posture toward the case. Mertz (2007) recently detailed the shifting of discursive categories that is required in reading legal cases and demonstrated the enculturation process that is at the heart of law school pedagogy. The think-aloud protocols demonstrated that some students make this transition while others struggle against more conventional readings.

Another set of dysfunctional strategies, and perhaps one triggered by the multiple-choice format, were test-based errors. One student commented that the "test makers like E," as he incorrectly selected choice E on an item. Others claimed to have found "tricks" in the test, such as a supposedly deliberate substitution of fax for letter, while some played a "match game" choosing an option because it matched words used in the text rather than focusing on the deeper meaning of an option. Finally, dismissed choices were accompanied by comments such as, "That's not really a relevant question when you're trying to figure out the problem at hand." "That's irrelevant" was the most highly used form of elimination, and often what was eliminated was the keyed response.

We do not intend in this section to unduly criticize the reading and reasoning skills of those who volunteered for our study. Indeed, as stated earlier, we were greatly impressed by the sound strategies and exemplary reasoning skills demonstrated by most of the participants. Probably one of the most fruitful strategies observed in the think-aloud condition was the practice of reading the stem and speculating on a possible answer *before* reading over the choices. In this case, "matching" was not a game, but a systematic and knowledge-based way of dealing with the problem of the item while keeping sharply in mind the purpose of the task. Students who used this strategy commented, "That's what I was saying only they said it better" or "Yeah, that's what I'm looking for."

Students' Ideas About Case Reading and Reasoning

In the final part of the think-aloud session students were asked to sit face to face with the proctor and respond to a few interview questions. We tried to determine how sensitive students were to the structure of the test. Most noted that they detected differences between what we called determinate and indeterminate items. Some categorized the former as "getting" or "go back" items, and the latter as "strategic" or "client's need" items. Only a few noted the single- versus cross-case items. Regarding perception of difficulty, most found the indeterminate items harder. Two mentioned that difficulty was a function of whether or not an issue was "seen" in the first reading. If it was, it made an item easier to deal with. Some found synthesis requirements and speculations on opposing counsel's arguments more difficult to deal with. A couple of students mentioned that they felt that personal opinion was competing with the reading task. One realized that he was "drawing too much on outside stuff." This comment might explain the lure of "seductive details" noted in the previous section and speaks to the developmental process that characterizes the acquisition of legal literacy.

In self-assessing their performance on the test, only two mentioned that the think-aloud condition may have interfered with their reading abilities. Students voiced more confidence in their performance on TV1 items than on TV2 items. In commenting on TV2, one student said that they were "pretty involved questions, not about specific details." In terms of students' confidence in their case-reading skills in general, most believed that these were "under development." Some differentiated between how they read cases for classes and how they read (or speculated about reading) cases for practice. One student mentioned working at a law firm while completing his first year. He said, "I work at a firm downtown. When I read cases for them I'm very paranoid that I'm missing something."

A majority of the students interviewed saw the relation between the kind of reading demanded by the test and the kind of reading they do in legal writing classes. Two thought the tasks of the test were more like law school tests than law school classes. One commented, "Very few instructors focus on this type of reading. It's find the holding, find the rule of law. That's a little more rudimentary than this."

In terms of their perception about the instruction they were receiving about case reading in law school, most were critical. Some felt that they (professors) "Throw us to the wind," that instruction is "almost zero—you're kind of expected to figure it out." Some spoke about silence on reading after orientation. A few mentioned specific professors who stress reading, and most thought that legal writing teachers were more direct in their instruction on reading skills necessary for law practice.

Finally, students were asked whether they would welcome additional instruction in reading. The group was about half split, but almost all said they would welcome more feedback on both reading and writing. The naysayers generally were not unconcerned about their reading development or lack thereof; rather, most saw time as the enemy of close reading. As one maintained, "When you're trying to understand a massive amount of information, you can't really spend this kind of time."

Implications

On Case Reading and Reasoning

The "issue" of reading in law school is relatively recent. Lundeberg's 1987 study that used think-aloud protocols to capture the striking differences between expert and novice readings of a case remains a cornerstone of the work that has followed. It is somewhat ironic that the conversation began with nonlawyers such as Lundeberg, Stratman, and Evensen (Deegan, 1995). However, the issue has been taken up by more voices from within the legal education community. Oates (1997) provided one of the earliest pictures supplied by legal educators of what reading successes and reading struggles look like.

Two books that set out to assist both teachers and students in understanding and practicing the new literacies necessary to both the study and practice of law became available in 2005. Schwartz's *Expert Learning for Law Students* devotes a chapter to legal reading and case briefing. Schwartz characterizes the "expert" learner of law as one who engages in such reading strategies as "text lookback," monitoring progress (especially through question posing), "dialoguing," and evaluating cases as they read. McKinney (2005) devotes an entire book (and accompanying online exercises) to *Reading Like a Lawyer*. Her primary concern is the initial shepherding of students into legal discourse, but she attempts to keep her student audience aware that reading cases involves what we have referred to in our project as sensitivity to both determinacies and indeterminacies. McKinney maintains: "Being able to accurately identify component parts of a case … is a prerequisite step to being able to discuss the case at a deeper, more meaningful level" (p. 84). It is this second goal that she refers to as artful and exciting reading. She refers to the "'space' in the case (the things that are "indeterminate" about the text, the questions not yet answered, the facts not dealt with adequately)" (p. 85).

A very different perspective toward legal reading and the reading of cases is offered by Mertz, a law professor and anthropologist, who devotes a chapter of her 2007 book *The Language of Law School* to "Learning to Read Like a Lawyer." Mertz traces the methodical instruction afforded in the Socratic classroom that transforms student reading. She asserts that "law school classroom discussions provide a kind of prism through which we can discern core features of legal readings and texts" (p. 59). She goes on to name the skills involved in this process, among them identifying facts, discerning analogies, differentiating between holdings and dicta, etc., but ends by calling these a "deceptively simple-sounding set of tasks" (p. 61).

Of course, the learning of these skills or tasks is highly constrained and directed in the classroom, where the professor controls and dominates the discourse and where, as Mertz and others have pointed out, relatively few student voices are heard. Schwartz (2003) referred to the participation of the great majority of students in law classrooms as "vicarious learning" and challenged the largely untested assumption that the uncalled-upon students are paying attention and sensitive to the subtle discursive moves on display.

Students who have been interviewed about their perceptions of case reading, how cases are treated in traditional classrooms, and the instruction available on case reading speak of getting "lost in cases" (Sullivan, Colby, Wegner, Bond, & Shulman, 2007, p. 41). Indeed, Wegner's (Sullivan et al., 2007) empirical investigation of law school teaching describes "what students learn in the first-year curriculum and what they could learn but typically do not" (p. 21). Among her list of skills she includes:

> *Legal Literacy.* Students are trained to develop legal literacy through emphasis on vocabulary, close reading, and textual interpretation, all of which contribute to their ability to develop their knowledge and comprehension of the field. Faculty often model important ways of "thinking about thinking" particularly with regard to testing one's own knowledge and understanding, but rarely cue students explicitly about what they are doing or elaborate on the importance of such skills.

Similar to Schwartz's criticism noted above, Wegner faults the legal education community for assuming that learning occurs vicariously.

We understand that the case reading practiced in law school classrooms, particularly within the first year, corresponds more to what Sullivan et al. refer to as the *cognitive apprenticeship* of legal education than the *practical apprenticeship* of it. The Carnegie study of professional preparation in law, published with Sullivan as first author, urges the legal education community to think more broadly of their goals and the desirability of incorporating these two apprenticeships with a *formative apprenticeship* that acknowledges the ethics of legal practice. The respondent in the Carnegie study who was quoted above as "lost in cases" went on to explain that when engaged in a more nontraditional, simulation-based classroom "found myself discovering new skills and I enjoyed that—applying the ideas and seeing them in different ways. Everything (including cases) really started to come together" (p. 41). The Carnegie study urges legal educators to look to other professional fields for "instances of pedagogies that teach complex practical reasoning and judgment, blending the cognitive and practical apprenticeships" (p. 99). There is no reason, given the tenor of the book, why the *formative apprenticeship* could not be included in this statement. Going on to explain how this might be

done, Sullivan et al. point to the need for more "significant work" that would include "assessing the skills and qualities of the practical and ethical-social apprenticeships" (p. 175).

We see our reading and reasoning tests as modest assessment efforts that might bridge the cognitive and practical apprenticeships through more direct instruction in reading.

What Does Our Research Demonstrate About Students' Case Reading and Reasoning Abilities?

If we look only at the statistical results of our study that, in the end, involved over 300 law students, it looks as though our law students' case-reading abilities need attention. The aggregate averages for both versions of the test were calculated at around 60%; students do not appear to improve in their case reading and reasoning abilities over their 3 years of legal education; and test scores do not correlate very highly with law school GPAs. However, we need to look more closely at these results. First, the fact that performance on both versions of the test was normally distributed attests to the proposition that this ability appears to constitute an *individual difference* among law students. In other words, most students fall within the middle range, but some excel while others falter. Extensive research in educational psychology and instructional systems should convince even the most skeptical that individual differences can be diminished through instruction.

In terms of improvement we might argue, as we did in our first study, that the kind of case reading we are operationalizing through the multiple-choice test, is one not stressed or practiced extensively in law schools. The literature discussed above largely aligns with this assertion. We did try to survey the third-year students in our sample to see if experiences such as clinics, externships, or legal work might have contributed to improved abilities, but analyses of these data provided no insights. We agree with legal educators who stress the importance of developing students' skills in dealing with determinacies in cases, but do not see early attention to more practice-related skills with indeterminacies as necessarily interfering with the development of the former. There is much written within the reading research literature that speaks to the importance of *flexibility* in reading (Feltovich, Spiro, Coulson, & Myers-Kelson, 1995; Spiro, Coulson, Feltovich, Anderson, 1994) and readers' sensitivities and abilities to deal with varying purposes (Gee, 1996). This issue seems to be largely absent from educational scholarship within the legal education community. Early works by Hofer (1987) and Stratman (1990, 1994, 2002) provide the assumption regarding the function of purpose in case reading governing the present studies, and the role of purpose in case reading might be usefully revisited by legal educators. Finally, with regard to the low correlations between the test and standard prognostic and performance assessments, it must be remembered that correlations can also run negatively. Our analyses yielded only positive correlations, and we included some graphics in our report to indicate that for at least the high and low scorers, there seemed to be an unmistakable correspondence between LSAT/LGPA scores and test performance. Within the think-aloud data, coincidentally, the only student to get all 8 items correct had one of the highest GPAs (3.72), but not one of the highest LSAT scores (157). Conversely, one of the lowest scorers (2 out of 8) had the lowest GPA among the group (2.53), but an LSAT score of 161. To sort out these factors is not easy. Indeed, the modest correlation typically reported between the LSAT and first-year GPA (reported by LSAC to have a median of about .39) continues to perplex many legal educators.

Our optimism about law students' case reading and reasoning derives mostly from the think-aloud data we collected and analyzed. The fact that we so often found at least one student who could read and reason in the ways we had predicted gives us confidence that these skills can be enacted in a traditional assessment form. The think-alouds also provide solid evidence that our theoretical constructs conform to empirical data, thus strengthening our claim for test-construct validity. Although numerous types of validity can be found in the psychometrics literature, construct validity is seen as a most vital standard. Messick (1989), who argues for a unified theory of validity, states that, "construct validity is based on an integration of any evidence that bears on the interpretation or meaning of the test scores" (p. 17). He goes on to explain that validation is an "argument" or a collection of evidence to support a given interpretation of a test score. We see the think-aloud data, especially when added to our statistical and expert data, as giving force and authority to our argument about the validity and equivalence of our two test versions. The same source of data, however, also pointed out to us the shortcomings of our efforts; however, overall we believe we established that our test is measuring something important to the literacy and professional development of law students.

In addition, these more qualitative analyses allowed us to specify the hierarchy of skills associated with the general task of reading cases and the specific task of reading these particular cases. Finally, the analyses provided a window into problematic areas of students' efforts to read and reason through cases. We feel confident that this research contributes to the knowledge base of learning in law school

Uses of the Tests

One of the reasons we credit for the generous cooperation we received from law schools and law students during this research is that they perceived the usefulness of our enterprise. The call for increased assessment, particularly formative assessment, in law schools has been loud and clear in recent years, and one of the stumbling blocks to such efforts has been a lack of tools or instruments of assessment. To conduct think-alouds of student reading is an onerous task and

probably best left to researchers. However, many law teachers have successfully used demonstration think-alouds in classes where they are attempting to portray expert reading. Academic support personnel, who often work one-on-one with struggling students, also make conscious or instinctive uses of think-alouds, but rarely have the chance or the time to transcribe interactions and subject them to the microanalysis expected in formal research studies.

We hope that our tests might provide one tool that could be more efficiently used to diagnose student reading and reasoning performance. The tests might be used in standard, group administrations to pre- and post-test case-reading instruction. They might also be used selectively with small groups of students in need of tutorial services. Additionally, they might be used with their justifications in one-on-one tutorials where the tutor would be listening for correspondence between a student reasoning through choices and the hypothesis related to that activity. Again, this latter use would be more structured than asking a student to think aloud while reading a case. Essentially, the student would be cued by the test choices to certain theoretically justified ways of thinking.

We would also hope that this report could be used in conjunction with the tests per se, especially our tables that provide lists of constructs, skills, and subskills of case-reading tasks. Along with those, we would invite readers of this report to make use of the descriptions of dysfunctional strategies we detected among our student participants as a way to anticipate and correct such errors.

Finally, we see this research as applicable for doctrinal classrooms where the daily "case dialogue" (see Wegner as cited in Stuckey et al., 2007) assumes understanding of particular aspects of cases. Law professors might try their hand at composing single, multiple-choice items that focus on aspects of an assigned case deemed essential to the day's work. Technologies available in most law school classes would allow the teacher to present an item on a PowerPoint slide at the beginning of class and have students "key in" choices. If a majority of students "miss" the item, it would be obvious that some direct teaching of case reading and reasoning would be necessary. Likewise, teachers of legal reading and writing might productively use the tests as they were constructed in early sessions with first-year students or attempt to emulate some of our item types to comport with cases they are using for open or closed memos.

Future Research

We propose five lines of research that could follow up on the work begun through this project. First, despite the extensive field-testing of our two tests, these still remain in need of further testing. We hope to solicit law schools interested in using the tests, particularly in orientation sessions or pre-first-year programs. We would invite these groups to provide us with anonymous data relevant to their use. Second, we intend to disseminate this report at annual meetings of legal educators. We presented preliminary findings from this study at the Legal Writing Institute biannual meeting in 2006 (Atlanta) and final results at their 2008 meeting in Indianapolis. Third, we hope to produce an item-writing manual to serve as an addendum to this report that would be published by a publisher interested in work in legal education. We see this as useful most especially to skills-based and doctrinal teachers as mentioned above. Fourth, we would like to continue this research with legal educators who are interested in researching their own practice. Specifically, we would like to work with these people to construct interventions aimed at improving law students' case reading and reasoning skills and to devise systematic ways of evaluating these interventions. Again, our tests could be among the tools necessary to enact such research. Our last research goal, the fifth, we see as the most difficult, but significantly facilitated if linked with collaborative efforts among law schools: that is, to trace the development of these skills through law school and perhaps into professional practice. We see this as a huge challenge, but one worth pursuing.

As we stated in our first LSAC report, we did not set out to compose the definitive test of case reading. Ours is a prototype that might allow for continued efforts to better understand these crucial skills of learning and practicing law.

References

American Bar Association Section of Legal Education and Admission to the Bar. (1992). *Legal education and professional development—An educational continuum: Report of the Task Force on Law Schools and the Profession: Narrowing the gap (McCrate Report).* Chicago, IL: American Bar Association.

Blasi, G. L. (1995). What lawyers know: Lawyering, expertise, cognitive science, and the function of theory. *Journal of Legal Education, 45*(3), 313–397.

Bryden, J. (1984). What do law students learn? A pilot study. *Journal of Legal Education, 34,* 479–506.

Chinn, C., & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49.

Christensen, L. M. (2007). Legal reading and law school success: An empirical study. *University of Seattle Law Review, 30,* 1–47.

Deegan, D. H. (1995). Exploring individual differences among novices reading in a specific domain: The case of law. *Reading Research Quarterly, 30*(2), 154–170.

Evensen, D. (2004). To group or not to group: Students' perceptions of collaborative learning activities in law school. *Southern Illinois University Law Journal, 28,* 343–421.

Fajans, E., & Falk, M. (1993). Against the tyranny of paraphrase: Talking back to texts. *Cornell Law Review, 78,* 163–205.

Feltovich, P., Spiro, R., Coulson, R., & Myers-Kelson, A. (1995). The reductive bias and the crisis of text in the law. *Journal of Contemporary Legal Issues, 6*(1), 187–212.

Gee, J. P. (1996). *Social linguistics and literacies: Ideology in discourses.* Bristol, PA: Taylor & Francis Ltd.

Graesser, A., Langston, M., & Baggett, W. (1993). Exploring information about concepts by asking questions. In G. V. Nakamura, R. M. Taraban, & D. Medin (Eds.), *The psychology of learning and motivation: Vol. 29. Categorization by humans and machines* (p. 411–436). Orlando, FL: Academic Press.

Graesser, A., & Olde, B. (2003). How does one know whether a person understands a device? The quality of the questions a person asks when the device breaks down. *Journal of Educational Psychology, 95*, 524–536.

Graesser, A., & Person, N. (1994). Question asking during tutoring. *American Educational Research Journal, 31,* 104–137.

Gross, P. (1973). On law school training in analytic skill. *Journal of Legal Education, 25,* 261–311.

Hacker, D., Dunlosky, J., & Graesser, A. (Eds.) (1998). *Metacognition in educational theory and practice.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hofer, P. (1987). Cognitive strategies for interpreting law. *Dissertation Abstracts International, 48,* 5-B. (UMI No. 87-16612).

Hoeflich, M. H. (1991). On reading cases: The law student in Wonderland. *Syracuse Law Review, 42*, 1163–1187.

Karabenick, S. A., Kempler, T. M., & Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42,* 139–151.

Kissam, P. C. (2000). The ideology of case method/final examination law school. *University of Cincinnati Law School, 70*, 137–189.

Lundeberg, M. A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in legal case analysis. *Reading Research Quarterly, 22*(4), 407–432.

Lustbader, P. (1997). Construction sites, building types, and bridging gaps: A cognitive theory of the learning progression of law students. *Willamette Law Review, 33,* 315–381.

Maki, R. (1998). Test predictions over test material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.) *Metacognition in educational theory and practice* (pp. 117–144), Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McKinney, R. A. (2005). *Reading like a lawyer: Time-saving strategies for reading law like an expert.* Durham, NC: Carolina Academic Press.

Mertz, E. (1998). What difference does difference make? The challenge for legal education. *Journal of Legal Education, 48,* 46–87.

Mertz, E. (2007). *The language of law school: Learning to "think like a laywer."* New York: Oxford University Press.

Messick, S. (1989). Validity. In R. L. Linn (Ed.) *Educational Measurement* (3rd ed.). New York: Macmillan.

Mills, C., Diehl, V., Birkmire, D., & Mou, L.-C. (1995). Reading procedural tests: Effects of purpose for reading and predictions of reading comprehension models. *Discourse Processes, 20*, 79–107.

Minnis, M. (1994). Toward a definition of law school readiness. In V. John-Steiner, C. Panofsky, & L. W. Smith (Eds.), *Sociocultural approaches to language and literacy* (pp. 345–390). Cambridge: Cambridge University Press.

Moore, M. (1981). The semantics of judging. *Southern California Law Review*, *54*(2), 151–294.

Munro, G. S. (2000). *Outcomes assessment for law schools.* Spokane, WA: Institute for Law School Teaching, Gonzaga University Law School.

Oates, L. C. (1997). Beating the odds: Reading strategies of law students admitted through alternative admissions programs. *Iowa Law Review, 83*(1), 139–160.

Otero, J. (1998). Influence of knowledge activation and context of comprehension monitoring of science texts. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 145–164). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Otero, J., & Graesser, A. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction, 19*, 143–175.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 3*(4), 229–235.

Palasota, A. (1991). Expertise and the law: Some recent findings from the cognitive sciences about complex human information processing. *Thurgood Marshall Law Review, 16,* 598–626.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66,* 181–221.

Schlag, P. (1989). Missing pieces: A cognitive approach to the law. *Texas Law Review, 67*(6), 1195–1251.

Schwartz, M. S. (2001). Teaching law by design: How learning theory and instructional design can inform and reform law teaching. *San Diego Law Review, 38,* 346–362.

Schwartz, M. S. (2003). Teaching law students to be self-regulated learners. *Detroit College of Law Review*, 447–505.

Schwartz, M. H. (2005). *Expert learning for law students*. Durham, NC: Carolina Academic Press.

Senger, C. J. (1993). Thinking aloud protocols: A diagnostic tool for teaching legal problem solving. *Thomas M. Cooley Law Review, 10*, 367–382.

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. K. (1994). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed.). Newark, DE: International Reading Association.

Stratman, J. (1990). The emergence of legal composition as a field of inquiry: Evaluating the prospects. *Review of Educational Research, 60*(2), 153–235.

Stratman, J. (1994). Investigating persuasive processes in legal discourse in real-time: Cognitive biases and rhetorical strategy in appeal court briefs. *Discourse Processes, 17*, 1–57.

Stratman, J. (2002). When law students read cases: Exploring relationships between professional legal reasoning roles and problem detection. *Discourse Processes, 34*(1), 57–90.

Stratman, J. (2004). How legal analysts negotiate indeterminacy of meaning in common law rules: Toward a synthesis of linguistic and cognitive approaches to investigation. *Language and Communication, 24,* 23–57.

Stratman, J., Evensen, D., & Oates, L. (2005). Developing an assessment of 1st year law students' critical case reading and reasoning ability. Newton, PA: Law School Admission Council.

Stuckey, et al. (2007). *Best practices for legal education: A vision and a road map.* Columbia, SC: Clinical Legal Education Association.

Sullivan, W. M., Colby, A., Wegner, J. W., Bond, L., & Shulman, L. S. (2007). *Educating lawyers: Preparation for the profession of law.* San Francisco, CA: Jossey-Bass.

Wade, S. E., Schraw, G., Buxton, B., & Hayes, M. (1993). Seduction of the strategic reader: Effects of interest on selective attention and recall. *Reading Research Quarterly, 28,* 93–114.

Weinstein, I. (1998). Lawyering in the state of nature: Instinct and automaticity in legal problem solving. *Vermont Law Review, 23*(1), 1–57.

Wetlaufer, G. B. (1990). Rhetoric and its denial in legal discourse. *Virginia Law Review, 76,* 1545–1597.

Wightman, L. F. (1993). *Predictive validity of the LSAT: A national summary of the 1990–1992 correlation studies*. (Technical Report 03-01). Newtown, PA: Law School Admission Council.

Wiley, J., & Voss, J. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for texts. *Journal of Educational Psychology, 91*(2), 301–311.

Woolley, S. L., Benjamin, W. J., & Woolley, A. W. (2004). Construct validity of a self-report measure of teacher beliefs related to constructivist and traditional approaches to teaching and learning. *Educational and Psychological Measurement, 64,* 310–331.

Acknowledgments

We would like to thank Christian Anderson and Ty Glade for their help with data management.

This study received funding from the Law School Admission Council (LSAC). The opinions and conclusions contained in this report are those of the author and do not necessarily reflect the position or policy of LSAC.

Appendix 1: TV1. Cases and Items Used in Field-Testing

Mackey Plumbing Co. v. Pepper (2002)
Court of Common Pleas of Allegheny County, Pennsylvania

An assumpsit action was heard by a board of arbitrators in Allegheny County. The board's final decision as to the arbitration award to be made was docketed on October 10, 2002. Mackey (the defendant) disagreed with the arbitrator's decision to award Pepper $2394 in damages for an improperly installed drainage pump in Pepper's home. To proceed with his appeal of this decision, Mackey paid $162 to the county for his (previous) use of the county arbitrators and timely paid the $25 filing fee. Mackey also paid $237 of the $327 that he owed to Pepper's counsel for accrued record costs, but this payment arrived late, 2 days after the 20 day limit. For this failure to pay all record costs and to do so within the stipulated 20 day limit, the plaintiff (Mr. Pepper) moved to quash Mackey's appeal of the arbitrator's award ($2394 in damages). Act of June 16, 1836, Pa. Law 715, Sec. 27, 5 Pa. S., Sec. 71.

We find for the plaintiff; appeal quashed.

Woodson, Jason.

1. Arbitration and Award

The right to appeal an arbitrator's decision is a substantial right but it is within the court's discretion to decide what reasonable conditions can be imposed upon this right, or when the difference between stated conditions and a defendant's attempted compliance should be considered a mere technicality and disregarded. Defendant's (Mackey's) failure here to pay the complete amount of record costs accrued and to do so on time justifies quashing his appeal and constitutes more than a technicality. P.S. Const. Art. 5, Sec. 9.

2. Arbitration and Award

The right to appeal an adverse decision of an arbitrator is rightfully subject to conditions that all costs be paid within twenty (20) days. Act of June 16, 1836, Pa. Law 715, Sec. 27, 5 Pa. S., Sec. 71.

WOODSON, Judge: (Opinion)

(1) The present motion raises the question of whether nonpayment of record costs, and failure to do so in a timely manner, should disallow the defendant from making an appeal from the arbitrator's decision under Pennsylvania rules. The particular issue in this motion, then, is whether the defendant's appeal should be quashed. The defendant in order to perfect his appeal from the arbitration decision was first required to pay $327 in record costs to plaintiff's attorney. Defendant had already timely paid $162 for the assistance of the county arbitrator and the $25 filing fee, as required. The plaintiff's (Pepper's) attorney concedes that the defendant (Mackey) had faxed him a letter requesting to know the amount that he owed for record costs. The plaintiff's handwritten letter in pencil in response to this request stated the following:

> "In response to your request to know the amount of record costs outstanding, please pay $327 and do so by November 3rd."

Defendant claims this pencil note was extremely difficult to read as the amount of record costs appeared to have been erased and rewritten several times; for this reason, defendant states he thought the correct amount to have been $237 and not $327. Defendant claims that his error was an honest one of transposing the numbers "2" and "3." He further claims that he wrote to the plaintiff to obtain clarification but no answer was forthcoming. Defendant claims he would have telephoned the plaintiff to verify the correct amount owed but did not do so on the assumption the plaintiff would contact him if the payment he eventually sent was insufficient. Defendant claims also to have called the Prothonotary's office to ascertain the exact amount, to have been put on hold, and never to have received a response. The plaintiff, in turn, claims to have made two phone calls to the defendant to inform him of his error, but was unable to reach him. The plaintiff made no other effort to contact the defendant before this motion to quash.

Defendant also claims that he relied upon the plaintiff's notice that payment could be received by November 3rd (also indicated in the pencil note), which is actually beyond (by two days) the twenty day limit allowed by law. Payment should have been received by the plaintiff or Prothonotary by October 31st.

Our inspection of the pencil note suggests it may indeed have been somewhat difficult to decipher (as to the amount owed) yet we cannot see that the defendant nonetheless made a valid or honest attempt to comply. The analysis in Black and Brown, Inc. v. Home For The Accepted, 335 Pa.Super. A.2d 722 (1975), in which the shortfall in costs was $74, requires a "valid attempt to make … full payment." We therefore hold that a shortfall in record payment costs of $90 in the present case cannot be excused as de minimis, nor should this shortfall be excused by the court as a technicality thwarting the Pennsylvania constitutional right of appeal under Article V. Section 9 simply because the defendant erroneously interpreted plaintiff's communication.

We find for the plaintiff; appeal quashed.

(2) We further regret that defendant relied upon miscalculations of the plaintiff as to the date when record costs were due, but defendant should have known the rules and acted upon them instead. We agree with the analysis in Meta v. Yellow Cab Company of Philadelphia, 222 Pa.Super. 469, 294 A. 2d 898 (1972), that "the right of appeal is properly subject to a fixed temporal condition of twenty days," and that here as in that case "it serves a meaningful purpose to know when a matter has finality." We do not see that the defendant made an honest attempt to meet the deadline. We find for the plaintiff; appeal quashed.

Meta v. Yellow Cab Company of Philadelphia (1972)
Superior Court of Pennsylvania

This is an appeal from a decision in compulsory arbitration proceedings. The Court of Common Pleas, Philadelphia County, quashed the defendant's (Yellow Cab's) appeal and defendant appealed. The Superior Court, No. 748 October Term, 1971 (Judge Packel) held here that the requirement that a party appealing from an arbitration decision pay the other party's costs of record within 20 days of decision was de minimis; thus, defendant's payment to plaintiff's attorney of only $10.00 for record costs instead of $17.75 required by law did not justify quashing the appeal.

We find for the defendant. Reversed.

Judge Hoffman dissented and filed an opinion in which Judge Watkins and Judge Jacobs joined.

1. Arbitration and Award

Right of appeal from decision of arbitrators is a substantial right but is not absolute and can be subjected to reasonable conditions. P.S. Const.. Art. 5, Sec. 9.

2. Arbitration and Award

Requirement that party appealing from decision of arbitrators pay record costs of other party within 20 days of decision is de minimis; thus, defendant's payment to plaintiffs' attorney of only $10.00 for record costs of compulsory arbitration proceedings instead of $17.75 as required by law did not justify quashing the appeal from the decision of the arbitrators filed by the defendant who had paid the filing fee and $120.00 to reimburse county for fees of arbitrators. 5 P.S., Sec. 71; P.S. Const. Art. 5, Sec. 9.

3. Arbitration and Award

The right to appeal from a decision of arbitrators is properly subject to conditions that the appeal be taken within 20 days of the decision and that security be given. 5 P.S.Sec. 71; P.S. Const. Art. 5, Sect. 9.

This case heard before WRIGHT, P. J., and WATKINS, JACOBS, HOFFMAN, SPAULDING, CERCONE, and PACKEL, JJ.

PACKEL, Judge: (Opinion)

The greatness of the common law as a system of jurisprudence is ascribable to the principle that a rule which has no reason for its present existence should not defeat substantial rights. A corollary of that principle is the recognized practice of the courts to disregard matters de minimis.

[1] A right of appeal is unquestionably a substantial right. In 1968 it was expressly made a constitutional right, Pa. Const. Art. V., Sec. 9, P.S. As to compulsory arbitration proceedings, the constitutional right to appeal from the decision

of the arbitrators was held to be mandated in view of the constitutional right to jury trial. Smith Case, 381 Pa. 223, 112 A.2d 625 (1955). The right to appeal is not absolute and can be subjected to reasonable conditions.

The specific issue in this case is whether the defendant's appeal from compulsory arbitration was properly quashed because only $10 instead of $17.75 was paid to plaintiff's attorney for record costs. The amount awarded by the arbitrators for the plaintiffs was $3000. The defendant in order to take the appeal had paid the filing fee and $120 to reimburse the county for the fees of the three arbitrators. The court below, after argument, had denied the motion to quash the appeal, but more than a month later it reconsidered the matter and quashed the appeal. It is admitted that prior to the end of the 20 day period for taking an appeal defendant's counsel called plaintiff's counsel to obtain his costs and also wrote to him:

> "Because I have not yet received a copy of your bill of costs, I am enclosing our draft for $10.00 in payment of your costs of record. If this amount is not adequate, I will be glad to review it with you or, upon receipt of your bill of costs, will promptly consider the items listed therein."

In another case currently before the Court, Holmes v. Broodno, Pa.Super., 294 A.2d 903 (1972), the award by the arbitrator was in the amount of $9000, and the appeal was quashed because $20.00 was not paid as record costs, even though $122.50 had been paid to take the appeal. It was agreed by both sides that the defendant's attorney within the 20 day period called, and wrote to, plaintiff's attorney, asking for the amount of plaintiff's costs, and that during the 20 day period he received no word from plaintiff's counsel.

Ancient cases have dealt with the issue when arbitration was compulsory only if demanded by one of the parties. A series of subsequent cases in the past eight years have dealt with the issue in connection with arbitration compulsory on all parties. The issue has become of much more importance because of the increase in the amount involved in compulsory arbitration in several steps from the original limit of $1000 to the very recent limit of $10,000. The old and the new cases are to be given due regard but it must be in the light of the changed economic conditions and different concepts as to the sporting theory of justice which mechanically called for absolute technical compliance.

The requirement for the payment of costs in order to effect an appeal from arbitrators has an old history. The Act of March 20th, 1810, 5 Sm. Laws of Pennsylvania 131, 135 contains the proviso in Section XI: "Provided also, that no appeal shall be allowed to either party until the appellant pay all the costs that may have accrued on such suit or action." The Act of June 16, 1836, P.L. 715, Section 27, 5 P.S. Section 71, as amended, and presently effective, provides that an appealing party

> "shall pay all the costs that may have accrued in such suit or action … Such appeal shall be entered, and the costs paid, and recognizance filed, within twenty days after the day of the entry of the award of the arbitrators on the docket … [A]ny party appealing shall first repay to the county the fees of the members of the Board of Arbitrators herein provided for, but not exceeding fifty percent of the amount in controversy."

In light of the exaltation of technicality during earlier periods of our jurisprudence it was characteristic that the first cases made it clear that, under the contemporary arbitration acts, an appeal would be dismissed or quashed for failure to pay record costs within twenty days. Walter v. Bechtol, 5 Rawle 228 (1835). (An earlier attack upon the constitutionality of requiring the payment of costs in an arbitration appeal was unsuccessful in McDonald v. Schell, 6 S. & R. 240 [1820]). The ancient rigidness of this rule is seen in Ellison v. Buckley, 42 Pa. 281 (1862), which held that payment of costs must be in cash. The early obsession with form is further illustrated in the shocking case of Thompson v. White, 4 S. & R. 135 (1818), where an appeal was dismissed because the affidavit on appeal stated that the appellant "believes injustice has been done" instead of stating that he "<u>firmly</u> believes injustice has been done"!

Eventually our courts recognized that the harshness created by the hypertechnical application of these procedural rules was effecting more injustice than serving any substantial purpose. This more humane approach is seen in Schrenkeisen v. Kishbaugh, 162 Pa. 45, 50, 29 A. 284, 286 (1894), where Justice Mitchell reversed an order dismissing an arbitration appeal and that dismissal was based on the failure to pay the appeal costs directly to the prothonotary, stating:

> "To hold that there was an inexorable necessity that the money should be paid first to the prothonotary and then by him to the constable, and that a payment directly by the party liable to the party entitled finally to receive it was not a valid payment, would be putting form in place of substance, and technicality before right."

Shortly after that decision the rule requiring payment of costs in cash was repudiated and the use of checks was permitted, Burns v. Smith, 180 Pa. 606, 37 A. 105 (1897).

In our more recent times, in Beth-Allen Sales Co. v. Hartford Insurance Group, 217 Pa.Super. 42, 268 A.2d 203 (1970), our court held that an arbitration appeal need not be quashed for non-compliance with the detailed requirement of recognizance. There Judge Hoffman followed the modern view of the sufficiency of substantial compliance (217 Pa. Super. at 47, 268 A.2d at 206):

> "This decision is based on the belief that where a party has made an honest effort to file his appeal in accordance with the statute, and has substantially complied with the requirements, justice will not permit his appeal to be dismissed with prejudice."

He underscored the undesirability of reliance upon immaterial formalities (217 Pa. Super. at 48, 268 A.2d at 206):

> "To quash the appeal in these circumstances would result in a return to the supremacy of form over substance and the exaltation of technical detail over justice, an approach which courts in all areas have been opposing for many years."

Notwithstanding the foregoing, the last word on the specific issue is in Manton v. Marini, 218 Pa.Super. 298, 280 A.2d 403 (1971), Wright, P. J. dissenting. There the plaintiff moved to quash because the defendant had sent a check to plaintiff's counsel for $32.50 instead of $34.50. This Court affirmed the quashing of the appeal even though the defendant had paid all the other costs, even though the arbitrator's finding for the plaintiffs' was in the amount of $7902, and even though the defendant had filed a bond in the amount of $15,804. This decision, all because of the non-payment of $2.00!

With inflation making the few dollars of record costs worth less, and with the increase in the amount of money which may be involved in arbitration proceedings, it becomes more manifest that record costs as compared to the dollars in controversy are matters de minimis. Ordinarily regular record costs in Philadelphia rarely exceed $30. Bills of costs can be in varying amounts but normally are much less than the $35 which must be paid by the appealing party for each member of the arbitration panel, plus an extra $15 for the chairperson.

[2,3] The right of appeal is properly subject to a fixed temporal condition of twenty days. It serves a meaningful purpose to know when a matter has finality. Likewise, a requirement of security as a condition of appeal is a reasonable requirement within the legislative province. Commonwealth v. Philadelphia Eagles, Inc. 437 Pa. 25, 261 A.2d 309 (1970). The requirement of payment to counsel for the other side of a relatively small amount with all kinds of variations as to what that "small" amount is, serves no real function as a condition of appeal and can raise a serious question of constitutionality. In any event, the requirement can truly be viewed as a matter de minimis. Although attorneys are supposed to know the law and to know how to take appeals, there is no reason to continue to maintain a legal trap which may cut down some appeals, but which puts counsel in a demeaning position, creates bad public relations and tends to cause additional malpractice litigation.

The time has come to prohibit the use of a meaningless de minimis condition to quash appeals. That position stands as a blemish upon a great program, initiated in Pennsylvania and being followed elsewhere, for the improvement of the administration of justice. More consonant to the demands of some legal technicians, the same result can be reached by stating very simply that the 20 day time limit and the furnishing of security are properly jurisdictional requirements but the requirement for the payment of costs is directory rather than mandatory.

The order of the court below quashing the appeal from the decision of the arbitrator is reversed.

HOFFMAN here files a dissenting opinion in which WATKINS and JACOBS join.

HOFFMAN, J. (dissenting):

I agree with the majority that the right of appeal is a substantial right which may be subject to reasonable conditions. Nowhere, however, does the majority establish that the requirement of paying the record costs within 20 days is an unreasonable condition.

The Pennsylvania legislature has decreed that a party may appeal from an award of arbitrators under the following excerpted rules:

> "II. Such party, his agent or attorney, shall pay the costs that may have accrued in such suit or action …
> IV. Such appeal should be entered, and the costs paid, and recognizance filed, within 20 days after the day of entry of the award of arbitrators on the docket … "

Thus, as stated in the Smith case, 381 Pa. 223, 112 A.2d 625 (1955), "There can be no valid objection, therefore, to the provisions of the Act of 1836, unchanged by the Act of 1952, regarding the payment of the accrued costs and the giving of a recognizance for the payment of the costs to accrue in the appellate proceedings as the condition for the allowance of an appeal from the award of the arbitrators."

So, in the instant matter, the appellant failed to comply with the aforementioned statute in that all costs were not paid prior to the expiration of the twenty day period. It is admittedly harsh to deny this appellant his right to appeal because he mistakenly paid $10.00 in costs instead of $17.75 required by law. This appellant, however, cannot plead ignorance of the costs; the appellant should have checked with the Prothonotary's office to determine the true costs owed. "It is the act, not this Court, which laid down the rule that the party appealing 'shall pay all costs which may have accrued in such suit or action.'" Fleisher v. Kaufman, 206 Pa. Super. 378, 381, 212 A.2d 846, 848 (1965).

This court cannot simply disregard the legislature's command—payment of costs within the twenty day period has been held mandatory to perfect the appeal. If this requirement is to be changed, the legislature, and not our Court, must amend the statute. Since the appellant did not pay the costs, his appeal should be dismissed.

Black and Brown, Inc. v. Home For the Accepted, Inc. (1975)
Superior Court of Pennsylvania

In an assumpsit action, Home For the Accepted, Inc. (the appellant) lost a decision before a board of arbitrators. The appellant tried to appeal, but despite compliance with other requirements, the appellant made no effort to pay accrued costs to the appellee (Black and Brown, Inc.) prior to the expiration of the 20 day appeal period. Based on this failure to tender costs, the appellee then moved to quash Home's appeal. The Court of Common Pleas of Philadelphia County, at No. 507 September Term, 1971, Trial Division, Law, Victor J. Di Nulile, J. granted the appellee's motion and ordered Home's appeal quashed. Home now appeals this motion to this Court.

The Superior Court, No. 844 October Term, 1974, Van der Voort, J. held that nonpayment of costs warranted dismissal of the appeal. Affirmed.

[1] Arbitration

Requirement that record costs be paid during appeal period is mandatory, but a valid attempt to make such timely and full payment, coupled with substantial though incomplete compliance with the requirement, should not result in harsh finality of an order quashing an appeal from arbitration; rather, the courts should examine the appellant's attempts at compliance in order to determine whether an honest effort has been made to meet the requirements of the statute; overruling Meta v. Yellow Cab Company of Philadelphia, 294 A.2d 903. P.S. 71.

[2] Arbitration

Nonpayment of costs warranted dismissal of appeal from award of arbitration, where no timely attempt to tender costs was made by appellant despite express notice of the requirement that this be done. 5 P.S. 71.

Before WATKINS, President Judge, and JACOBS, HOFFMAN, CERCONE, PRICE, VAN der VOORT and SPAETH, JJ.

VAN der VOORT, Judge (opinion):

This appeal again raises the issue of whether the nonpayment of costs, in an appeal from an award of arbitration, should support a dismissal of the appeal. Our review requires a re-examination of the holding in Meta v. Yellow Cab Company of Philadelphia, 222 Pa.Super. 469, 294 A. 2d 898 (1972).

The record in the present case shows that on September 9th, 1971, a Complaint in Assumpsit was filed by appellee (Black and Brown, Inc.) against appellant (Home for the Accepted, Inc.) in the Court of Common Pleas. Appellee had previously filed a mechanic's lien for alleged non-payment on cabinets and appliances furnished to the appellant. The assumpsit action was heard by a Board of Arbitrators and on December 3, 1973 an award was filed in favor of the appellee for $2,475 plus interest. The award was docketed on December 28, 1973.

The parties agreed that on December 10th, 1973, appellant's trial counsel inquired of, and was informed by, counsel for the appellee, that actual accrued costs amounted to $102. Appellant's trial counsel notified his client of the costs and the necessity for payment of the costs if an appeal was to be taken. Appellant disregarded this notice and dismissed its trial

counsel without notice to appellee. Appellant did not inform new counsel of its prior notice as to costs. An appeal from the arbitrator's award was made on January 10th, 1974. But despite compliance with other requirements for making an appeal, appellant made no effort to pay accrued costs to the appellee prior to the expiration of the appeal period.

On February 15, 1974, appellee filed a Motion to Quash the appeal based on the failure of appellant to tender costs. On March 14, 1974, appellant's counsel offered appellee's counsel a check for $102, but such offer was refused. On March 22, 1974, the lower court granted appellee's Motion and ordered the appeal quashed. This present appeal by appellant followed.

Now the appellant maintains that the action of the lower court was erroneous in light of the holding of Meta v. Yellow Cab Co. of Philadelphia. In that case, the defendant's attorney paid to the plaintiff's attorney only $10 for record costs associated with the arbitration proceedings, rather than the actual amount due, $17.75. Judge Packel, in the majority opinion, joined by then President Judge Wright, and Judges Spaulding and Cercone, held essentially that the "… 20 day time limit and the furnishing of security are properly jurisdictional requirements, but the requirement for the payment of costs is directory rather than mandatory" (underline added). In a Dissenting Opinion, Judge Hoffman, joined by President Judge Watkins and Judge Jacobs, argued essentially that the requirement was mandatory that costs be timely paid to perfect the appeal, and that the Court should not disregard the clear legislative command by holding otherwise.

The lower court in the present case distinguished Meta from the facts of this case since there was no attempt by the appellant in this case to pay any costs during the appeal period. In our review of the instant appeal we find significant the fact that no timely attempt to tender costs was made by the appellant (Home for the Accepted, Inc.), despite express notice of the requirement that this be done.

Our recognition in this regard leads us to conclude that the holding in Meta must be overruled insofar as it declares that the statutory requirement of the payment of costs in this type of appeal is directory rather than mandatory. We simply cannot condone a complete refusal and failure to pay the record costs in an appeal from arbitration.

To be sure, both the majority and the minority views in Meta recognized the harshness of denying a party the right of appeal when he mistakenly paid $10 in record costs, rather than the actual record costs of $17.75. The majority pointed out that in other cases, our Court has followed a principle of the sufficiency of substantial compliance. We now adopt that rationale and hold that the requirement that record costs be paid during the appeal period is mandatory— but with the caveat that a valid attempt to make such a timely and full payment, coupled with substantial though incomplete compliance with the requirement should not result in the harsh finality of an order quashing an appeal from arbitration. Rather, our courts should examine the appellant's attempts at compliance in order to determine whether an honest effort has been made to meet the requirements of the statute. Our overruling of the holding of Meta applies also to its progeny, e.g., Holmes v. Broodno, 222 Pa.Super. 478, 294 A.2d 903 (1972).

Lastly, returning to the facts of this case, we find no attempt at substantial compliance by appellant, but rather a situation where no semblance of compliance exists. The lower court, in view of this record, was correct in quashing the appeal.

Affirmed.

ITEM 1: SINGLE-CASE, DETERMINATE

Which one of the following statements best summarizes the *reasoning* in the *Black and Brown* decision?

a) The court concludes that the appellant's appeal from an adverse arbitration decision should be quashed because while the defendant's payment in *Meta* was only $10 short, here the appellant was short by a much larger amount, $102, and therefore the appellant's failure to meet the statute cannot be excused as *de minimis*.

b) The court concludes that the appellant's appeal from an adverse arbitration decision should be quashed because the appellant dismissed its trial counsel without notice to the appellee.

c) The court concludes that the appellant's appeal from an adverse arbitration decision should be quashed because the appellant failed to notify its new counsel as to the amount of record costs owed.

d) ***The court concludes that the appellant's appeal from an adverse arbitration decision should be quashed because the appellant paid none of the record costs owed by the statutory deadline and because the *Meta* decision, if allowed to stand, would allow defendants to simply ignore the prepayment requirement altogether.

e) The court concludes that the appellant's appeal from an adverse arbitration decision should be quashed because the amount in controversy in the original arbitration hearing was not sufficiently large for the record cost requirement to be waived as *de minimis*.

ITEM 2: SINGLE-CASE, DETERMINATE

Which one of the following statements best summarizes the court's *reasoning* in *Mackey*?

a) ***Citing *Black and Brown*, the court concludes that Mackey's attempts to learn what he owed and to make his payment on time are neither "valid" nor "honest." The court states that Mackey should have known the procedural rules and acted upon them, but otherwise offers little reasoning concerning why Mackey's attempts to learn his costs and the correct due date are neither "valid" nor "honest."

b) The court reasons that if the rule requiring payment of opposing counsel's record costs were relaxed, litigants could not learn when no further appeals may be taken by the losing party in arbitration.

c) Because Mackey did not produce physical evidence of timely mailing, Mackey's appeal cannot be perfected and must be quashed.

d) The court reasons that the fact that the Prothonotary's office put Mackey on hold is irrelevant since Mackey could have easily continued trying to reach them within the statutory time limit.

e) Following *Black and Brown*, the court reasons that Mackey's attempts to learn what he owed and to make his payment on time are neither "valid" nor "honest" because Mackey tries to have it both ways: he argues that he could not read the pencil note, and also that he mistakenly juxtaposed the numbers. The court thus finds this inconsistent argument less than honest.

ITEM 3: SINGLE-CASE, DETERMINATE

Which one of the following statements best summarizes the *reasoning* in the minority opinion (dissent) in *Meta*?

a) Citing the *Smith* case, the minority dissents because it sees no reason to question legislative acts that have remained unchanged for over 100 years.

b) The minority dissents because it claims that courts are supposed to follow strictly the legislature's wishes and only usurp the legislature's function when they have good reason.

c) ***The minority dissents for the following reasons. It argues that, in order to vacate the statute, the majority should have made (but failed to make) an argument that the record cost statute is unreasonable, and it argues that the defendant had a responsibility to learn the correct amount of costs.

d) The minority dissents because the majority cited *Beth-Allen Sales v. Hartford Insurance Group* when it should have cited the earlier *Smith* and *Fleisher v. Kaufman* cases instead; the dissent claims that the majority failed to discuss these two cases at all.

e) The minority dissents because they perceive that the principle of the sufficiency of substantial compliance would render the finality of adjudication and efficient processing of disputes impossible, since "substantial compliance" is, at best, a vague standard and would allow too much subjectivity in interpretation.

ITEM 4: SINGLE-CASE, DETERMINATE

Which one of the following questions best expresses the *legal issue* that the *Mackey* court sees itself as addressing?

a) Is it within the court's discretion to decide when the difference between stated requirements for making an appeal from arbitration and a defendant's partial compliance with those requirements can be disregarded as a mere technicality?

b) Under Pennsylvania rules, what reasonable conditions may a court impose upon the right of appeal from arbitration?

c) ***Should a defendant's nonpayment of record costs, coupled with an attempt to pay them two days late, prevent a defendant from making an appeal from an adverse arbitration decision under Pennsylvania rules?

d) Under Pennsylvania rules, may defendants who find themselves in the following factual situation be allowed to appeal an adverse arbitration decision:

- Received an unclear communication from the plaintiff concerning the amount owed and the due-date for paying record costs; and
- Tried but failed to reach the Prothonotary's office to learn the exact amount of record costs owed; and
- Paid $237 of the $327 due for record costs two days late; and
- Met all other procedural and monetary requirements for appeal?

e) Should a defendant's partial payment of record costs be considered more or less important than payment of other costs, such as fees for the arbitrators and security?

ITEM 5: CROSS-CASE, DETERMINATE

Assume for the moment that you decide to submit a motion to the Common Pleas Court to reconsider its decision against your client, Mackey Plumbing Co. Which one of the following statements best summarizes the similarities and differences between the facts in *Mackey* and the facts in *Meta*?

a) Except for the different amounts of record costs involved, the fact patterns in these two cases are identical: both defendants tried and failed to contact the Prothonotary's office; both paid their record costs late; and both failed to pay the full amount that they owed.

b) The only factual similarity of any importance is that both defendants failed to comply with the record cost statute. The key difference is that the defendant in *Meta* was allowed to proceed with an appeal.

c) The *Mackey* case is distinguishable from *Meta* because Mr. Mackey paid about 72% of what he owed, whereas Yellow Cab paid only 57% the amount owed. The cases are similar in that both defendants argued that they had to deal with apparently deliberate mis-statements from the plaintiffs when trying to learn the true amount of costs owed.

d) The relevant similarity between the facts of these cases is that both defendants violated the statute because both defendants failed to pay the full amount, while the relevant difference is the large gap between what Mackey owed ($90) and what Yellow Cab owed ($7.50).

e) ***The relevant factual similarity is that the payments of both defendants fell short of the amount owed. However, a comparison also shows an important difference in that Mackey paid most ($237 of $327) of the full amount that he owed two days late, whereas Yellow Cab paid just over half ($10 of $17.50) of the amount owed on time.

ITEM 6: CROSS-CASE, DETERMINATE

Which one of the following questions best summarizes the legal issue raised by *all three* of the cases you have read?

a) Under what circumstances may losing parties in Pennsylvania arbitration decisions validly rely upon the record cost amounts provided to them by opposing counsel rather than the amount officially recorded?

b) Under what circumstances may a Pennsylvania court declare a party's failure to comply with procedural statutes for perfecting an appeal to be *de minimis* and disregarded?

c) ***In what ways, if any, may a party fail to comply with the procedural statute requiring a losing party in arbitration to "pay all costs that have accrued on such suit or action … within 20 days after the entry of the award of the arbitrators on the docket" and yet proceed with an appeal?

d) When attempting to appeal an adverse arbitration decision to a regular appellate court in Pennsylvania, what portion of the total record cost owed to opposing counsel must a losing party pay?

e) Under what conditions may a court decide to void a statute because it has clearly put form over substance or has otherwise compromised justice by elevating mere technical detail?

ITEM 7: CROSS-CASE, DETERMINATE

Which one of the following statements best summarizes the similarities and differences between the courts' reasoning in *Meta* and the court's reasoning in *Black and Brown*?

a) The reasoning used in the two cases is largely similar: whereas the *Meta* court reasons that striking down the record cost requirement will improve the administration of justice by restoring the supremacy of substance over form, the *Black and Brown* court also recognizes the harshness of denying a party the right of appeal for accidentally underpaying the amount of record costs owed to opposing counsel.

b) The reasoning used by these two courts is different: whereas the *Meta* court reasons that it is following the principle of the sufficiency of substantial compliance invoked in earlier cases, the *Black and Brown* court explicitly reasons that the right of appeal is a substantial right which is subject to reasonable conditions.

c) ***The reasoning used by the two courts differs: although both courts refer to the principle of the sufficiency of substantial compliance in their decisions, the *Meta* court further reasons that the record cost requirement is a matter *de minimis* because in comparison with the amount of money in controversy in arbitration decisions, record costs are usually quite small. The *Black and Brown* court does not discuss this disparity in amounts involved.

d) The reasoning used by the two courts is the same: both courts agree that the record cost prepayment requirement serves no real function as a condition of appeal and that record costs in comparison with dollars in controversy make the latter costs *de minimis*.

e) The reasoning used by the two courts differs: while both courts invoke the principle of the sufficiency of substantial compliance as a rationale for their decisions, the *Meta* court further argues that the requirement is a legal trap that "puts counsel in a demeaning position, creates bad public relations, and tends to cause additional malpractice litigation."

ITEM 8: CROSS-CASE, DETERMINATE

As an advocate representing Mackey Plumbing Co., you may want to find out more about the constitutionality of the record cost statute. As far as this constitutionality issue is concerned, which one of the following statements seems most accurate, based upon the three cases you have read?

a) None of these three cases makes any comment, one way or another, about the constitutional status of this requirement. No case cited or discussed in any of these three cases appears to have raised the constitutionality of the statute as an issue on appeal, and thus it appears likely that the constitutionality of the statute has never been challenged previously.

b) The *Mackey* case is the only case among the three in which the defendant directly challenged the statute on constitutional grounds, and is the only case of the three that directly answers the question as to whether the prepayment statute is constitutional.

c) The *Meta* case is the only case among these three in which the defendant directly challenges the constitutionality of the record cost statute, as can be seen in this court's complaint that "The requirement of payment to counsel for the other side of a relatively small amount with all kinds of variations as to what that 'small' amount is, serves no real function as a condition of appeal and can raise a serious question of constitutionality."

d) ***The record cost procedural statute appears to have been directly challenged on constitutional grounds at least once (as indicated in the majority opinion in *Meta*) in the "ancient" *McDonald v. Schell* case of 1820. However, there is no direct challenge presented to the statute's constitutionality in any of these three cases (*Meta*, *Black and Brown*, and *Mackey*) and, at least on their basis, the statute remains constitutional.

e) The *Black and Brown* case is the only case among these three in which a defendant directly challenges the statute on constitutional grounds, by basing an appeal upon *Meta*. Further, the *Black and Brown* court rejected the *Meta* court's argument that the statute is unconstitutional by arguing that the *Meta* court never demonstrated the unreasonableness of the statute.

ITEM 9: SINGLE-CASE, INDETERMINATE

Courts sometimes use phrasing that leaves room for interpretation by other courts. However, while an opinion may contain a number of statements that are ambiguous, some of these ambiguities may be more significant than others. Assuming you decide to appeal the *Mackey* decision to Superior Court, which one of the following ambiguities presented by the *Black and Brown* opinion is most relevant to you?

a) The *Black and Brown* court speaks both of a defendant's "valid attempt" and of a defendant's "honest effort" to meet the requirement as possible mitigating circumstances. Assuming that the court uses these two phrases to mean different things, would a defendant who failed to comply with the statute but still wished to appeal need to show that (i) she made both a "valid attempt" and an "honest effort," or, (ii) would showing that she satisfied one or the other of these elements suffice to prevent her appeal from being quashed?

b) Why doesn't the court more precisely indicate what statements are to be included as part of its holding? For instance, consider the court's statement that "courts should examine the appellant's attempts at compliance in order to determine whether an honest effort has been made to meet the requirements of the statute." Is this statement to be considered dicta, or is it considered part of the holding?

c) Why does the *Black and Brown* court in its holding speak both of a "valid attempt to make … timely and full payment" and of "substantial though incomplete compliance with the requirement" as possible mitigating factors in a defendant's behavior? Does the court mean that a "valid attempt to make … timely and full payment" is the same thing as "substantial though incomplete compliance with the requirement;" or, does the court, by using these two different phrases, mean to imply substantially different standards for forgiving a defendant's non-compliance?

d) ***Why does the *Black and Brown* court say in the first part of its holding that "the requirement that record costs be paid during the [20 day] appeal period is mandatory," and then add the "caveat that a valid attempt to make such a timely and full payment, coupled with substantial though incomplete compliance with the requirement should not result in the harsh finality of quashing an appeal …."? Does the court's combination of these two statements mean that (i) the defendant must still satisfy the statutory requirement by making at least a partial payment to the plaintiff within the 20 day limit or else the defendant's appeal will be quashed without further examination of the facts; or, (ii) that the defendant's appeal may not be quashed without first examining the circumstances to see if the defendant made a "valid attempt" at timely payment?

e) Why does the *Black and Brown* court speak both of a defendant's "valid attempt" and of a defendant's "honest effort" to meet the requirement as possible mitigating circumstances? For instance, is a "valid attempt" the same thing as an "honest effort," or does the court, by using these two different terms, mean to imply substantially different tests for forgiving a defendant's non-compliance?

ITEM 10: SINGLE-CASE, INDETERMINATE

All opinions are interpretations, and interpretations invite questioning. Assuming that you decide to appeal the *Mackey* decision to Superior Court, which one of the following questions presented by the *Mackey* opinion is most relevant to this task?

a) Why does the court state that the issue in the case is whether an appeal may be quashed for "nonpayment" when the facts show that Mr. Mackey partially paid the amount owed, albeit two days late?

b) Why does the *Mackey* court say that Mackey's shortfall in payment constitutes "more than a technicality"?

c) Why does the court rely on its own "inspection" of the pencil note rather than calling for or inviting an expert witness to offer testimony concerning its legibility?

d) Why doesn't the *Mackey* court reprimand the plaintiff for communicating the wrong due date for the record costs?

e) ***Why does the court note each of Mackey's claims about his attempt to contact the Prothonotary's office yet offer no reasons for concluding that Mr. Mackey did not make either an "honest effort" or "a valid attempt to make … full payment"?

ITEM 11: SINGLE-CASE, INDETERMINATE

Assume for the moment that you decide to submit a motion to the Common Pleas Court to reconsider its decisions against your client, Mackey Plumbing Co. Which one of the following "unknowns" in the facts of the *Mackey* case would be most relevant to your motion?

a) Did Mackey have counsel, or was he acting *pro se*?

b) Does Mackey still have a copy of the dated fax he sent to the plaintiff?

c) Has Mackey Plumbing Co. ever previously lost an arbitration hearing and subsequently failed to pay record costs?

d) Does Mackey have some visual impairment?

e) ***Is there a way of finding out whether the plaintiff's communication was deliberately versus accidentally misleading?

ITEM 12: CROSS-CASE, INDETERMINATE

Because the *Mackey* case was decided in a lower level court (Court of Common Pleas), the judge in that case would need to examine decisions in higher appellate courts to learn the applicable law. In Pennsylvania this court would be Superior Court. Given this circumstance, which one of the following questions about the relationship between the *Mackey* decision and the other two Superior Court decisions (*Meta* and *Black and Brown*) is most important for you to think about?

a) Is it possible that the *Mackey* court cited *Meta* and *Black and Brown* as it did in order to deliberately provoke an appeal challenging the record cost statute as unconstitutional?

b) Does the *Mackey* court's use of *Meta* and *Black and Brown* suggest that a rule is needed forbidding plaintiffs from communicating their record costs directly to defendants?

c) Can the *Mackey* court properly conclude from the decisions in either *Meta* or *Black and Brown* that "it serves a meaningful purpose to know when a matter has finality"?

d) ***Can the *Mackey* court, without qualification, properly conclude from the decisions in either *Meta* or *Black and Brown* that record costs must be paid in 20 days?

e) Can the *Mackey* court properly conclude that "it is within the court's discretion to decide what reasonable conditions can be imposed" upon the right of appeal?

ITEM 13: CROSS-CASE, INDETERMINATE

Most case analysis takes place in a specific context, with a specific problem in view. Assuming you decide to appeal the Mackey decision to Superior Court, which one of the following questions about the *Mackey* court's reliance on the *Black and Brown* decision is most relevant?

a) Why does the *Mackey* court apparently misstate the amount of costs that the defendant owed in *Black and Brown*: the *Black and Brown* court states that defendant owed $102, whereas the *Mackey* court states the amount the defendant owed there was $74?

b) ***Why does the *Mackey* court only quote the part of the holding in *Black and Brown* that requires "a valid attempt to make … full payment"?

c) Why does the *Mackey* court express its holding by applying the principle of *de minimis*, given that there is no mention of *de minimis* in the *Black and Brown* holding or reasoning?

d) Why doesn't the *Mackey* court discuss *Beth-Allen Sales Co. v. Hartford Insurance Group* (1970), a case which the *Meta* court discussed as endorsing the principle of the sufficiency of substantial compliance?

e) Why doesn't the *Mackey* court more fully review the procedural facts in the *Black and Brown* case?

ITEM 14: CROSS-CASE, INDETERMINATE

An advocate must routinely think through the way opposing counsel may analyze the law. Assuming that you have decided to appeal the Common Pleas Court decision (in *Mackey*) to Superior Court, which one of the following questions about opposing counsel's approach to the *Mackey* decision's relation to precedent is most relevant to you?

a) How might opposing counsel compare Mackey's situation with Yellow Cab's situation (in *Meta*)?

b) ***How might opposing counsel compare Mackey's situation with the defendant's situation in *Black and Brown*?

c) How might opposing counsel make use of the dissent in the *Meta* decision?

d) To what extent will opposing counsel consider Mackey's compliance with other procedural requirements for appeal (e.g., such as payment of the arbitrators' fee) in reviewing the Common Pleas Court decision quashing his appeal?

e) Given that the *Black and Brown* court stressed that the appellant in that case had "express notice" that record costs be paid within the 20 day limit, how likely is it that opposing counsel will argue that Mackey similarly had "express notice" that he had to pay his record costs?

Appendix 2: TV2. Cases and Items Used in Field-Testing

JUDD **ALEXANDER et al.** vs. **PRIMERICA HOLDINGS, INC.**, formerly known as PRIMERICA CORPORATION

No. 91-5712

UNITED STATES COURT OF APPEALS FOR THE THIRD CIRCUIT

September 8, 1991, Argued
October 28, 1991, Filed

PRIOR HISTORY: Appeal from the United States District Court for the District of New Jersey. (D.C. Civil No. 89-05151)

DISPOSITION: Reverse and remand.

JUDGES: Before: STAPLETON and MANSMANN, Circuit Judges and POLLAK, District Judge.

Honorable Louis H. Pollak of the United States District Court for the Eastern District of Pennsylvania, sitting by designation.

OPINION BY: MANSMANN

OPINION: OPINION OF THE COURT

MANSMANN, Circuit Judge.

At issue here is whether a particular clause in an ERISA summary plan description reserved a company's right to reduce employee welfare benefits. The district court opined that it did. Because we conclude that the clause is ambiguous, we will reverse the district court's grant of summary judgment, and we will remand the case for a determination by the district court, as fact finder, as to whether the company promised lifetime fixed-cost benefits to its retirees.

I.

The plaintiffs, Judd Alexander and Richard Edwards, are retired salaried employees of American Can Company. As retirees, they received health and life insurance from the American Can Salaried Retirees Group Insurance Plan. In this class action, they contest a tenfold increase in their monthly premiums, instituted by Primerica Holdings, Inc., a successor in interest to American Can. In essence, the retirees claim that the increase violated the terms of the Plan. Primerica argues that the Plan reserved the right to reduce benefits and, therefore, that the Plan did not promise any vested benefit.

Because a formal plan document does not exist, our knowledge of the Plan arises from three summary plan descriptions. These summary plan descriptions provide that, upon retirement, salaried employees and their spouses will receive the Plan's Basic Medical coverage and that the death of a retiree does not affect a spouse's right to benefits. The Plan's Major Medical Benefits will terminate, however, if a retiree ceases to contribute to the Plan.

The following section of one summary plan description figures prominently in this litigation:

> **Extent and Limit of Coverage**
>
> The Company expects to continue this Plan indefinitely, but necessarily reserves the right to amend, modify, or discontinue the Plan in the future in conformity with applicable legislation. The Plan does not provide for benefit payments in any case or under any condition not identified and provided for in this booklet. The Group Insurance Policy and the certificates thereunder issued by the Insurance Company are consistent with the terms and conditions outlined in this booklet.
>
> If any changes become effective under the Medicare Program by reason of current or future Governmental legislation or regulations, then full consideration will be given to appropriate modification in this Retired Group Insurance Plan.

This section appeared with a slight variation in one other summary plan description. Primerica argues that the amendment clause in the section's first sentence reserves an unqualified right to reduce or terminate benefits. To the contrary, the retirees argue that it reserves only a limited right.

In 1989, shortly after it acquired American Can, Primerica raised monthly premiums from $5 to $50 per person. The retirees then filed this class action, claiming that the increase violated the terms of the Plan and that the consequent reduction of vested benefits violated ERISA.

Primerica, in a motion for summary judgment, argued that the summary plan descriptions notified the employees of the Plan's absolute right to reduce or terminate benefits, and that therefore the Plan did not promise vested benefits. In response, the retirees argued that the summary plan descriptions only notified employees that the Plan would necessarily change if required by law. The district court held that the clause was unambiguous, granted summary judgment for the defendants, and dismissed the complaint. The retirees appealed.

II.

The district court held that because the amendment clause unambiguously reserved the right to the Plan to reduce or discontinue benefits, the Plan did not promise lifetime benefits. Whether an ERISA plan is ambiguous is a question of law. *Taylor v. Continental Group*, 933 F.2d 1227, 1232 (3d Cir. 1991). We review the district court's decision accordingly.

A.

The contested amendment clause, which is at the center of this appeal, reads:

> The Company [American Can] expects to continue this plan indefinitely, but necessarily reserves the right to amend, modify, or discontinue the Plan in the future in conformity with applicable legislation.

The defendants suggest that this clause means that American Can might modify the Plan at any time and for any reason. The retirees suggest that the clause means that American Can will only modify the Plan if necessary to conform with applicable legislation. We find that the clause is ambiguous, that is, "subject to reasonable alternative interpretations." *Taylor*, 933 F.2d at 1232.

For instance, the word "necessarily" might mean "understandably," as the defendants proposed at oral argument, but the word might also mean "as a necessary result or consequence." See *Webster's Third New International Dictionary* (unabridged), at 1510 (1981). The phrase "in conformity with applicable legislation" may either limit the right to modify (as the retirees urge) or notify employees that any change will conform to the law (as the defendants urge). The word "legislation" may imply "change in the law," which would support the retirees' position, but perhaps it simply means "law," which would support the defendants' position. Similarly, the word "indefinitely" either means "continuing without limit" or "undetermined."

The variant of the amendment clause does not eliminate the ambiguity. It provides:

> American expects to continue this Plan indefinitely, but necessarily reserves the right to amend, modify, or discontinue the Plan in conformity with applicable legislation and also subject to any applicable collective bargaining agreement.

The additional words, "subject to … " do not elucidate the clause's meaning.

Nor is the ambiguity eliminated by the Medicare clause, which assures participants that, if Medicare law changes, "full consideration will be given to appropriate modification" of the Plan. Interpreted one way, the Medicare clause conforms to the retirees' position; it reassures the employees that although American Can would reduce benefits if the law required, the company might also increase them if Medicare law created shortfalls. Interpreted another way, the Medicare clause conforms to the defendants' position; it reassures that American Can will give the retirees' needs "full consideration," even without the obligation to maintain benefits.

We also note that ERISA plans, like contracts, are to be construed as a whole, see *Hansen v. Continental Ins. Co.*, 940 F.2d 971, 981 (5th Cir. 1991), and context colors the meaning of words. Empirical research has demonstrated that a summary not unlike this one can lead readers to conclude that benefits are for life. See F. Namtarts, Contract Disclaimers in ERISA Summary Plan Documents: A Deceptive Practice? 10 *Indus. Rel. L.J.* 350 (1988).

Of the arguments advanced by both sides, none persuades us that the amendment clause has only one meaning. For example, the defendants have asserted that American Can could have written: "The Plan may only be amended to conform to applicable changes in legislation." By the same logic, the retirees have argued that American Can could have written: "The Plan may be amended at any time." This argument shows only that the clause, if drafted differently, would have been unambiguous.

The defendants have questioned the implication in the retirees' interpretation that legislation might result in a benefit reduction. But, in fact, one can easily imagine legislation that might result in a reduction of benefits. For example, a tax on benefits would reduce them; nationalization of health care might eliminate private insurance altogether in an effort to contain costs.

Additionally, the defendants have made much of evidence that indicates that American Can increased benefits on one occasion, with a slight increase in premiums from about $3 to $5 per month. We reject the defendants' argument that the company's past practice of changing benefits somehow renders the amendment clause unambiguous. As we explain below, a benefit increase does not necessarily contradict the retirees' position.

Summary plan descriptions must warn employees of adversity. See 29 U.S.C. § 1022. Whether the clause reserved a limited or unlimited right, it addressed the right to reduce benefits. A lawyer reading the words "necessarily reserves the right to amend, modify or discontinue the Plan" might give the emphasized words their broadest meaning: enhance or reduce. It is unclear, however, why a company would reserve the "right" to enhance benefits. Neither ERISA nor common sense requires notice of future benefit enhancements. The average Plan participant, under the spell of common sense, would understand the clause to reserve either a limited or unlimited right to reduce benefits. An increase in benefits therefore sheds no light on the meaning of the clause.

A prior record of benefit reductions, however, would show action consistent with the defendants' position that American Can had reserved the right to reduce or discontinue benefits at will. The district court, as the trier of fact, might reasonably infer that, because benefits had been reduced in the past, the Plan had not promised irreducible benefits. The defendants support their allegation of past reductions with a snippet of testimony, which culminates in the statement that the cost of benefits "changed by a matter of $ 1, a dollar and change a month in return for an offsetting increase in the protection, particularly on the catastrophic side." The testimony shows a cost increase that was minuscule in comparison to the one contested here, and which was accompanied by a benefit increase. According to the defendants' own submission, American Can explained the change as an increase in benefits, not a reduction, and the employees perceived it that way. On remand, it will be for the district court to determine if this evidence is sufficient to show that American Can always acted consistently with a reservation of authority to reduce benefits and increase premiums and the relevance thereof.

B.

We further conclude that the amendment clause is ambiguous in light of *Hamilton v. Air Jamaica*, 945 F.2d 74 (3rd Cir. 1991); *Alday v. Container Corp. of America*, 906 F.2d 660 (11th Cir. 1990), *cert. denied*, 112 L. Ed. 2d 668, 111 S. Ct. 675 (1991); *Musto v. American General Corp.*, 861 F.2d 897 (6th Cir. 1988), *cert. denied*, 490 U.S. 1020, 104 L. Ed. 2d 182, 109 S. Ct. 1745 (1989); and *Moore v. Metropolitan Life Ins. Co.*, 856 F.2d 488 (2d Cir. 1988). In *Hamilton*, plan documents reserved the right in individual cases and generally, "to alter, reduce or eliminate any pay practice, policy or benefit, in whole or in part, without notice." In *Alday*, plan documents reserved the right to "terminate, suspend, withdraw, amend or modify the plan in whole or in part at any time." 906 F.2d at 662. In *Musto*, the insurance policy in question provided:

> This Policy, or any insurance coverage hereunder, may be amended or discontinued at any time by an announcement duly published by the Company. The Company reserves the right to determine new premium contributions from time to time and at any time. Termination of this Policy, or any insurance coverage hereunder, or any amendments hereto shall not require the consent of any Employee or beneficiary, nor shall such action require individual notice to any such person. Only the President, a Vice-President, the Secretary or an Assistant Secretary has power on behalf of the Company to make or modify this contract of insurance. 861 F.2d at 901-02.

In *Moore*, various summary plan descriptions had reserved, since 1915, the right to change the plan's terms "at any time." 856 F.2d at 490. In each case, the court determined that the clause had reserved the right to reduce benefits.

But the American Can clause differs from those in *Hamilton*, *Alday*, *Musto*, and *Moore*. By way of example, only one of many summaries in *Moore* did not contain a reservation clause, and that summary referred prominently to the plan

document itself. *Moore*, 856 F.2d at 490. In contrast, the American Can plan document does not exist, and there are merely three summaries, of which only two contain the clause. Also, the phrase "at any time," which appears in the *Alday*, *Musto*, and *Moore* documents, does not appear in the American Can documents. In short, the *Alday*, *Musto*, and *Moore* plans used clauses more straightforward than the one we consider here.

In opining that the American Can clause was unambiguous, the district court equated it with two phrases from *Musto* and *Moore*. These two phrases, however, do not sufficiently resemble the American Can clause. A slight similarity does not warrant construing a different phrase to be unambiguous as a matter of law.

C.

The district court granted the defendants' motion for summary judgment on the theory that an unqualified reservation of the right to reduce benefits was inconsistent with a promise of vested, lifetime benefits. Because the summary plan descriptions do not clearly reserve the right to reduce benefits, we will reverse the district court's grant of summary judgment dismissing the complaint. We will remand for the district court's interpretation of the summary plan descriptions in light of all relevant evidence and for the district court's further consideration of the retirees' claims.

Joan Bridell, Plaintiff, v. RIBIER CLOTHIERS, INC., et al., Defendants

Civil Action No. H-89-1995

UNITED STATES DISTRICT COURT FOR THE SOUTHERN DISTRICT OF TEXAS, HOUSTON DIVISION

742 F. Supp. 392; 1990 U.S. Dist. LEXIS 11761; 12 Employee Benefits Cas. (BNA) 1838

November 4, 1991, Decided

DISPOSITION: GRANTED.

JUDGES: Norman W. Black, United States District Judge.

OPINION BY: BLACK

OPINION: ORDER

NORMAN W. BLACK, UNITED STATES DISTRICT JUDGE

Pending before the Court is a motion for summary judgment filed by Defendants Ribier Clothiers, Inc., Brighter Days Design Company and National American Life Insurance Company. After consideration of said motion and the response filed by Plaintiff Joan Bridell, the Court is of the opinion that there is no genuine issue of material fact in this case, and that, for the reasons discussed below, Defendants' motion should be granted.

Statement of the Case

Joan Bridell brought this claim of discrimination and violation of her rights under ERISA against her employer, Ribier Clothiers, Inc., Brighter Days Design Company and the insurance company that administers her medical plan, National American Life Insurance Company. She alleges that because Defendants changed portions of the group medical care plan they violated ERISA, 29 U.S.C. § 1001 *et seq*.

Bridell began working for Defendant Ribier Clothiers, Inc. in 1982. In December of 1987 she was diagnosed as having AIDS and informed Defendant of her diagnosis. In July of 1988 Defendants made changes in their health insurance program. Previously employees that were diagnosed with AIDS were entitled to maximum benefits of $1,000,000.00 during their lifetime. The new plan decreased AIDS benefits to a maximum of $ 5,000.00 and also discontinued benefits for chemical dependency. In addition, in July 1988, the medical coverage changed from fully insured to self-insured. Bridell, after learning of the changes in the plan, filed suit.

Summary Judgment

Summary judgment is authorized if there is no genuine issue as to any material fact and the moving party is entitled to judgment as a matter of law. Fed. R. Civ. P. 56. The United States Supreme Court has interpreted this rule to mandate the entry of summary judgment after an adequate time for discovery against a party who fails to make a showing sufficient to establish the existence of an element essential to that party's case, and on which that party will bear the burden of proof at trial. *Celotex Corp. v. Catrett*, 477 U.S. 317, 91 L. Ed. 2d 265, 106 S. Ct. 2548 (1986).

Discussion

The paramount purpose of ERISA is to protect the solvency of employee benefit plans to ensure that the valid claims of employees and their beneficiaries will be paid. *Orchard Creek Hospital, et al. v. The Life Insurance Company of Virginia, et al*., Cause No. H-88-3500 (S.D. Tex. March 1990). ERISA sets forth uniform standards dealing with reporting, disclosure, and fiduciary responsibility. *Shaw v. Delta Air Lines, Inc*., 463 U.S. 85, 91, 77 L. Ed. 2d 490, 103 S. Ct. 2890 (1983). Furthermore, ERISA does not mandate that employers provide any particular benefits, and does not itself proscribe discrimination in the provision of employee benefits. *Id*.

The purpose of the changes in the group medical plan made by Defendants jibes with the ultimate purpose of ERISA: protection of the plan. Defendants made changes in the plan because in past years the plan suffered serious financial losses. Defendants were faced with either dropping the plan altogether or making changes. The alterations were not made to discriminate against Bridell or anyone who was diagnosed with AIDS.

Defendants had the right to make changes in the plan and complied with the requirement of ERISA. According to ERISA every employer must provide its employees with a summary plan description ("SPD"). 29 U.S.C. § 1022(a)(1). The SPD must be written in a manner calculated to be understood by the average plan participant. 29 U.S.C. § 1022(a)(1). Defendants followed these ERISA provisions and in the SPD that they handed out to their employees they included a footnote on the last page stating, "The Plan Sponsor may terminate or amend the Plan or terminate any benefit under the Plan *at any time* pursuant to relevant changes in federal legislation" (italics added). Therefore, beneficiaries were put on notice that from year to year their group medical plan could change.

Indeed, Bridell's allegations show no *promised* benefit, for there is nothing to indicate that defendants ever promised that the $1,000,000 coverage limit was permanent. Bridell has not and could not reasonably contend that the SPD's earlier reference to "lifetime medical benefits of up to $1,000,000" for eligible employees and their spouses somehow renders their reservation of rights in the disclaimer ambiguous. Moreover, there is no allegation or evidence that any other oral or written representations were made to Bridell that the $1,000,000 coverage limit would never be lowered. Defendants therefore broke no promise to Bridell. In a recent Third Circuit decision, a similar SPD disclaimer permitting a benefit plan administrator "to alter, reduce or eliminate any pay practice, policy, or benefit, in whole or part, without notice," was found to permit the defendant to sharply reduce the severance pay of an employee relative to what the company expressly stated elsewhere in the document. *See Hamilton v. Air Jamaica, Ltd.,* 945 F. 2d 74 (3d Cir. 1991).

In another similar case an insurance company completely canceled a group plan and one of the beneficiaries filed suit against the insurance company in an attempt to force it to provide insurance benefits. *Mackenzie v. Travelers Ins. Co*., 752 F.2d 1350 (8th Cir. 1985). The Court held that the insurance company had the right to terminate the plan. *Id*. The insurance company had complied with the ERISA requirements, and provided an SPD which was distributed to the employees. The SPD tracked the language of the plan document. The court in *Mackenzie* held that neither the employer nor the insurance company was under any contractual obligation to refrain from terminating the plan. *Id*. at 1351. Furthermore, the Court found that ERISA does not create liability on the part of an employer who changes the kind of health plan it provides to its employees where no contract prohibits or prevents such change. *Id*. at 1352. The contract provided by Defendants in the case before this Court does not prohibit or prevent any changes from being made in the group plan. Just as the Defendants can make the benefits larger for AIDS patients they can also make the benefits smaller.

Finally, the Plaintiff claims that the motion for summary judgment should be denied because there is evidence of discrimination. Although an analysis of Plaintiff's claim of discrimination is not necessary, even using this analysis Plaintiff's claim fails.

In a Third Circuit case the court set out three requirements to prove discrimination under ERISA. *Gavalik v. Continental Can Co*., 812 F.2d 834 (3rd Cir. 1987). An employee must demonstrate (1) prohibited employer conduct, (2) taken for the purpose of interfering (3) with the attainment of any right to which the employee may become entitled. *Id*. at 852. Plaintiff cannot prove that Defendants' conduct was prohibited. The Defendants had the right to change or terminate the

medical plan under ERISA. In addition, the action was not taken to interfere with the rights of Bridell but to ensure the future existence of the plan.

Bridell was not entitled to health benefits whose terms never change. She was not denied any compensation for expenses related to treatment for AIDS during 1987 when the plan paid up to $1,000,000 for AIDS patients, nor was she denied medical compensation for insurance claims once Defendants changed the health plan in 1988. Therefore, Bridell's claim of discrimination must fail.

Since Defendants have complied with the regulations of ERISA, legally have the right to make changes in its group medical plan and did not act in a discriminating manner, Plaintiff's claim must fail.

For the reasons stated above, it is therefore ORDERED that Defendants' motion for summary judgment is GRANTED.

HAMILTON, Robert L., Appellee v. **AIR JAMAICA, LTD.**, Appellant

No. 90-1933

UNITED STATES COURT OF APPEALS FOR THE THIRD CIRCUIT

April 15, 1991, Argued
September 30, 1991, Filed

PRIOR HISTORY:

On Appeal from the United States District Court for the Eastern District of Pennsylvania; D.C. Civil Action No. 89-06133.

DISPOSITION: Reversed.

JUDGES: Stapleton, Greenberg, and Higginbotham, Circuit Judges.

OPINION BY: STAPLETON

OPINION: OPINION OF THE COURT

STAPLETON, Circuit Judge

Air Jamaica, Ltd. appeals from a judgment in favor of its former employee, Robert Hamilton, for additional severance pay in accordance with the terms printed in Air Jamaica's employee handbook ("the Handbook"). This case requires us to determine whether the Handbook is an ERISA plan and, if so, whether Air Jamaica's reservation of the right to determine non-pension employee benefits at the time they accrue is consistent with ERISA. We conclude that the Handbook is an ERISA plan and hold that Air Jamaica's reservation of the right to determine benefits as they accrue is consistent with ERISA and precludes Hamilton's claim. Accordingly, we will reverse the judgment for Hamilton.

I.

The undisputed facts as stated by the district court are as follows:

1. Plaintiff began working for defendant, Air Jamaica, Ltd., as a sales representative in August, 1976. Plaintiff was promoted to account executive in 1985 and was working in that position in Philadelphia when Air Jamaica terminated his employment on May 31, 1989.

2. Air Jamaica terminated plaintiff's position as part of a corporate reorganization brought on by budgetary problems.

3. At the time of his termination, plaintiff was an "exempt" employee, as defined on page 2 of the "Policies and Procedures" section of Air Jamaica's Employee Handbook.

4. Air Jamaica distributed the Handbook—a written document outlining the components of Air Jamaica's employee

benefits program, including vacation benefits and separation pay benefits—to all of its employees early in 1989. Plaintiff received his copy of the Handbook early in March, 1989.

5. The Handbook contains a description of a policy to pay employees' separation pay. The Handbook states that "if termination occurs due to job elimination or reorganization, a non-exempt employee will receive 2 weeks a year for each full year worked and an exempt employee will receive four weeks for each full year worked."

6. The separation pay policy described in the handbook was not the policy that Air Jamaica intended to be included therein. Air Jamaica has a longstanding separation pay practice, as evidenced by its conduct prior to and after plaintiff's termination, to pay separation pay in the amount of two weeks pay for each full year of employment, with a maximum payment of three months.

7. Air Jamaica retained the consulting firm of Foster-Higgins, to help prepare the Handbook. The incorrect statement of Air Jamaica's separation pay policy was written into the Handbook by Linden Dunn, a Foster-Higgins employee, during production at Foster-Higgins. Air Jamaica failed to discover the mistake during proofreading of the Handbook.

8. Air Jamaica's preparation of the Handbook, including supervision over and copy-editing of its contents, was frustrated by a tight deadline, personal tragedy and natural disaster. The officer who spearheaded the project was murdered in the summer of 1988 in the course of a labor dispute with Air Jamaica pilots. His assistant died in the same week of natural causes. In September, 1988, Jamaica was ravaged by Hurricane Gilbert, shutting down Air Jamaica operations for a full two weeks and putting the Handbook project on hold until late in the year. During the entire course of the project, Air Jamaica officers were preoccupied with the company's medical benefits package, which was undergoing change, and a desire to release the Handbook concurrent with this change at the beginning of 1989.

9. Air Jamaica discovered the mistake in the Handbook regarding Air Jamaica's separation pay policy after the decision to terminate plaintiff's employment. Mr. Simpson first became aware of the error on May 25, 1989, while in the Philadelphia office to personally inform plaintiff of the termination.

10. Mr. Simpson gave plaintiff a letter on May 25, 1989, memorializing the terms of plaintiff's termination, to become effective on May 31, 1989. The letter indicated that plaintiff would be paid a "terminal gratuity" based on the formula of "two weeks pay for each year of employment to a maximum of three months' payment".

11. Air Jamaica sent a memorandum, dated May 30, 1989, to all affected employees, advising them that there was an error in the Handbook's description of the separation pay policy. The memorandum advised that the Handbook should state that employees terminated due to job elimination or reorganization will receive "a terminal gratuity based on two weeks' basic pay per year for each full year worked up to maximum of three months' pay."

12. Plaintiff received the errata memorandum on May 30, 1989.

13. Air Jamaica paid defendant a terminal gratuity amounting to three months pay.

14. Plaintiff made a demand to Air Jamaica on June 3 requesting a severance payment in accord with the policy described in the Handbook. By letter of July 14, Air Jamaica indicated that its policy is to pay severance in the amount of two weeks per year worked with a maximum payment of three months.

15. The Handbook contains a disclaimer, whereby Air Jamaica reserved the right in individual cases and generally, "to alter, reduce or eliminate any pay practice, policy or benefit, in whole or in part, without notice." The Handbook further contains a "statement of understanding", which employees are instructed to date, sign and return to their supervisors, indicating that they have reviewed the policies and procedures section of the Handbook and understood the disclaimer contained therein.

16. Plaintiff did not sign and return the statement of understanding.

In view of these facts, the district court held that, although the severance policy printed in the manual appeared to be a printer's error, under the Employee Retirement Income Security Act of 1974, § 402(b)(3), 29 U.S.C.A. § 1102(b)(3) ("ERISA") the employer was bound by that policy. *Hamilton v. Air Jamaica, Ltd.*, 750 F. Supp. 1259, 1267 (E.D. Pa. 1990). It held that a written plan must control over an unwritten plan and therefore concluded that the terms set forth in the Handbook were Air Jamaica's severance plan under ERISA.

The district court then considered whether Air Jamaica's reservation of the right to determine benefits on a case by case

basis barred Hamilton's claim, but held that giving full effect to such a provision would be inconsistent with ERISA. The court further held that Air Jamaica, through the same explicit reservation, had reserved its right to amend the plan. Nevertheless, the court concluded that the amendment here—one day before Hamilton stopped working–was untimely and that recognizing it would frustrate ERISA's policy of protecting employees' legitimate expectations. The district court therefore entered judgment for Hamilton on his claim for severance pay under ERISA.

This court has jurisdiction over the final judgment of the district court pursuant to 28 U.S.C. § 1291. The district court had jurisdiction over the ERISA claims under 28 U.S.C. § 1331 and 29 U.S.C. § 1132(f) and had pendent jurisdiction over the state law claims. The proper application of ERISA to the undisputed facts is a question of law over which we exercise plenary review.

II.

Air Jamaica contends that it is not bound by the severance pay provisions of the Handbook which were included only because of a printer's error. Instead, Air Jamaica argues it is bound only by its "actual" *un*written severance pay policy that it consistently followed and of which its employees were aware. We join the district court in rejecting that argument. The district court correctly held that "where an employer has published a plan document containing representation of intent to provide benefits, the employer cannot rely on [unwritten] evidence to the contrary", 750 F. Supp. at 1267, and that, in the absence of any other written description of benefits, the Handbook constituted Air Jamaica's ERISA plan.

In reaching that conclusion, the district court relied heavily on *Hozier v. Midwest Fasteners, Inc.*, 908 F.2d 1155 (3d Cir. 1990), in which this court joined six other courts of appeals that refuse to recognize oral modifications to ERISA plans. *Hozier* strictly interpreted 29 U.S.C. § 1102(a)(1)'s requirement that "every employee benefit plan shall be established and maintained pursuant to a written instrument." Indeed, all the cases decided before *Hozier* had rejected oral modifications that increased benefits, and the district court concluded that the policies behind ERISA made it even more important to disallow oral modifications that decrease benefits. *Id.* at 1163.

Hozier held that the requirement of 29 U.S.C. § 1102(a)(1) that all ERISA plans be written prevents an unwritten amendment from being an enforceable part of an ERISA plan. Therefore, it follows that Air Jamaica's unwritten severance pay policy could not be its ERISA plan and that the Handbook was its ERISA plan.

III.

Having determined that the Handbook constituted an ERISA plan, the district court was forced to decide whether Air Jamaica's printed disclaimer reserving the right to determine benefits on a case-by-case basis prevented Hamilton from having an enforceable claim to the severance benefits described in the Handbook. The district court concluded that Air Jamaica's reservation should not be given effect because it conflicted with ERISA:

> Read literally, the Air Jamaica disclaimer could be taken to mean that Air Jamaica is under no obligation at any time to award benefits in accord with its written promises. To the contrary, Congress enacted ERISA to provide security in benefits packages. *Hozier* stands for the rule that plan participants are entitled to rely, and to continue to rely, on an employer's current writing until those writings are in fact amended. 750 F. Supp.at 1270.

While we agree that an employee is entitled to rely on the employer's current writing until it is duly amended, we nevertheless conclude that nothing in ERISA prevents Air Jamaica from providing its employees with benefits on a case by case basis—as long as that limitation is explicitly stated as part of the plan.

We find the district court's arguments against enforcing the reservation unpersuasive. First, Air Jamaica's reservation is not an attempt, as the district court asserted, to avoid "award[ing] benefits in accord with its written promises." On the contrary, the reservation is part of the written promise and a limitation upon it. While ERISA was enacted to provide security in employee benefits, it protects only those benefits provided in the plan. For example, even if an oral amendment of an ERISA plan would increase employee benefits, ERISA does not provide for the enforcement of that amendment because it is not part of the plan. *See, e.g., Nachwalter v. Christie*, 805 F.2d 956, 960 (11th Cir. 1986) (refusing to enforce oral promise to use more favorable valuation date in determining employee's share of plan assets). Therefore, unless the reservation can be shown to violate the policies behind ERISA, it must be enforced as part of the plan.

Second, no credible argument can be made that ERISA required Air Jamaica to provide Hamilton with better or more definite benefits. ERISA does not require Air Jamaica to provide its employees with a benefits plan, *Viggiano v.*

Shenango China Div. of Anchor Hocking Corp., 750 F.2d 276, 279 (3d Cir. 1984), nor does ERISA require that Air Jamaica provide any particular set of benefits, if it decides to establish a welfare benefits plan. Pension plans are subject to participation, vesting and funding requirements that do not apply to welfare benefit plans. *Hlinka v. Bethlehem Steel Corp.*, 863 F.2d 279, 283 (3d Cir. 1988). "ERISA mandates no minimum substantive content for employee welfare benefit plans, and therefore a court has no authority to draft the substantive content of such plans." *Blau v. Del Monte Corp.*, 748 F.2d 1348, 1353 (9th Cir. 1984), *cert. denied*, 474 U.S. 865, 88 L. Ed. 2d 152, 106 S. Ct. 183 (1985).

Third, it is not true, as the district court claimed that permitting employers to reserve the right to make individual benefit determinations will render the protections of ERISA illusory. While ERISA imposes no substantive requirements on what welfare benefits are provided, it requires that welfare benefit plans be established, maintained, and disclosed in accordance with its procedural safeguards. The reporting requirements serve "two salutary purposes … to ensure that 'the individual participant knows exactly where he stands with respect to the plan' and 'to enable employees to police their plans.'" *Hozier*, 908 F.2d at 1170. Full disclosure also permits employees to bargain further or seek other employment if they are dissatisfied with their benefits. Enforcing Air Jamaica's reservation does not subvert any of these objectives. Unlike employees who rely on some informal "understanding" about what their benefits are, Air Jamaica's employees are on notice that they have no guaranteed benefits.

Our conclusion here is a logical extension of *Hlinka, supra*. In that case, we found an employer's reservation of the right to award early retirement benefits to a certain class of employees only when the employer believed "that [the particular employee's] retirement would … be in its interest" was not prohibited by ERISA. *Id.* at 283. Having concluded that ERISA permits an employer to reserve the right to make early retirement benefit determinations on an individual basis, nothing suggests we should treat severance pay differently.

Reservations expressed by the employer to make welfare benefits determinations on an individual basis are not prohibited by ERISA and allowing such reservations furthers the interest of employees. Employers are understandably more willing to provide employee benefits when they can reserve the right to decrease or eliminate those benefits. To the extent that employees have sufficient bargaining power to obtain guaranteed benefits, ERISA will enforce those rights and will ensure—through its disclosure requirements—that employees know what benefits they will receive. Therefore, allowing employer reservations of the right to make individual benefit determinations takes nothing away from employees who can command guaranteed benefits and will allow other employees to obtain benefits the employer would refuse to provide on a guaranteed basis.

In short, we conclude that Air Jamaica had no obligation to provide Hamilton with a definite amount of severance pay and that its reservation of the power to determine severance benefits on a case by case basis is therefore valid.

IV.

We conclude that the Handbook constituted Air Jamaica's ERISA plan. We also hold that Air Jamaica's reservation in that plan of the right to make individual benefit determinations is consistent with ERISA and bars Hamilton's claim. Accordingly, we will reverse the judgment below for Hamilton.

ITEM 1: SINGLE-CASE, DETERMINATE

After the *Hamilton* court dealt with the threshold issue of whether or not Air Jamaica's Handbook constituted an ERISA plan, it went on to reason about a second issue. Which one of the following statements best summarizes the court's reasoning concerning this second issue?

a) The court reasoned that when a disclaimer to limit or eliminate benefits is part of a written benefits plan, the disclaimer itself constitutes an amendment and, as such, meets the reporting requirements of ERISA.

b) The court reasoned that an employer's obligation to provide employees with written amendments to published benefit plans is applicable only to vested pension benefits and that, according to the requirements of ERISA, the substance and procedural provisions of welfare benefits are left to the discretion of individual employers.

c) The court reasoned that "employers are understandably more willing to provide employee benefits when they can reserve the right to decrease or eliminate those benefits" through disclaimers included in summary plan documents. Thus, disclaimers are not only consistent with, but also advance the goals of ERISA legislation.

d) ***The court reasoned that if the plan includes an explicit disclaimer reserving the right to change or cancel the plan, then the disclaimer must be viewed as an integral part of the company's promises while serving as an acceptable limitation upon them. By using a disclaimer, the company sponsoring the plan has satisfied ERISA's procedural and reporting safeguards by letting individual participants know exactly where they stand.

e) The court reasoned that the May 30, 1989 letter functioned as a written amendment to the Air Jamaica plan, and, as such, curtailed any subsequent employee claims to previously published severance benefits in its employee Handbook. Thus, all procedural requirements of ERISA were met in this case.

ITEM 2: SINGLE-CASE, DETERMINATE

Which one of the following statements best summarizes the court's reasoning in *Alexander v. Primerica*?

a) ***The court reasoned that the reservation of rights clause in the benefits plan is ambiguous and fails to clarify whether the company is reserving a limited or an unlimited right; moreover, the court reasoned that the company's prior behavior in changing plan benefits does not render its disclaimer unambiguous.

b) The court reasoned that Primerica cannot use its disclaimer to reduce the retirees' health insurance premiums because under ERISA and common law principles, a company cannot reduce benefits below the level that individual employees were receiving on the date that they first began drawing on their retirement benefits.

c) The circuit court reasoned that it remanded the case to the district court because unambiguous language easily could have been employed by American Can Company in its disclaimer and because ambiguous language must be construed against the drafter.

d) The circuit court reasoned that it remanded the case to the district court because, when a formal benefits plan does not exist, employers must be held to the terms outlined in written summary descriptions and to practices established through the conduct of the company in relevant areas.

e) The court reasoned that arguments made by the defendants jeopardized the logic of their own case by implying that their earlier premium increase was not balanced by a much larger benefit increase.

ITEM 3: SINGLE-CASE, DETERMINATE

Which one of the following statements best expresses the issue(s) which the District Court of New Jersey, on remand, will have to decide in view of the *Alexander* court's analysis and decision?

a) The district court must decide whether or not American Can's previous changes to its welfare benefit plan were consistent with ERISA's goal of preserving the solvency of its plan and hence preserving the continued availability of benefits already promised to employees.

b) The district court must decide whether the benefit plan documents provided by American Can unambiguously reserve the right to alter participants' benefits or whether these documents unambiguously promise lifetime benefits that cannot be altered.

c) The district court must decide whether American Can's contested disclaimer phrase, "in conformity with applicable legislation," satisfies ERISA's requirement that summary plan descriptions adequately warn employees of adversity and conditions under which their benefits may be cancelled or denied.

d) The district court must decide whether, given the arguments of both the plaintiffs and the defendants, the benefit plan documents provided by American Can to retired employees have only one or multiple meanings.

e) ***The district court must decide whether American Can's prior behavior in raising the cost of plaintiffs' medical insurance premiums supports the company's claim that it unambiguously reserved the right to reduce plan benefits and never promised lifetime, fixed-cost benefits in its benefit plan documents.

ITEM 4: SINGLE-CASE, DETERMINATE

Which one of the following statements most explicitly articulates the substantive reasoning used by the district court to support its decision to award summary judgment to the defendants in the *Bridell* case?

a) The court gave two reasons: first, because Bridell failed to present any evidence showing that Ribier engaged in discriminatory practices; and second, because Ribier's summary plan document was "written in a manner calculated to be understood by the average plan participant," as required by ERISA.

b) The court reasoned that Ribier Clothing Inc. should be allowed to reduce its medical benefits for two reasons: first, because of the steep, unpredictable rise in the cost of such benefits; and second, because "the paramount purpose of ERISA is to protect the solvency of employee benefit plans."

c) ***The court advanced two reasons for awarding summary judgment: first, because Bridell was unable to convincingly demonstrate that an immutable, promised benefit had been revoked, and second, because Bridell failed to present any evidence showing that Ribier engaged in prohibited conduct or that the company interfered with any right to benefits under ERISA to which she was entitled.

d) The court awarded summary judgment to Ribier Clothing Inc. because Bridell failed to sufficiently establish the existence of an element essential to her case and on which she would need to bear the burden of proof at trial.

e) The court reasoned that summary judgment should be awarded to Ribier Clothing Inc. for two reasons: first, because "Bridell has not and could not reasonably contend that the SPD's earlier reference to 'lifetime medical benefits for eligible employees and their spouses' somehow renders [Ribier Clothing's] reservation of rights in the footnote disclaimer ambiguous;" and second, because ERISA does not require employers to provide any welfare benefits at all.

ITEM 5: CROSS-CASE, DETERMINATE

Which one of the following statements best summarizes the differences in reasoning between the district court and circuit court decisions in *Hamilton*, so far as these differences are visible in the *Hamilton* opinion?

a) Whereas the district court, following *Hozier,* reasoned that Air Jamaica's past actual practice in awarding severance benefits could not substitute for a written plan document, the circuit court reasoned that the district court misinterpreted *Hozier* and argued that cases in most other federal circuits clearly show that past company behavior can be an enforceable indicator of its intended severance pay policy.

b) ***Whereas the district court reasoned that allowing Air Jamaica to enforce its reservation of rights disclaimer would "frustrate ERISA's policy of protecting employees' legitimate expectations," the circuit court reasoned that employees reading Air Jamaica's Handbook, together with its disclaimer, "are on notice that they have no guaranteed benefits" and that, consistent with ERISA, they "know exactly" where they stand.

c) Whereas the district court reasoned that a fundamental purpose of ERISA is to allow employees to form reasonable expectations about their benefits and that orally expressed amendments to benefit plans interfere with such expectations, the circuit court reasoned instead that the fundamental purpose of ERISA is to insure the solvency of employer welfare plans and therefore Air Jamaica has a unilateral right to reduce its severance benefits to help assure such solvency.

d) Whereas the circuit court following the *Viggiano* and *Hlinka* cases reasoned that "ERISA does not require Air Jamaica to provide its employees with a benefit plan" at all, the district court reasoned that the simple fact that Air Jamaica employees were given a benefits Handbook with a sign-off space implies a promise to provide at least some minimal set of welfare benefits.

e) Whereas the district court reasoned from *Hozier* that disclaimers render welfare benefit plans internally contradictory and are tantamount to promising nothing, the circuit court reasoned that the Air Jamaica Handbook's reservation of rights, expressed in its disclaimer, provides employees the leverage to extract legally binding promises from employers.

ITEM 6: CROSS-CASE, DETERMINATE

Which one of the following choices best summarizes the legal issue(s) addressed by *all three* of the cases you have just read?

a) ***Under what conditions does a company's use of a reservation of rights clause in its benefits plan documents satisfy ERISA's competing goals of accurately disclosing the substance of employee benefits while also protecting a company's unilateral right to offer or not offer such welfare benefit plans?

b) Under what conditions does an employers' use of reservation of rights clauses in ERISA summary plan documents violate ERISA's fundamental goal of providing welfare benefit plans in a non-discriminatory manner?

c) Under what conditions may employers satisfy ERISA's goal of protecting employees' benefits while at the same time denying welfare benefits explicitly identified in both benefit plan documents and in specific, individual employee cases?

d) Under what conditions may an employer's prior history and behavior in increasing or reducing plan benefits to employees be used to clarify the meaning of disputed terminology in published welfare benefit plans?

e) Under what conditions may ERISA's goal of protecting employers' unilateral right to offer or not offer welfare plan benefits to employees be given more weight than ERISA's goal of protecting employees' reasonable expectations both for such benefits and accurate disclosure concerning them?

ITEM 7: CROSS-CASE, DETERMINATE

Assume for the moment that you decide to prepare an appeal to the Fifth Circuit to re-examine the district court decision against your client, Joan Bridell. Which one of the following summaries of the similarities between the facts in *Bridell* and the facts in *Alexander* would be most accurate *as well as* most useful for you on appeal?

a) The main factual similarity between these cases is that the employers in each predicated their reductions in welfare plan benefits upon the presence of a reservation of rights disclaimer placed within the benefit plan documents that were distributed to their employees.

b) The main factual similarity is that the employers in each case previously increased their employees' health insurance premiums and thus set a clear precedent that they had a unilateral right to further increase or decrease benefits at any time.

c) The chief factual similarity between these cases is that in each the plaintiffs were objecting to reductions in their health-related "welfare" benefits, claiming that the reductions violated ERISA-protected written promises concerning these benefits made by employers.

d) The main factual similarity between these cases is that the employers in both cases predicated their reductions in employees' welfare plan benefits upon a formal amendment process and upon the urgent need to preserve the solvency of their plans.

e) ***The main factual similarity between these cases is that the employer's disclaimer in *Alexander* used the phrase "in conformity with applicable legislation," and the employer's disclaimer in *Bridell* used the phrase "pursuant to changes in federal legislation."

ITEM 8: CROSS-CASE, DETERMINATE

Which one of the following statements best summarizes the view of the courts in the three cases you read concerning the fundamental legality, under ERISA, of employers' use of disclaimers to reserve the right to change their welfare benefit plans?

a) It is clear from *Hamilton* that employers are free to use, phrase, and locate their disclaimers in any manner they choose in order to deny the provision of welfare plan benefits that may be described elsewhere in their plan documents. At the same time, the other two cases, *Bridell* and *Alexander*, make it clear that use of reservation of rights disclaimers is not illegal in principle under ERISA.

b) There is no indication in either *Bridell* or *Alexander* that the use of disclaimers in summary plan documents to reserve the right to change or eliminate welfare plan benefits necessarily violates ERISA's disclosure goals; however, the court's analysis in *Hamilton* clearly shows that use of such disclaimers is legal under ERISA only if they are presented together with a place for employees to sign and thereby provide informed consent.

c) While *Hamilton* and *Bridell* offer no indication that the use of disclaimers to deny the provision of welfare plan benefits is illegal in principle under ERISA, the *Alexander* decision clearly offers a direct challenge to such use by reasoning that they invariably confront employees with ambiguity and obscurity and therefore create doubt as to their interpretation.

d) Although *Hamilton* and *Alexander* are silent as to whether or not employers' uses of disclaimers to reserve certain rights concerning their welfare benefit plans are illegal in principle under ERISA, it is clear from *Bridell*'s discussion of the *Hamilton* case that such disclaimers must be presumed to be legal unless they can be shown to be discriminatory toward specific classes of employees.

e) ***The *Hamilton* case in its reasoning implies that a reservation of rights disclaimer is legal under ERISA as long as it is published within a summary plan document and not merely communicated orally to employees; however, the *Bridell* and *Alexander* cases in their reasoning each imply that a reservation of rights disclaimer may be illegal, hence unenforceable, under ERISA if its meaning is demonstrably ambiguous.

ITEM 9: SINGLE-CASE, INDETERMINATE

When you are researching among cases, you may find that they contained unanswered questions that, if logically extended to your case, could frame issues upon which you could build arguments for your client. Which one of the following questions remains unanswered in *Hamilton*, and, as such, might best provide you with an issue relevant to your appeal for Bridell?

a) On what bases might an employer be permitted to extend, curtail, or deny welfare benefits specified in summary document plans to discrete groups of employees?

b) ***To what extent might employers be held liable for ensuring that employees are cognizant of meanings contained in reservation of rights clauses within published summary plan documents?

c) Under what circumstances will the established behavior of a company supersede published documents relative to the provision of welfare benefits?

d) When is it permissible for an employer to withhold promised benefits to individuals?

e) Does ERISA allow for employees to individually contract with their employers to obtain particular welfare benefits?

ITEM 10: SINGLE-CASE, INDETERMINATE

Assume that you are considering ways in which you might help Bridell to appeal the decision of summary judgment against her. Which one of the "unknowns" in the facts of the *Bridell* case would provide the best grounds for a possible motion?

a) Was there any evidence that the "financial losses" suffered by Ribier had not resulted from extraordinary expenses related to their benefit plan, but were the result of high risk investments that failed to yield expected returns?

b) Are there any records of conversations between Ribier's senior managers and its insurers concerning the projected costs related to the lifetime coverage of the average AIDS patient?

c) ***Are there any company records (letters, recorded phone calls, emails, etc.) that contain individual employee questions or confusions about how the SPD disclaimer affects "lifetime benefits," and can it be shown that some company personnel responded evasively or unclearly to these?

d) Did Bridell inform her employer, Ribier, of her AIDS diagnosis in a memo in which she also made reference to the terms of the existing plan and did a company representative acknowledge that memo in writing?

e) Did Ribier ever change the provisions of its medical plan either increasing or decreasing benefits?

ITEM 11: SINGLE-CASE, INDETERMINATE

Although you did not represent Bridell at the District Court level, you have agreed to take Bridell's case to the Fifth Circuit. In your research you find that Section 510 of ERISA states: "It shall be unlawful for any person to discharge, fine, suspend, expel, discipline, or discriminate against a participant or beneficiary for exercising any right to which [s]he is entitled under the provisions of an employee benefit plan…" 29 U.S.C. § 1140. In light of this provision, which one of the following questions, if found in the affirmative, would be most likely to thwart Ribier's position in a new case?

a) ***Can Bridell prove that the reduction in AIDS benefits from a lifetime cap of $1,000,000 to $5,000 was enacted to retaliate against her for filing a large claim?

b) Can Bridell point to previous Ribier employees diagnosed with AIDS-related illnesses who received lifetime benefits well beyond $5,000?

c) Can Bridell provide evidence that she had not received reimbursements for the claims she filed with National American Life Insurance prior to July 1988?

d) Can Bridell demonstrate that AIDS is the only catastrophic illness to which a $5,000 limit is applied?

e) Can Bridell prove that in 1987 Ribier Clothiers, Inc. stock increased significantly, but that employees nearing retirement were discharged as a purported economic measure?

ITEM 12: CROSS-CASE, INDETERMINATE

Assume that your petition to appeal the *Bridell* case has been successful. As you prepare your case you will be thinking about how opposing counsel might look to other cases to effectively develop its case. In light of this task, which one of the following questions regarding opposing counsel's argument strategy would concern you the most?

a) To what extent will opposing counsel try to establish that just as medical benefits could be reduced for retirees as they were in *Alexander,* so can they be reduced for active employees like Bridell?

b) How might opposing counsel compare the disclaimer evidenced in Bridell's plan with the disclaimer found in Hamilton's plan?

c) To what extent might opposing counsel draw upon both *Hamilton* and *Alexander* to synthesize a rule concerning promises that inhere in benefit plan documents?

d) How might opposing counsel rely on the *Alexander* court's hypotheticals about possible federal legislation affecting the provision of private health care plans?

e) ***To what extent will opposing counsel attempt to show that the case by case justification put forward in *Hamilton* rightfully extends to Bridell's case?

ITEM 13: CROSS-CASE, INDETERMINATE

The *Alexander* court warns any court or lawyer who reads its opinion that "slight similarity" is insufficient to the purpose of comparing facts from one case to the facts in another case. In addition, the facts a lawyer selects for comparison must be significant within the context of the case at hand. Given these two considerations, which one among the following questions might present the strongest challenge to the *Bridell* court's analogical reasoning?

a) ***The *Bridell* court states that Ribier's disclaimer is "similar" to the one examined in the *Hamilton* case. Does the fact that Air Jamaica's disclaimer appeared in the body of its plan's text and the fact that Ribier's disclaimer appeared in a footnote render this assertion of similarity suspect?

b) Air Jamaica's disclaimer explicitly stated that their benefits were subject to change "without notice," resulting, according to the court, in employees clearly knowing where they stood with regard to expected benefits. According to the *Bridell* court, Ribier employees should have expected the same possibility since they were put on notice that their plan benefits could change "at any time." Are the phrases "at any time" and "without notice" sufficiently similar to be taken as equivalent in meaning?

c) Air Jamaica's disclaimer assigns executive control regarding any changes to welfare benefit plans to the "Plan Sponsor." Ribier reserves such actions to "a benefit plan administrator." Are these terms substantively similar and meaningful enough to be considered appropriate synonyms and, thus, the basis of an apt analogy?

d) In *Hamilton*, the benefit at issue was a severance benefit. In *Bridell*, at issue was the amount of coverage allowed for a catastrophic disease, AIDS. Are these factual elements similar enough to carry a defensible analogy?

e) The disclaimer issued in Air Jamaica's plan specifically allowed the company to "reduce…any…benefit." The language contained in Ribier's disclaimer sanctioned the company to "terminate any benefit." Since the AIDS benefit was not "terminated" but actually "reduced" did the *Bridell* court err in counting this aspect of the two cases analogous?

ITEM 14: CROSS-CASE, INDETERMINATE

All court opinions involve the construction of arguments; in turn, arguments invite questions. In light of this, which one of the following lines of questioning related to Bridell's previous defeat in district court is best supported by the cases you have read?

a) ***In assessing whether or not Ribier lawfully changed particular terms in its medical plan, the district court acknowledged one requirement of ERISA that an SPD "be written in a manner calculated to be understood by the average plan participant." Did counsel for Bridell consider obtaining expert evidence concerning the visibility and understandability of the footnoted disclaimer and its relationship to the phrase in the main part of the SPD referring to "lifetime medical benefits"?

b) In deciding on what motive accounted for the changing of Ribier's welfare benefits plan, the court accepts the defendant's claim that their change was prompted by a desire "to ensure the future existence of the plan." Did counsel for Bridell consider investigating the financial reports of Ribier to ascertain whether the company had indeed incurred major financial losses?

c) The difference between the benefit cap afforded to Bridell as an AIDS patient before the change ($1,000,000) and the one offered after the change in the plan ($5,000) was considerable. Did counsel for Bridell consider arguing about the harshness and the unreasonableness of this reduction in terms of a catastrophic disease like AIDS?

d) We know that Ribier provided its employees with an SPD that contained a disclaimer expressing the company's reservation of the right to change the plan. Did counsel for Bridell check to confirm that this reservation was also included in any other plan document or previously distributed SPDs that outlined welfare benefits for employees?

e) The Bridell court maintained that the changes in the welfare benefits plan "were not made to discriminate against Bridell or anyone who was diagnosed with AIDS." It then went on to illustrate how Bridell's claims related to her AIDS condition were not denied either before or after the benefit changes. Did counsel for Bridell consider arguing that Ribier might have dealt with its plan solvency problem by spreading out reductions across a large number of different illnesses rather than focusing them upon benefits for AIDS patients?