DOCUMENT RESUME

ED 387 513 TM 023 731

AUTHOR Sawyer, Richard
TITLE Eliciting Utility Functions for Validating Course Placement Systems.
PUB DATE 21 Apr 95
NOTE 78p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 19-21, 1995). For the 1994 version, see ED 373 067.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Academic Achievement; College Faculty; *College Students; Grades (Scholastic); Higher Education; *Outcomes of Education; Remedial Instruction; Student Attitudes; *Student Placement; Teacher Attitudes; *Validity
IDENTIFIERS Decision Theory; *Utility Functions

ABSTRACT

Course placement systems in postsecondary education consist of an assessment component to estimate students' probability of success in standard first-year courses and an instructional component in which underprepared students are taught the skills and knowledge they need to succeed in the standard courses. Student success is usually defined in terms of course grades. Using a decision theory model to judge the effectiveness of course placement systems, the feasibility of eliciting students' and instructors' preferences for the different outcomes of course placement systems was studied with groups of 129 and 141 students and 9 instructors. The results suggest that about half of the respondents to a paper-and-pencil instrument provide sufficient information to develop coherent preferences for the outcomes of a course placement system. The elicited preferences differed significantly according to the method used (value function versus hypothetical lotteries). Responses of students and instructors were similar. Appendix A contains the questionnaire, Appendix B (two tables) discusses eliciting von Neumann-Morgenstern utilities for course grades, and Appendix C (two tables) discusses constructing summary value functions for course placement outcomes. (Contains 4 tables, 1 figure, and 31 references.) (SLD)

ED 387 513

Eliciting Utility Functions for Validating Course Placement Systems

Richard Sawyer

Research Division,
American College Testing
P.O. Box 168
Iowa City, Iowa 52243

Paper presented at the annual meeting of the
National Council on Measurement in Education, San Francisco

April 21, 1995

Abstract

Course placement systems in postsecondary education consist of an assessment component (to estimate students' probability of success in standard first-year courses), and an instructional component (in which underprepared students are taught the skills and knowledge they need to succeed in the standard courses). The effectiveness of a placement system depends on students' ultimately succeeding in the standard courses. Success is usually defined in terms of course grades.

Using a decision theory model to judge the effectiveness of course placement systems, I studied the feasibility of eliciting students' and instructors' preferences for the different outcomes of course placement systems. The results suggest that about half of the respondents to a paper-and-pencil instrument provide sufficient information to develop coherent preferences for the outcomes of a course placement system. The elicited preferences differed significantly according to the method used (value function vs. hypothetical lotteries). Students and instructors' responses were similar.

Eliciting Utility Functions for Validating Course Placement Systems[1]

Richard Sawyer

A typical and important use of college entrance tests is course placement, i.e., matching students with instruction appropriate to their academic preparation. For example, students whose academic skills are insufficient for them to be successful in a standard first-year English course might, on the basis of their test scores and other characteristics, be advised or required to enroll in a remedial English course. On the other hand, students with an unusually high level of academic preparation might be encouraged to enroll in an accelerated course or in a higher-level course.

Most colleges and universities enroll students who are not academically prepared to do work at a level traditionally expected of first-year students. The percentage of postsecondary institutions with some form of placement and remedial instruction has steadily increased in the past decade, and is now about 90% (Woods, 1985; Wright and Cahalan, 1985; McNabb, 1990; "Colleges and Universities Offering Remedial Instruction," 1994). One suggested explanation is that American high schools have become less effective in preparing students for college (The National Commission on Excellence in Education, 1983; The Carnegie Foundation for the Advancement of Teaching, 1988; Singal, 1991). Another explanation is that more students from disadvantaged backgrounds are attending college (Munday, 1976; College Entrance Examination Board, 1977; Carriuolo, 1994).

During the past three decades, several authors have proposed using decision theory to validate educational selection systems. Two different general approaches are those proposed by Cronbach and Gleser (1965) and by Petersen and Novick (1976). Cronbach and Gleser adapted linear regression methodology to estimate the expected costs and benefits of using a test score or other predictor variable for classifying or selecting personnel. Their technique continues to be widely applied in industrial/organizational settings. Petersen and Novick (1976) developed a "threshold" model based on Bayesian decision theory. Ben-Shakhar, Kiderman, and Beller (1994) compared these two approaches, and illustrated them using data from an admission selection problem.

[1] I wish to express my appreciation to Dan Anderson, Jerry Dallam, and Chuck Hinz for their help in collecting data for this study; Mark Houston for calculating the utility functions; and Mark Houston, Alan Nicewander, and Julie Noble for their comments on earlier drafts of this paper.

Sawyer (in press) proposed a statistical decision theory model for validating course placement variables such as tests. The model can be used to compare the effectiveness of alternative placement variables in identifying underprepared students, and to determine appropriate cutoff scores on these placement variables. Sawyer (1994) proposed a decision theory model for measuring the effectiveness and worth of remedial instruction. In this paper, alternative methods are investigated for eliciting decision makers' preferences for course placement outcomes.

Background

Remedial Instruction

At many postsecondary institutions, there are two levels of first-year courses: a "standard" course in which most students enroll, and a "remedial" course for students who are not academically prepared for the standard course. At some institutions, the lower-level course may be given other names, such as "college-preparatory," "compensatory," "developmental," or "review." Carriuolo (1994) articulated differences in the meanings of "remedial" and "developmental." At some institutions, there may be courses that require more knowledge and skills than the lowest-level remedial course, but less than the standard course. In this paper, only a single lower-level course is considered, and it is designated "remedial," to be consistent with Willingham's (1974) nomenclature. Often, remedial courses do not carry credit toward satisfying degree requirements.

Though essential to placement, testing is but one component of an overall system. To be educationally effective, a placement system must satisfy all of the following requirements:

1. Students who have small chance of succeeding in the standard course (underprepared students) are accurately identified.
2. Appropriate remedial instruction is provided to these underprepared students.
3. Both the students who originally enrolled in the standard course, and the students who were provided remedial instruction, eventually do satisfactory work in the standard course.

Note that accurately identifying underprepared students (Requirement 1) is necessary, but not sufficient, for a placement system as a whole to be effective. Accurate identification is not an end, but only a mechanism for effectively allocating remedial instruction (Requirement 2). On the other hand, providing remedial instruction is itself only a means to achieve the larger goal that students succeed in college: Even if underprepared students are accurately identified and are provided remedial instruction, if they eventually drop out or fail in the standard course, then little will have been accomplished by the placement system. On the contrary, both the institution's and the students' resources will have been wasted. Van der Linden (1991) noted that a defining characteristic of course placement systems is that students take different treatments (courses), and the success of each treatment is measured by the *same* criterion variable.

One might argue that failure in the standard course can lead to positive results for students, such as their selecting and succeeding in another educational program better matched with their talents and interests. While this statement is undoubtedly true for some students, they would have done better to select their preferred educational programs in the first place, through appropriate counseling. This scenario illustrates that effective counseling is important for effective placement. This paper does not, however, attempt to model the effect of counseling on the outcomes of placement.

The need for an institution to serve students who by traditional standards are academically unprepared for college imposes a fourth requirement on placement systems. Even if a large proportion of the underprepared students are accurately identified, are provided remedial instruction, and ultimately succeed in the standard course, the overall result still might not be satisfactory. This would occur if an institution diverted resources to instruction in the remedial course to such an extent that the performance of students in the standard course was adversely affected. In other words, institutions should consider the tradeoffs they must make in allocating their finite resources when they provide remedial placement systems; such considerations may relate to institutional mission and policy, as much as to costs and to grades. There is controversy about the proper role of remedial placement in the missions of

postsecondary institutions. Mac Donald (1994) argued that by overexpanding its remedial programs, the CUNY system seriously degraded the quality of its standard-level undergraduate programs. Lively (1993) reported on efforts in different states to eliminate remedial instruction from four-year public institutions by designating that role to two-year colleges.

A Decision Theory Model for Course Placement

The decision problem can be formally defined as follows: One must select a particular decision d from a set D of possible decisions. A particular outcome θ occurs, from among a set of possible outcomes Θ. A utility function $u(d,\theta)$ assigns a numerical value to the desirability of decision d when the outcome is θ. The exact outcome θ that occurs is unknown to the decision maker, but there is some probabilistic information available about the likely values of θ. In a Bayesian decision model, this information is described by a subjective probability distribution on Θ; the subjective probability distribution quantifies the decision maker's personal beliefs about the likely values of θ, given both prior beliefs and any relevant data collected. The Bayesian optimal strategy is to choose the decision d that maximizes the expected value of $u(d,\theta)$ with respect to the subjective probability distribution on Θ (Lindley, 1972).

To apply this structure to course placement, suppose there is a cutoff score K on a placement test, and that:

* test scores are obtained for all first-year students at an institution;
* students whose test scores are less than K are provided remedial instruction before they enroll in the standard course, and students whose test scores are greater than or equal to K enroll directly in the standard course; and
* the actual final performance in the standard course is known for all students (i.e., for students who are provided remedial instruction, as well as for those who are not).

The final performance in the standard course of students who first enroll in the remedial course will, of course, become known later than the performance of students who enroll directly in the standard course. For each student, four possible events could occur, as shown in Table 1 below.

Table 1
Events Associated with Identifying and Providing Remedial Instruction to Underprepared Students

Event	Test score	Course into which student is placed	Eventual performance in standard course
(1)	$\geq$ K	Standard	Successful
(2)	$\geq$ K	Standard	Unsuccessful
(3)	< K	Remedial	Unsuccessful
(4)	< K	Remedial	Successful

Each student is classified either as being adequately prepared for the standard course (if his or her test score equals or exceeds the cutoff score K), or as needing remedial instruction (if the score is less than K). Because the classification for any student depends on K, the set of decisions (*D*) in this case is the set of possible values of K. The goal is to find the "best" value of K, and to quantify the effectiveness of the associated instruction.

At an institution without a placement system, the events in Table 1 could be observed as follows:

* Randomly assign students, regardless of their test scores, either to enroll directly in the standard course or to enroll first in the remedial course.
* Observe the students' eventual performance in the standard course, and note which of them succeed and which do not succeed.

For each value of K, there would be a set of proportions associated with the events (1) - (4). Let us suppose, temporarily, that data are collected this way; the modifications required when there is prior selection resulting from an existing placement system are described on p. 9.

Let $p_1(K)$, $p_2(K)$, etc., denote the observed proportions of students corresponding to events (1), (2), etc., in the entire group of students when the cutoff score is K. (Then, for example, $p_1(K) + p_4(K)$ is the proportion of students who are ultimately successful, and $p_2(K) + p_3(K)$ is the proportion of students who

are ultimately unsuccessful. The overall usefulness of the predictions can then be evaluated in terms of the costs and benefits associated with each event (1) - (4). A function that assigns a value to outcomes such as these is called a utility function. One class of utility functions would assign different values to each event, and weight their sum:

$$u(K;\theta) = w_1p_1(K) + w_2p_2(K) + w_3p_3(K) + w_4p_4(K) \qquad (1)$$

where $0 \leq w_1,\ldots,w_4 \leq 1$. Such a function would quantify the different costs and benefits of each outcome. Consider, for example, the trade-offs a student must make in his or her utility. Although students pay tuition to take remedial courses (just as they do to take other courses), remedial courses often do not carry college credit. From a student's perspective, the weights $w_1, \ldots, w_4$ must balance the benefit in performance in the standard course against the extra time and money spent on taking the remedial course.

In principle, utility functions are person-specific, and hence need to be elicited separately for each student, counselor, teacher, or administrator. In practice, this is not feasible, and we must look for utility functions that reasonably approximate the preferences of different groups of people.

Other models

In the model described in Table 1, there are only two results in the standard course: "Successful" and "Not successful." In practice, "Successful" usually means completing the standard course with a particular grade (e.g., C) or higher. A more basic decision theory model, defined directly in terms of the grade received, would describe people's preferences more accurately. For example, instead of designating each student as "Successful" or "Unsuccessful" in the standard course, one could specify the student's completion of the course and final grade (e.g., A-F). In this case, there would be 10 outcomes (rather than 4) in the model; such a model is described on p. 16. If we considered that some students withdraw before completing the standard course, then there would be 12 outcomes in the model.

The adequacy of the model in Table 1 therefore assumes that the decision maker's preferences for particular grades have a step-function relationship. Petersen and Novick (1976) called such a function a

"threshold utility." One goal of this study was to obtain evidence about the appropriateness of threshold utility functions in course placement.

Expected Utility Functions

In practice, a utility function cannot be directly computed for the group of students for whom placement decisions are to be made, because the actual outcomes (students' test scores and eventual performance in the standard course) are not yet known. In (1), for example, the actual proportions $p_1(K)$, $p_2(K)$, etc., are not known for a particular group of students before they are tested and complete the standard course. These proportions must instead be estimated in some way from data on past students, under the assumption that future students will be similar to past students.

The "expected utility function" is a formal mechanism for dealing with the uncertainty of outcomes in a decision theory model. It is from the expected utility function that decisions on the effectiveness of a placement system can be made. In Bayesian models, an "expected utility function" is the average (expected) value of a utility function $u(d,\theta)$ with respect to the decision maker's subjective probability distribution for the outcomes Θ. In the example previously given,

$$u'(K) = E_\theta[\, u(K,\theta)\,] = w_1\hat{p}_1(K) + w_2\hat{p}_2(K) + w_3\hat{p}_3(K) + w_4\hat{p}_4(K) \qquad (2)$$

where $\hat{p}_1(K) = E_\theta[\, p_1(K)\,]$, $\hat{p}_2(K) = E_\theta[\, p_2(K)\,]$, etc., are estimated from a past group of students. In the Bayesian model, the estimates $\hat{p}_1(K)$, $\hat{p}_2(K)$, etc., are the expected values of the corresponding observed proportions with respect to the decision maker's subjective probability distribution for students' test scores and course grades. In the terminology of Bayesian statistical inference, the subjective probability distribution for test scores and course grades is specified by a "predictive density" for their joint distribution. The predictive density is based on prior beliefs about the joint distribution and on data obtained from a particular group of past students. Although simple in concept, Bayesian statistical methods can be mathematically formidable in real applications. When prior beliefs are vague or as sample sizes become large, however, Bayesian estimates are, for practical purposes, similar to much simpler

estimates based on classical sampling theory (i.e., estimates based only on an assumed model and on data; DeGroot, 1970).

Sawyer (in press) described a simple procedure, based on sampling theory, for estimating the cell probabilities $\hat{p}_1(K)$, $\hat{p}_2(K)$, etc. The first step is to estimate the relationship between success in the standard course and a placement test score using a logistic regression function:

$$P[Y=1 \mid X=x] = (1 + e^{-\alpha-\beta x})^{-1} \tag{3}$$

where Y = 1, if a student is successful,

= 0, if a student is unsuccessful;

and X is the student's score on a placement test or other placement variable. The numbers α and β in Equation (3) are unknown parameters that are estimated from data on the test scores and on the success/failure variable Y for a group of enrolled students. The regression function $P_S(x)$ of students who enroll directly in the standard course and the regression function $P_R(x)$ of students who enroll first in the remedial course are estimated separately.

Once estimates *a* and *b* have been obtained for the unknown parameters α and β, the conditional probabilities of success $\hat{P}_S(x)$ and $\hat{P}_R(x)$ can be estimated by substituting *a* and *b* in Equation (3). From the estimated conditional probabilities, the proportions for the four events described in Table 1 can be easily calculated. For example, the proportion of students associated with Event (1) in Table 1 can be estimated by:

$$\hat{p}_1(K) = \sum_{x \geq K} \hat{P}_S(x) * n(x) / N \tag{4}$$

where $\hat{P}_S(x)$ = estimated P [Y = 1 | X = x] for students who enrolled directly in the standard course,

K = the minimum score required for enrollment in the standard course (cutoff score),

n(x) = the number of students in the placement group whose test score is equal to x, and

$N = \sum_{\text{all } x} n(x)$, the total number of students in the placement group.

The proportions for Events (2), (3), and (4) can be estimated similarly.

Note that the summations in Equation (4) are based on the x-values (e.g., test scores) of all the students in the placement group (the set of students for whom placement decisions are made), not just the students who complete the course. In practical terms, the placement group will usually consist of all first-time entering students with test scores, regardless of which course they actually enroll in. Of course, one could also define a placement group for students in a particular program of study (e.g., business) or with particular background characteristics (e.g., minority students).

At an institution with an operational placement system with cutoff score K_0, we can estimate $P_S(x)$ only from data with $x \geq K_0$, and we can estimate $P_R(x)$ only from data with $x < K_0$. The reason is that students whose test scores are below the cutoff score K_0 do not enroll directly in the standard course, and therefore do not have performance data unaffected by remedial instruction. Sawyer (1993) noted, however, that the logistic regression model (3) can be conveniently extrapolated to test scores below the current cutoff score K_0. Schiel & Noble (1993) compared logistic regression functions estimated from truncated subsets of a data set that was not subject to prior selection. They found that when the truncation involved less than 15% of the population, the resulting errors were small, but that large amounts of truncation (e.g., 50%) resulted in large errors. Houston (1993) did computer simulations to examine the effects of truncation on the accuracy of estimated conditional probabilities of success. He found increases in standard error of 6%, 30%, and 43% when the placement group was truncated at the 25th, 50th, and 75th percentiles, respectively, as compared to the standard error associated with no truncation.

Optimal cutoff scores

If the expected utility $u'(K) = E_\theta[u(K;\theta)]$ attains a maximum value at some cutoff score K_0, then using K_0 as a cutoff score will result in a greater expected utility for the group than using any other cutoff score. Furthermore, if K_0 is between the minimum and maximum possible scores on the test or other placement variable, then the effectiveness of the placement system as a whole is supported. On the other hand, if u' is an increasing function, then the effectiveness of the placement variable is called into question — the

placement variable is not able to discriminate between students who should enroll directly in the standard course and those who should first take the remedial course. Finally, if u' is a decreasing function, then the effectiveness of both the placement variable and the remedial course is called into question. Of course, all of these inferences depend on the validity of the success criterion variable.

Eliciting Utility Functions

If the decision model and optimal cutoff score are to be useful in real applications, the utility function must accurately describe the preferences of the decision makers. In the model described by Table 1, for example, we need some way to quantify students' and instructors' preferences for success in the standard course, as balanced against the extra time and cost associated with taking the remedial course.

There is a vast literature on eliciting (i.e., assessing) utility functions. One important characteristic distinguishing various utility theories is whether they are deterministic or stochastic:

* A *value function* measures the satisfaction of any sort of "want" without regard to uncertainty. For example, some economists model the satisfaction that an individual receives from consuming commodities. The key characteristic of a value function is that it assigns numerical values to the subjective worth of outcomes without regard to uncertainty (Yates, 1990).

  A simple example of eliciting a value function would be to ask an individual to rank each possible outcome on the following Likert scale:

  1="dislike very much", 2="dislike", 3="dislike a little", ..., 7="like very much")

  Note that in this example, the assignment of values to outcomes is done outside any context of uncertainty or risk.

* A *von Neumann-Morgenstern* utility, in contrast, is explicitly defined in terms of uncertainty. The standard assumption in von Neumann-Morgenstern (abbreviated hereafter as vN-M) theories is that the decision maker has a preference relation $\prec$ over the set Π of *probability distributions* on the outcome space Θ (rather than on Θ itself), and that $\prec$ satisfies an appropriate set of axioms (e.g., transitivity). Then it can be shown that there exists a real function u on Θ, such

that for distributions p, $q \in \Pi$, $p \prec q$ if, and only if, $E_p[u] < E_q[u]$. The function u is unique up to positive, linear transformations; therefore, one can without loss of generality assign the value 0 to the least favorable outcome and the value 1 to the most favorable outcome. Note that vN-M utility functions are defined in terms of probability; therefore, their elicitation is naturally done in reference to hypothesized probability distributions. See Farquhar (1984) for a comprehensive review of different strategies for eliciting vN-M utility functions.

The principal advantage of value functions is that they are easy to elicit, because they do not require any reference to uncertainty or risk. The principal advantage claimed for vN-M utility functions is that they are more realistic, because they reflect the decision maker's feelings about both the inherent worth of the outcomes, and the risk involved in making choices. (On the other hand, this realism is elicited in the context of hypothetical situations!) Although both value functions and vN-M utility functions can formally be used in expected utility models (Yates, 1990), they are not the same, and can lead to different decisions. I shall follow Yates' convention in reserving the term "utility" to refer specifically to a vN-M utility function, and the term "value function" for a function that does not consider risk.

One class of methods for eliciting vN-M utility functions is called "probability equivalence" methods. Probability equivalence involves asking a decision maker to determine the probability p for which he or she is indifferent to obtaining Outcome θ_k with certainty, and a gamble involving Outcome θ_i with probability p and Outcome θ_j with probability 1-p. Farquhar (1984) denotes this relationship as $\theta_k \sim [\ \theta_i, p, \theta_j\ ]$. Different probability equivalence methods involve different ways of choosing the outcomes in the hypothetical gambles. Novick and Lindley (1979), for example, order the n outcomes $0 = u(\theta_0) < u(\theta_1) < \ldots < u(\theta_{n-1}) < u(\theta_n) = 1$. They then make comparisons involving the n-1 adjacent outcomes: $\theta_i \sim [\theta_{i+1}, p_i, \theta_{i-1}]$. Finally, they solve the resulting system of linear equations: $u(\theta_i) = p_i * u(\theta_{i+1}) + (1-p_i) * u(\theta_{i-1})$, where i=1,...n-1. Novick and Lindley also consider additional gambles involving more distant comparisons, such as $\theta_i \sim [\theta_{i+2}, r_i, \theta_{i-2}]$, to check the consistency (also called "coherence") of the elicited utilities.

Decision theory provides an intellectually attractive method for studying the effectiveness of remedial instruction. Its practical feasibility in this application, however, needs to be proven. Among the feasibility issues, eliciting the preferences of students and instructors is certainly crucial: If these decision makers are unable to provide information that accurately reflects their preferences and is inexpensive to collect, then the method will be only a toy of statisticians, rather than a practical means for improving postsecondary education.

The purpose of this study was to obtain preliminary answers to the following questions:

1. Is it feasible to elicit utilities by a paper-and-pencil questionnaire?
2. How do different analytic schemes for eliciting utilities affect the results?
3. Is the threshold utility a reasonable approximation to students' or instructors' utilities?
4. How do the utilities of students differ from those of instructors?

The reasons for posing these questions are discussed below.

Question 1 has implications for the feasibility of eliciting utilities in a routine and large-scale settings. Sophisticated interactive computer systems (e.g., Isaacs & Novick, 1978) have been developed for eliciting utilities; these systems have internal mechanisms for detecting and correcting inconsistencies in decision makers' responses, either by asking for additional information, or by smoothing, or both. It would certainly be more economically and practically feasible, however, for institutions to administer paper-and-pencil questionnaires than to maintain or subscribe to an interactive computer system.

Question 2 also has implications for eliciting utilities in large-scale settings. Value functions are much easier to elicit than vN-M utility functions. If the elicited value functions of most decision makers closely approximated their elicited vN-M functions, then one would need to elicit only the value functions.

To answer Question 3, let $u(G)$ denote an instructor's value function or vN-M utility for grade G, and let "<<" denote "much less than." If we observed the following result:

$$u(F) < u(D) << u(C) < u(B) < u(A), \tag{6}$$

then a threshold utility with respect to the C-or-higher success criterion would be supported. If we found that

$$u(F) < u(D) < u(C) << u(B) < u(A), \quad (7)$$

then a threshold utility with respect to the B-or-higher success criterion would be supported.

Because students and instructors obviously have different roles in education, they may well have different preferences (Question 4). Two other groups, counselors and administrators, are also important decision makers in course placement systems, and could have utilities that differ in important ways from those of both students and instructors. Unfortunately, it was not possible in this study to administer a questionnaire to counselors and administrators. Future studies will include them.

Definitive answers to these questions undoubtedly depend on many educational and background variables, and could be the goal of an entire research agenda. For example, utilities of students and instructors at different types of institutions (e.g., 4-year liberal arts colleges, state universities) may differ from those of students and instructors at community colleges. This study, it is hoped, provides initial "order-of-magnitude" results, as well as guidance on how to design more sensitive studies in the future.

Data

I administered questionnaires to the following groups of people:

* Group 1: Students who enrolled in first-year remedial or standard English or mathematics courses at a community college in the midwest (n=129).
* Group 2: Students who enrolled in basic algebra or calculus at a public university in the midwest (n=141).
* Group 3: The instructors of the students in Group 2 (n=9). This group included 1 faculty member and 8 graduate students.

To make the questionnaire items more meaningful to the respondents, I developed a separate questionnaire for each group. The questionnaires are reproduced in Appendix A. (The two institutions are given the fictitious names "Midwest Community College" and "Midwest Public University.")

Background Information

Part 1 of each questionnaire asked about respondents' course taking (or course teaching) experience, as well as background information. These questions will be used in future studies to determine whether particular groups of respondents have particular difficulties in providing preference information, and whether their preferences differ substantially from each other.

Value Function

Part 2 of each questionnaire elicited a value function for the grades of B, C, and D in the standard course. To simplify the respondents' deliberations, this question ignored the possibility that a student might withdraw (W) or obtain an incomplete (I) in the standard course. Respondents were presented with a scale ranging from 0 to 100, and incremented in units of 10. In the student version of the questionnaires, the scale was intended to measure satisfaction with particular grades, with F indicating a satisfaction of 0, and A indicating a satisfaction of 100. In the instructor version of the questionnaires, the scale was intended to measure satisfaction with student academic performance levels associated with different grades. Respondents were asked to mark the letters D, C, and B over the points on the scale that reflected their satisfaction with these grades.[2] This method of eliciting value functions is called "Stevens' magnitude estimation with modulus" (Falmagne, 1985).

vN-M Utility

Part 3 of the student questionnaire elicited vN-M utility functions for the grades B, C, and D. I asked students to consider either earning a particular grade G for sure, or else entering a lottery involving grades G_0 and G_1, where $G_0 < G < G_1$. In the lottery, students would earn grade G_0 with probability p, or grade G_1 and probability $1-p$. Here is an example item:

[2] Respondents were warned not to confuse the "satisfaction scale" with the "percent correct scales" sometimes used to assign grades (e.g., A=90-100).

Suppose I offered you the choice of:
(1) Earning exactly a grade of D for sure, or
(2) Taking a chance, where you could earn either a C or an F.

How large would the chances of earning a C have to be before you would prefer taking a chance [Option (2)] to the sure grade of D [Option (1)]?

I would want the chances of earning a C to be _____% before taking a chance.

Students were given all possible logical combinations (n=10) of G, G_0 and G_1. In an earlier pilot study (Sawyer, 1994), the lotteries were also presented as choice tasks. In the choice tasks, several values of p were displayed in a table, and students were asked to mark the value of p for which they were indifferent between the sure grade and the lottery. I found that students responded no better to the choice tasks than to the direct elicitation items, and that the resulting utility functions were similar.

Part 3 of the instructor questionnaires also elicited vN-M utilities. The only difference between Part 3 of the instructor questionnaires and Part 3 of the student questionnaire is that the lotteries in the instructor questionnaire were stated in terms of grade distributions, rather than probabilities. For example:

Suppose I asked you to choose between the following two scenarios in your course:
(1) A uniform result, in which all students perform at D level, or
(2) A mixed result, in which P percent of the students perform at C level, and all the rest perform at F level.

How large would the percentage P of students who perform at C level have to be before you would prefer the mixed result [Scenario (2)] to the uniform performance at D level [Scenario (1)]?

I would want P = _____ before selecting the mixed result.

Of course, the phenomenon described by the second item is the expected result of the phenomenon described by first item, but the two items are not, strictly speaking, asking the same question. I elected to use the second style, because it is more natural from the perspective of instructors.[3]

To reduce potential order effects in eliciting vN-M utilities, I created two forms of the student and instructor questionnaires. The two forms differed only in the order of the lotteries. For example, Item 1 of Form A pertained to the comparison D ~ [C, p, F] and Item 10 pertained to the comparison

[3] Instructors nevertheless objected to the hypothetical lotteries, because of their artificial quality.

B ~ [A, p, C]. In Form B, the items were presented in reverse order. I then "spiralled" the two forms (i.e., collated them before distributing them).

Remedial Instruction

Part 4 of the questionnaire considered the effectiveness of remedial instruction. If grades A-F in the standard course are assumed to define the final results of a student's involvement with a course placement system, then the outcome space consists of 10 elements:

$\Theta = \{ (S,A), (R,A), (S,B), (R,B), (S,C), (R,C), (S,D), (R,D), (S,F), (R,F) \}$,

where S denotes taking the standard course directly, R denotes taking the remedial course before taking the standard course, and A, B, C, D, F are the grades a student eventually earns in the standard course. Eliciting vN-M utilities for all these outcomes seemed, on its face, to be infeasible in a paper-and-pencil format. Even eliciting a value function for 10 outcomes seemed impractical. Therefore, I elected to elicit a value function for the outcomes associated with taking the remedial course, *relative* to the outcomes associated with taking the standard course directly. The values associated with taking the standard course directly were taken to be those elicited in Part 2 of the instrument.

Administration

The student questionnaires were distributed by instructors at the end of one class, and then collected at the beginning of subsequent classes. This method of administration is obviously vulnerable to self-selection effects; unfortunately, the instructors were not willing to give away instructional time for students to complete their questionnaires in class. Given the choice between potentially biased data, and no data at all, I acceded to the conditions demanded by the instructors.

The instructors completed their questionnaires in the same manner.

Analysis

For both the student data and the instructor data, I computed coherence indicators for the value functions and coherence rates for the vN-M utility function. I summarized the distribution of the

coherence indicators and rates, the value functions, and the vN-M utilities over the total group of respondents and over respondent subgroups.

Value Function Coherence Indicator

I computed a coherence indicator (denoted "CHRIND1") for the responses to Part 2. For an elicited grade value function *gvf*, CHRIND1=1 if $0 < gvf(D) < gvf(C) < gvf(B) < 100$, and CHRIND1=0 otherwise.

Calculating vN-M Utilities and Coherence Rates

The data from each comparison in Part 3 can be represented by a linear equation. For example, the data from the comparison C ~ [B, p, D] can be represented by the linear equation $u(C) = p * u(B) + (1 - p) * u(D)$. Because there were 10 comparisons in Part 3, and because there are three "unknowns" ($u(B)$, $u(C)$, and $u(D)$), the responses to Part 3 could result in a maximum of 10 linear equations in 3 unknowns. It can be shown that of the resulting $120 = \binom{10}{3}$ systems of 3 linear equations in 3 unknowns, only 108 are of full rank. (Appendix B contains a listing of the 108 full-rank systems.) Therefore, each respondent could, in principle, have 108 different solutions for $u(B)$, $u(C)$, and $u(D)$.

In practice, of course, a respondent might not provide equivalence probabilities p for all 10 comparisons, and so might have fewer than 108 sets of elicited utilities. For each respondent, all the systems of linear equations for which there were valid data were solved. Some of these resulting solutions were "coherent" in the sense that $0 < u(D) < u(C) < u(B) < 1$. Incoherent solutions resulted from inconsistent responses to the items (for example if a respondent reported probabilities $p_1 < p_2$ in the comparisons B ~ [A, p_1, F] and C ~ [A, p_2, F]). The "coherence rate" for a respondent was defined as the number of coherent solutions divided by 108. The coherence rate is an indicator of how well utilities can be elicited from a respondent. Note that the coherence *rate* (CR) pertains only to the vN-M utility, while the coherence *indica!* ·· pertains to the value function. For each respondent, I calculated the CR and the mean of the utility values $u(B)$, $u(C)$, and $u(D)$ associated with coherent solutions.

Course Placement Outcomes

For each respondent, I computed a summary value function for course placement outcomes. The summary value function measures a respondent's preferences for taking the remedial course or not, relative to different grades in the standard course. I computed the summary value function by combining the information elicited from Part 4 of the questionnaire with the grade value function information elicited from Part 2. For example, as before, let (R,A) denote the outcome that a student takes the remedial course before taking the standard course and earns an A; let (S,B) denote the outcome that a student takes the standard course directly and earns a B; and suppose that a respondent's choices in Part 4 of the questionnaire indicate that (R,A) $\succ$ (S,B). Then, a summary value function *svf* can be imputed by interpolating between 1=*gvf*(A) and *gvf*(B): *svf*(R,A)=.50 + .50**gvf*(B).

It is possible to do such imputation consistently provided the respondent's choices in Part 4 are coherent (consistent). For example, the following two choices are incoherent:

a. (R,A) $\succ$ (S,B)
b. (S,C) $\succ$ (R,A).

Of the 2^{10} = 1024 possible sequences of choices, only 14 are coherent. For each respondent, I computed an indicator CHCSEQ: CHCSEQ=1, if the respondent's choice sequence was coherent; and CHCSEQ=0, otherwise. See Appendix C for details on the coherent choice sequences.

The respondents for whom a coherent summary value function can be computed are those for whom CHRIND1=1 and CHCSEQ=1. I therefore computed a coherence indicator for the summary value function CHRIND2=CHCSEQ*CHRIND1.

Summary Statistics

Each respondent had a grade value function with its coherence indicator CHRIND1; mean elicited vN-M utilities with their associated coherence rate CR; and a summary value function with its associated coherence indicator CHRIND2. I summarized the distribution of these statistics over all respondents and over the respondents in each Group (as defined on pp. 13-14), using the minimum, median, and maximum value of each statistic.

Results

Response Rate

Of the 191 community college students in Group 1 who were given the opportunity to respond to the questionnaire, 129 (67%) did so. Of the 320 university students in Group 2 who were given the opportunity to respond, 141 (44%) did so. All 9 of the mathematics instructors in Group 3 responded to the questionnaire.

The response rates for both student groups (particularly, for Group 2) leave open the possibility for self-selection bias. In other words, the results obtained from the sample collected might have been affected by the characteristics of the students who were inclined to respond to questionnaires.

The sample of mathematics instructors, while not influenced by self-selection biases, was very small. The results for the instructors should therefore be interpreted with caution.

Coherence

Table 2 on the following page summarizes the distribution, by Group, of the coherence indicators (CHRIND1 and CHRIND2), and of the coherence rate (CR). About 2/3 of the overall group provided coherent grade value functions (CHRIND1). About 1/5 of the incoherent grade value functions were classified as incoherent because the respondents stipulated that *gvf*(D)=0; if this particular response had been classified as coherent, then about 3/4 of the respondents would have provided coherent grade value functions. Nearly all of the remaining incoherent responses were due to incomplete data (e.g., respondent provided *gvf*(B), and *gvf*(C), but not *gvf*(D)). The public university mathematics students responded coherently more often (75%) than the community college students (57%).

The median coherence rate for the vN-M utility function was .35. Although the maximum observed CR was .90, about 6% of the total group had a zero CR; and only about 20% of the respondents had a

Table 2.
Distribution of Coherence Indicators and Coherence Rate,
by Respondent Group

Respondent group	N	Grade value function coherence indicator	vN-M utility for grades coherence rate			Summary value function coherence indicator
		Pct. [CHRIND1=1]	Min.	Med.	Max.	Pct. [CHRIND2=1]
Community college students	129	57	.00	.30	.89	94
Public university math. students	141	75	.00	.44	.89	64
Public university math. instructors	9	89	.00	.35	.89	56
Total group	279	67	.00	.35	.90	46

CR > .50 (not shown in Table 2). The public university mathematics students again performed better (median CR=.44) than the community college students (median CR=.35).

About 78% of the respondents supplied coherent sequences of choices in Part 4. A coherent summary value function could be constructed for about 46% of respondents,

Elicited Utilities

Table 3 on the following page shows the distributions, for the different respondent groups, of the elicited grade value function and the vN-M utility function. Figure 1 on the page following Table 3 pictorially displays these results for the total group of respondents. Both Table 3 and Figure 1 are based on the 188 responses for which CHRIND1 > 0 (grade value function) or the 262 responses for which CR > 0 (vN-M utility).

Total group. The most apparent result for the total group is that the median vN-M utility was significantly higher than the median grade value function, particularly for the grades C and D. For example, the median grade value function for C was .50, and the median vN-M utility function for C was .75. This result is also true of all three respondent groups. This result is consistent with one obtained in an earlier pilot study (Sawyer, 1994).

The vN-M utility was also more variable over respondents than the grade value function. For example, the elicited vN-M utility for the grade C ranged from .01 to .98; but the grade value function for C ranged from .10 to .85. This result was also true of all three respondent groups.

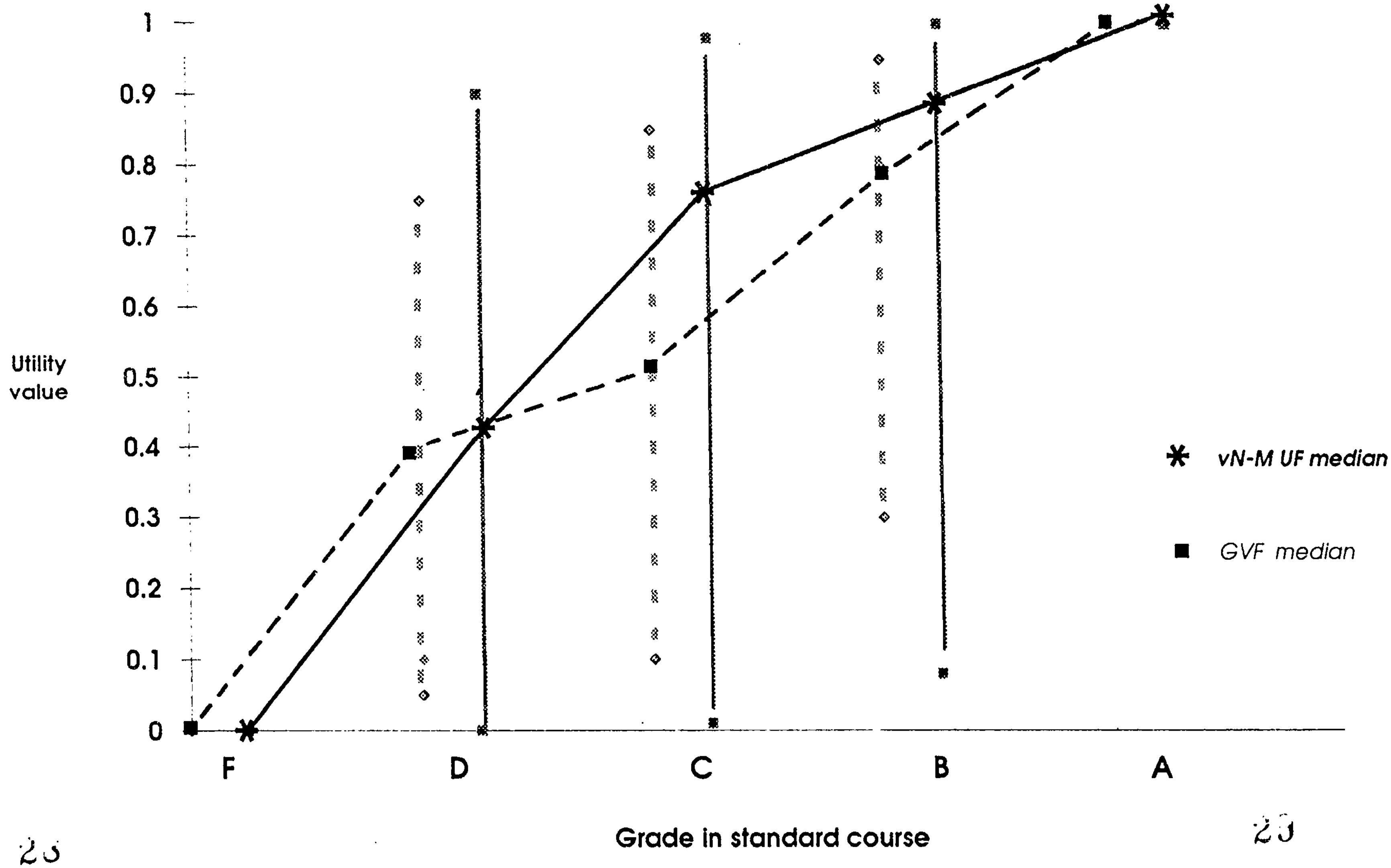
Suspecting that these two results might be related to the quality of individuals' responses to the hypothetical lotteries in Part 3 of the questionnaire, I studied the relationship between the difference $gvf(G) - vNM(G)$ and the coherence rate CR(G), for the grades G = B, C, and D. I found no relationship between these variables for any respondent group. Moreover, there was no relationship between these variables in any spiralled form of the questionnaire.

Group comparisons. The instructors' median grade value function was very similar to that of the students. The instructors' median vN-M utility was about .10 lower than that of the students. I found a similar result in the earlier pilot study (Sawyer, 1994).

Table 3.
Distribution of Elicited Preferences for Course Grades,
by Respondent Group and Method

Respondent group	Grade value function					vN-M utility function				
	N	Grade	Min.	Med.	Max.	N	Grade	Min.	Med.	Max.
Community college students	74	B	.30	.80	.95	119	B	.08	.90	.99
		C	.10	.50	.85		C	.01	.76	.98
		D	.05	.20	.75		D	.00	.50	.90
Public university math. students	106	B	.40	.78	.90	135	B	.24	.91	1.00
		C	.10	.40	.80		C	.02	.74	.98
		D	.05	.10	.60		D	.00	.47	.85
Public university math. instructors	8	B	.60	.80	.90	8	B	.33	.84	.92
		C	.40	.50	.60		C	.16	.60	.78
		D	.10	.25	.40		D	.08	.36	.54
Total group	188	B	.30	.80	.95	262	B	.08	.90	1.00
		C	.10	.50	.85		C	.01	.75	.98
		D	.05	.10	.75		D	.00	.48	.90

Figure 1.
Distribution of Elicited Grade Value Function and von Neumann-Morgenstern Utility Function



Summary Value Function

Recall that we are assuming that the outcomes of a course placement system are elements of the set $\Theta = \{ (S,A), (R,A), ..., (S,F), (R,F) \}$. Table 4 shows the distribution, over the total group of respondents, of the imputed summary value function CHRIND2 for this set. According to the medians in Table 4, the typical respondent would be as satisfied with enrolling in the standard course directly and earning a C, as he or she would with first taking the remedial course and earning an A or B in the standard course. Taking the remedial course and earning a C in the standard course was much less desirable; and taking the remedial course and earning a D in the standard course was valued hardly more that getting an F. The results for the instructors were similar to those of the students.

I also computed a hybrid "summary value function" by combining the elicited vN-M utility from Part 3 with the choice sequence in Part 4. I obtained result like those in Table 4, but shifted to higher numerical values.

Table 4.
Distribution of Summary Value Function for Course Placement Outcomes (Total Group)

Course placement outcome	Summary value function		
	Min.	Med.	Max.
(S,A)	1.00	1.00	1.00
(R,A)	.18	.54	.91
(S,B)	.40	.75	.95
(R,B)	.09	.55	.88
(S,C)	.10	.50	.85
(R,C)	.08	.28	.78
(S,D)	.05	.10	.75
(R,D)	.03	.05	.38
(S,F)	.01	.03	.19
(R,F)	.00	.00	.00

Note: Statistics are based on 152 cases with a coherent grade value function and a coherent choice sequence.

Discussion

As postsecondary institutions spend more resources on placing their first-year students into appropriate courses, they will be challenged to document the effectiveness of their placement systems. Evaluating complex systems requires presenting evidence on multiple measures. In course placement, for example, one could document per-student costs of testing and remedial instruction; survey students, faculty, and staff on their satisfaction with various components of the system; present statistics on success rates; etc. Decision theory provides another way to develop an indicator of program effectiveness: If the expected utility associated with the cutoff score on a placement variable significantly exceeds the expected utility associated with the minimum possible score, then one has evidence that the placement system is accruing a net benefit to its users.

About half of the university students and about two-thirds of the community college students surveyed completed the questionnaire. Of those who completed the questionnaire, about 2/3 provided enough information to develop a coherent grade value function, and slightly less than half provided enough information to develop a coherent summary value function. Although 94% of the respondents had a positive, vN-M coherence rate, the typical respondent provided enough information to elicit only 35% of the possible vN-M utility functions. These results suggest that institutions can not realistically expect to elicit utilities by routinely administering paper-and-pencil questionnaires to their instructors and students (see Question 1, p. 16). Institutions would instead need to provide special instruction and motivation to elicit valid utilities for most students and instructors. An interactive computer elicitation program might increase validity by giving respondents an opportunity to correct inconsistencies. The difficulty and expense of implementing such a computer program, however, would seem to make it unattractive to institutions, even today.

A more realistic strategy would be to elicit utilities for different groups of people at a variety of institutions, and attempt to make some kind of generalizations. For example, we might find that the summary value functions of students in community colleges fall into two general clusters, say for "risk-taking" and "risk-averse" students. Other community colleges could apply one or both of these summary value functions in developing their expected utility indicators.

The results of this study also suggest that there is a pronounced methodological effect on elicited utilities. The vN-M utility function was systematically higher than the grade value function (Question 2). Moreover, this difference

transcended differences on any other variables I investigated. One possible explanation for this result is that hypothetical lotteries bring out risk aversion in people: people will, for example, demand a very probability of an A before trading a certain grade of B for a lottery in which they might earn an A or an F. The method one prefers depends partly on one's philosophical orientation and partly on the intended use of the model: An orthodox Bayesian decision theorist would deny that the concept of utility has any meaning outside the context of probability, whereas most of the respondents in this study complained about the artificial quality of the vN-M lotteries. If the major purpose for eliciting a preference function is to develop weights for an indicator of overall program effectiveness that would be reported along with more specific indicators, then there would seem to be little accuracy lost (and maybe some to be gained) in using a value function.

Figure 1 suggests that neither the grade value function nor the vN-M utility function are well approximated by step functions. Therefore, the threshold utility is not a very accurate description of most people's preferences (Question 3). This result is a pity, because the threshold model is much simpler to work with mathematically and to explain!

Within the limitations of the data in this study, neither the grade value functions nor the summary value functions of students differ significantly from those of instructors. This conclusion should be considered more tentative than the others, given the small number of instructors.

Conclusions

A college course placement system consists of an assessment component and an instructional component. The effectiveness of the system as a whole depends on both components. Statistical decision theory can be used to describe the possible outcomes of course placement systems. By eliciting a preference function of the outcomes, and by averaging the function with respect to a probability distribution, one can evaluate the effectiveness of the course placement system and select optimal cutoff scores. Preference functions may be categorized according to whether they are deterministic (value functions) or stochastic (vN-M utility functions).

In a study at a midwestern community college and public university, about 2/3 of the respondents were able to supply coherent grade value functions, and slightly less than half were able to supply coherent summary value functions for course placement outcomes. Performance on hypothetical lotteries used to elicit vN-M utilities varied significantly

among individuals, but most respondents were able to supply enough information to calculate at least one utility function. The vN-M utility values were typically much larger than the grade function values. The median results for the instructors and students were very similar.

Future research

In fall 1995 I hope to elicit preference functions for large samples of students, instructors, and support staff at several institutions. I will revise the questionnaire to elicit only value functions I will attempt to obtain duplicate measurements for many of the respondents, so that I can estimate reliabilities.

I also hope to collect data on the placement variables and course grades of the students. By combining these two data sets, I will obtain evidence about the perceived effectiveness of the placement systems at these institutions.

References

Ben-Shakhar, G., Kiderman, I., & Beller, M. (1994). *Comparing the utility of two procedures for admitting students to liberal arts: An application of decision theoretic models.* Jerusalem: National Institute for Testing and Evaluation.

The Carnegie Foundation for the Advancement of Teaching (1988). *An imperiled generation: Saving urban schools*. Princeton, NJ: Princeton University Press.

Carriuolo, N. (1994, April 13). Why developmental education is such a hot potato. *The Chronicle of Higher Education*, Sec. 2, pp. 1-2.

College Entrance Examination Board (1977). *On further examination*. New York: Author.

Colleges and universities offering remedial instruction and tutoring. (1994, April 13). *Education Week*, XIII, No. 29, p.6.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana, Illinois: University of Illinois Press.

DeGroot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

Falmagne, J.-C. (1985). *Elements of psychophysical theory*. (Oxford Psychology Series No. 6). New York: Oxford University Press.

Farquhar, P.H. (1984). Utility assessment methods. *Management Science*, 30(11), 1283-1300.

Hancher, M. (1994, Feb. 2). The consequences of graue inflation. [Letter to the editor]. *Chronicle of Higher Education*, p. B-5.

Houston, W. M. (1993, April). *Accuracy of validity indices for course placement systems*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Isaacs, G. L. & Novick, M. R. (1978). *Manual for the computer-assisted data analysis (CADA) monitor*. Iowa City, Iowa: The University of Iowa Dept. of Educational and Psychological Foundations.

Lindley, D. V. (1972). *Bayesian statistics: A review*. (Regional Conferences Series in Applied Mathematics.) Philadelphia: Society for Industrial and Applied Mathematics.

Lively, K. (1993, Feb. 24). States step up efforts to end remedial courses at 4-year colleges. *The Chronicle of Higher Education*, p. A-28.

Mac Donald, H. (1994). Downward mobility. The failure of open admissions at City University. *City Journal*, Summer 1994, pp. 10-20.

McNabb, T. (1990). *Course placement practices of American postsecondary institutions*. (ACT Research Report Series 90-10). Iowa City, Iowa: American College Testing.

Munday, L. (1976). *Declining admissions test scores.* (ACT Research Report No. 71). Iowa City, Iowa: American College Testing.

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform.* (USED Stock No. 065-000-00177-2). Washington, DC: U.S. Government Printing Office.

Novick, M. R. & Lindley, D. V. (1979). Fixed-state assessment of utility functions. *Journal of the American Statistical Association*, *74*, 306-311.

Petersen, N.S. & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, *13*(2), 3-29.

Sawyer, R. (1994, April). Eliciting Utility Functions for Validating Course Placement Systems. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Sawyer, R. (in press). Decision theory models for validating course placement tests. *Journal of Educational Measurement.*

Schiel, J., & Noble, J. (1993, April). *The effects of data truncation on estimated validity indices for course placement.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

Shea, C. (1994, Jan. 5). Grade Inflation's Consequences. *The Chronicle of Higher Education*, p. A-45.

Singal, D. J. (1991, November). The other crisis in American education. *The Atlantic, 268*, pp. 59-69.

van der Linden, W. (1991). Applications of decision theory to test-based decision making. In Hambleton, R. L. & Zaal, J. N. (Eds.), *Advances in educational and psychological testing.* Boston: Kluwer.

Willingham, W. (1974). *College placement and exemption.* New York: College Entrance and Examination Board.

Woods, J. E. (1985). *Status of testing practices at two-year postsecondary institutions.* Iowa City, Iowa: American College Testing and the American Association of Community and Junior Colleges.

Wright, D. A. & Cahalan, M. W. (1985, April). *Remedial/development studies in institutions of higher education policies and practices, 1984.* Paper presented at the annual conference of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 263-828).

Whitney, D. R. (1989). Educational admissions and placement. In Linn, R. L. (Ed.), *Educational measurement* (3rd ed.). (pp. 515-526). New York: ACE and Macmillan.

Yates, J. F. (1990). *Judgment and decision making.* Englewood Cliffs, NJ: Prentice-Hall.29

Appendix A

Questionnaires

Students' Satisfaction With Grades and Course Placement Decisions

A research project by Richard Sawyer, ACT, with the cooperation of Midwest Community College

Fall, 1994

Purpose of this Study

There are benefits, risks, and costs associated with the decision to go to college. Part of ACT's work involves helping students decide which courses to take. I want to learn about the things you think about in making decisions about your courses.

I will ask you some questions about your academic work at Midwest, and about your preferences for different grades and course placement decisions. This questionnaire is not a test --- there are no right or wrong answers. I do not ask you for your name or other identifying information on the questionnaire, so your answers will be anonymous. I have written a number on the top of this page, but only to help me keep track of which questionnaires are given to which classes.

Your instructor will distribute this questionnaire in class. Please take it home, answer the questions, and then bring it back to the next class meeting, where your instructor will collect it.

There is a chance that you will receive a questionnaire from two different instructors. If you do receive two questionnaires, then complete only the first one you receive. At the top of the second questionnaire, just write "Second" and return it to the instructor at the next class meeting.

The questions are grouped into four parts. As soon as you finish one part, please continue on directly to the next part. The entire questionnaire should take 15-30 minutes to complete.

The information you give will be used to enhance the services ACT provides to students in the future. I sincerely appreciate your cooperation.

Part 1
Background Information

1. Please check (✓) the appropriate boxes to indicate whether you have taken, or are currently enrolled in, any of the courses in the table below. Also indicate either the grade you received in the course (if you have already taken it), or the grade you expect to receive (if you are currently enrolled in the course).

Course	Check here if you have already taken	Grade you received	Check here if you are currently enrolled	Grade you expect to receive
PA Reading				
PA Writing				
Elements of Writing				
Composition I				
Composition II				
College Writing				
PA Mathematics				
Beginning Algebra				
Intermediate Algebra				
Finite Mathematics				
Pre Calculus				
Mathematics for Dec. Making				
Statistical Ideas				
Other mathematics courses (please specify):				

2. What general program or major are you enrolled in at Midwest?

____

3. When did you first start taking courses at Midwest?

______________ (month and year)

4. When did you first enroll in your *current* program at Midwest?

______________ (month and year)

5. What is your gender?

___ Female

___ Male

6. What is your age?

_____ years

7. Which of the following statements best describes your goals about the grades you earn in courses at Midwest? *(Check one only.)*

___ I don't mind earning a few Ds, so long as I receive credit for all my courses.

___ It is important for me to earn only As, Bs, or Cs in my courses.

___ It is important for me to earn only As or Bs in my courses.

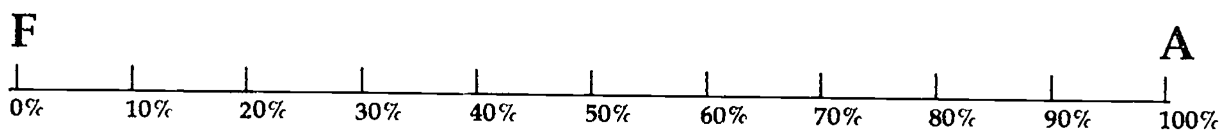
___ It is important for me to earn all As in my courses.

Part 2
Course Grades

Students want to earn as high a grade in a course as they can. Naturally, everyone would be more satisfied with an A than with a B, or with a B than with a C, and so forth --- but what about your *relative* satisfaction? Would you, for example, feel twice as satisfied with an A as with a B?

I want to find out your relative satisfaction with grades in the standard courses you take. (A "standard course" is a for-credit course that you need to pass to satisfy the requirements of your program at Midwest.) In answering the questions, please think of a standard course that is typical of those you are taking or have taken.

The line below is meant to suggest your relative satisfaction with the different letter grades. The letter grade of F is associated with 0% satisfaction, and the letter grade of A is associated with 100% satisfaction:



Please indicate on this line your relative satisfaction with the grades of B, C, and D by writing them above an appropriate point on the line. For example, if you would be about half as satisfied with a B as with an A, then you would write a "B" above the 50% mark.

NOTE: Your responses should reflect your <u>satisfaction</u> with particular grades in a standard course. Your responses do not have to correspond to a percent-correct grading scale (for example, where the grade A represents 90% or more correct).

Part 3
Course Grades (cont'd)

Sometimes we are given a choice between receiving a certain prize for sure, or else taking a chance on winning a better prize. For example, a contestant on a television game show may be given the choice either of winning a fancy color television for sure, or else having a 50% chance at winning $2,000 cash (and a 50% chance of winning nothing). If the contestant already has a color television, then he or she might be willing to take a 50% chance at winning $2,000 cash. On the other hand, the contestant might choose the certain prize of the color television if he or she doesn't have one.

The following ten questions ask about your satisfaction with different grades in this way. Each question asks you to think about either earning a certain grade for sure [Option (1)], or else taking a chance, in which you might earn a higher grade, but also might earn a lower grade [Option (2)].

In answering these questions, please think of any course that you need to pass to satisfy the requirements of your program at Midwest.

1. Suppose I offered you the choice of:
   (1) Earning exactly a grade of D for sure, or
   (2) Taking a chance, where you could earn either a C or an F.

   How large would the chances of earning a C have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of D [Option *(1)*]?

   I would want the chances of earning a C to be ______% before taking a chance.

2. Suppose I offered you the choice of:
   (1) Earning exactly a grade of D for sure, or
   (2) Taking a chance, where you could earn either a **B** or an F.

   How large would the chances of earning a **B** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of D [Option *(1)*]?

   I would want the chances of earning a B to be ______% before taking a chance.

3. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **D** for sure, or
   (2) Taking a chance, where you could earn either an **A** or an F.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **D** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

4. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **C** for sure, or
   (2) Taking a chance, where you could earn either a **B** or an F.

   How large would the chances of earning a **B** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

   I would want the chances of earning a B to be ______% before taking a chance.

5. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **C** for sure, or
   (2) Taking a chance, where you could earn either a **B** or a **D**.

   How large would the chances of earning a **B** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

   I would want the chances of earning a B to be ______% before taking a chance.

6. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **C** for sure, or
   (2) Taking a chance, where you could earn either an **A** or an F.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

7. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **C** for sure, or
   (2) Taking a chance, where you could earn either an **A** or a **D**.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

8. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **B** for sure, or
   (2) Taking a chance, where you could earn either an **A** or an **F**.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **B** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

9. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **B** for sure, or
   (2) Taking a chance, where you could earn either an **A** or a **D**.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **B** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

10. Suppose I offered you the choice of:
    (1) Earning exactly a grade of **B** for sure, or
    (2) Taking a chance, where you could earn either an **A** or a **C**.

    How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **B** [Option *(1)*]?

    I would want the chances of earning an A to be ______% before taking a chance.

Part 4
Course Placement

Let a "standard course" be a for-credit course that is required for your program. For example, many entering students may need to pass *Composition I* to satisfy the requirements of their programs at Midwest.

One purpose of a course placement system is to determine whether a student is ready to take a particular standard course. If a student is not ready to take the standard course, he or she can instead enroll in a "developmental course" to acquire the skills needed to succeed in the standard course. At Midwest, for example, *Elements of Writing* would be considered the developmental course for *Composition I.*

Taking a developmental course will tend to increase a student's chances of success in the standard course. However, taking a developmental course also has disadvantages---it increases the time required to complete your program, and it costs additional money. Therefore, the decision to take a developmental course involves a trade-off: an increased chance of eventually succeeding in the standard course, versus extra time and money.

I want to find out how *you* see these trade-offs.

The table below presents different situations in which you are asked to choose between either taking a developmental course *before* taking the standard course [Col. (1)], or *directly* enrolling in the standard course [Col. (2)]. Assume that the developmental course is 1 semester in length, and carries no program credit. (Some developmental courses at Midwest carry elective credit, and some carry no credit.)

For each situation, please check (✓) either the box in Col. (1) or the box in Col. (2), according to your preference:

Situation	Which would you prefer ?	
	Col. (1) Take the developmental course *before* taking the standard course. Then, earn this grade in the standard course:	**Col. (2)** Enroll directly in the standard course, and earn this grade:
1	A	B
2	A	C
3	A	D
4	A	F
5	B	C
6	B	D
7	B	F
8	C	D
9	C	F
10	D	F

***** This is the end of the questionnaire. Thank you for your help! *****

Students' Satisfaction With Grades and Course Placement Decisions

**A research project by Richard Sawyer, ACT,
with the cooperation of Midwest Public University
Fall, 1994**

Purpose of this Study

There are benefits, risks, and costs associated with the decision to go to college. Part of ACT's work involves helping students decide which courses to take. I want to learn about the things you think about in making decisions about your courses.

I will ask you some questions about your academic work at Midwest Public University, and about your preferences for different grades and course placement decisions. This questionnaire is not a test --- there are no right or wrong answers. I do not ask you for your name or other identifying information on the questionnaire, so your answers will be anonymous. I have written a number on the top of this page, but only to help me keep track of which questionnaires are given to which classes.

Your instructor will distribute this questionnaire in class. Please take it home, answer the questions, and then bring it back to the next class meeting, where we will collect it.

The questions are grouped into four parts. As soon as you finish one part, please continue on directly to the next part. The entire questionnaire should take about 15 minutes to complete.

The information you give will be used to enhance the services ACT provides to students in the future. I sincerely appreciate your cooperation.

Part 1
Background Information

1. Please check (✓) the appropriate boxes to indicate whether you have taken, or are currently enrolled in, any of the courses in the table below. Also indicate either the grade you received in the course (if you have already taken it), or the grade you expect to receive (if you are currently enrolled in the course).

Course	Check here if you have already taken	Grade you received	Check here if you are currently enrolled	Grade you expect to receive
Basic Algebra II				
Elementary Functions				
Quantitative Methods I				
Calculus I				
Other mathematics courses (please specify):				

2. What is your major at Midwest Public University? (If you have not yet selected a major, please write "None.")

____

3. When did you first start taking courses at Midwest Public University?

_______________ (month and year)

4. When did you first select your *current* major at Midwest Public University?

_______________ (month and year)

5. What is your gender?

___ Female

___ Male

6. What is your age?

_____ years

7. Which of the following statements best describes your goals about the grades you earn in courses at Midwest Public University? *(Check one only.)*

____ I don't mind earning a few Ds, so long as I receive credit for all my courses.

____ It is important for me to earn only As, Bs, or Cs in my courses.

____ It is important for me to earn only As or Bs in my courses.

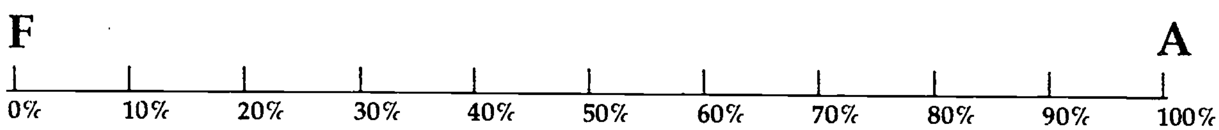
____ It is important for me to earn all As in my courses.

Part 2
Course Grades

Students want to earn as high a grade in a course as they can. Naturally, everyone would be more satisfied with an A than with a B, or with a B than with a C, and so forth --- but what about your *relative* satisfaction? Would you, for example, feel twice as satisfied with an A as with a B?

I want to find out your relative satisfaction with grades in the courses you take. In answering the questions, please think of any course that you need to pass to satisfy the requirements of your major at Midwest Public University.

The line below is meant to suggest your relative satisfaction with the different letter grades. The letter grade of F is associated with 0% satisfaction, and the letter grade of A is associated with 100% satisfaction:



Please indicate on this line your relative satisfaction with the grades of B, C, and D by writing them above an appropriate point on the line. For example, if you would be about half as satisfied with a B as with an A, then you would write a "B" above the 50% mark.

NOTE: Your responses should reflect your satisfaction with particular grades in a standard course. Your responses do not have to correspond to a percent-correct grading scale (for example, where the grade A represents 90% or more correct).

Part 3
Course Grades (cont'd)

Sometimes we are given a choice between receiving a certain prize for sure, or else taking a chance on winning a better prize. For example, a contestant on a television game show may be given the choice either of winning a fancy color television for sure, or else having a 50% chance at winning $2,000 in cash (and a 50% chance of winning nothing). If the contestant already has a color television, then he or she might be willing to take a 50% chance at winning $2,000 in cash. On the other hand, the contestant might choose the certain prize of the color television if he or she doesn't have one.

The following ten questions ask about your satisfaction with different grades in this way. Each question asks you to think about either earning a certain grade for sure [Option (1)], or else taking a chance, in which you might earn a higher grade, but also might earn a lower grade [Option (2)].

In answering these questions, please think of any course that you need to pass to satisfy the requirements of your major at Midwest Public University.

1. Suppose I offered you the choice of:
   (1) Earning exactly a grade of D for sure, or
   (2) Taking a chance, where you could earn either a C or an F.

   How large would the chances of earning a C have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of D [Option *(1)*]?

   I would want the chances of earning a C to be ______% before taking a chance.

2. Suppose I offered you the choice of:
   (1) Earning exactly a grade of D for sure, or
   (2) Taking a chance, where you could earn either a B or an F.

   How large would the chances of earning a B have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of D [Option *(1)*]?

   I would want the chances of earning a B to be ______% before taking a chance.

3. Suppose I offered you the choice of:
(1) Earning exactly a grade of **D** for sure, or
(2) Taking a chance, where you could earn either an **A** or an F.

How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **D** [Option *(1)*]?

I would want the chances of earning an A to be ______% before taking a chance.

4. Suppose I offered you the choice of:
(1) Earning exactly a grade of **C** for sure, or
(2) Taking a chance, where you could earn either a **B** or an F.

How large would the chances of earning a **B** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

I would want the chances of earning a B to be ______% before taking a chance.

5. Suppose I offered you the choice of:
(1) Earning exactly a grade of **C** for sure, or
(2) Taking a chance, where you could earn either a **B** or a **D**.

How large would the chances of earning a **B** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

I would want the chances of earning a B to be ______% before taking a chance.

6. Suppose I offered you the choice of:
(1) Earning exactly a grade of **C** for sure, or
(2) Taking a chance, where you could earn either an **A** or an F.

How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

I would want the chances of earning an A to be ______% before taking a chance.

7. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **C** for sure, or
   (2) Taking a chance, where you could earn either an **A** or a **D**.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **C** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance

8. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **B** for sure, or
   (2) Taking a chance, where you could earn either an **A** or an **F**.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **B** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

9. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **B** for sure, or
   (2) Taking a chance, where you could earn either an **A** or a **D**.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **B** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

10. Suppose I offered you the choice of:
   (1) Earning exactly a grade of **B** for sure, or
   (2) Taking a chance, where you could earn either an **A** or a **C**.

   How large would the chances of earning an **A** have to be before you would prefer taking a chance [Option *(2)*] to the sure grade of **B** [Option *(1)*]?

   I would want the chances of earning an A to be ______% before taking a chance.

Part 4
Course Placement

Let a "standard course" be a for-credit course that is required for your major. For example, some students may need to pass *Quantitative Methods I* to satisfy the requirements of their major at Midwest Public University.

One purpose of a course placement system is to determine whether a student is ready to take a particular standard course. If a student is not ready to take the standard course, he or she can instead enroll in a "developmental course" to acquire the skills needed to succeed in the standard course. At Midwest Public University, for example, *Basic Algebra II* would be considered a developmental course for the standard course *Quantitative Methods I.*

Taking a developmental course will tend to increase a student's chances of success in the standard course. However, taking a developmental course also has disadvantages---it increases the time required to complete your program, it costs additional money, and it may not carry credit toward your degree. Therefore, the decision to take a developmental course involves a trade-off: an increased chance of eventually succeeding in the standard course, versus extra time and money.

I want to find out how ***you*** see these trade-offs.

The table below presents different situations in which you are asked to choose between either taking a developmental course *before* taking the standard course [Col. (1)], or *directly* enrolling in the standard course [Col. (2)]. Assume that the developmental course is 1 semester in length.

For each situation, please check (✓) either the box in Col. (1) or the box in Col. (2), according to your preference:

Situation	Which would you prefer ? Col. (1) Take the developmental course *before* taking the standard course. Then, earn this grade in the standard course:	Col. (2) Enroll directly in the standard course, and earn this grade:
1	A	B
2	A	C
3	A	D
4	A	F
5	B	C
6	B	D
7	B	F
8	C	D
9	C	F
10	D	F

*** This is the end of the questionnaire. Thank you for your help! ***

Mathematics Instructors' Satisfaction With Grades and Course Placement Decisions

A research project by Richard Sawyer, ACT

Fall, 1994

Purpose of this Study

There are benefits, risks, and costs associated with the decision to take any college course. Part of ACT's work involves helping students decide which courses to take. I want to investigate mathematics instructors' satisfaction with the results of course placement decisions.

I will ask you some questions about your teaching responsibilities at Midwest Public University and about your satisfaction with different levels of student performance. This questionnaire is not a test---there are no right or wrong answers. Because I do not ask you for your name or other identifying information on the questionnaire, your answers will be anonymous.

The questions are grouped into four parts. As soon as you finish one part, please continue on directly to the next part. The entire questionnaire should take about 15 minutes to complete. I will collect your completed questionnaire, along with those of your students at your next class meeting.

The information you give will be used to enhance the services ACT provides to faculty and students in the future. I sincerely appreciate your cooperation.

Part 1
Background Information

1. Please check (✓) the appropriate boxes to indicate whether you are currently teaching, or have taught, any of the courses in the table below. If you have taught the course, please estimate roughly the percentages of different grades your students earned. Naturally, grades vary from term to term, depending on your students' performance; please try to approximate what the *typical* percentages are.

Course	Check here if you are now teaching	Check here if you have taught	Approximate percentage of students who earned . . .				
			A	B	C	D	F
Basic Algebra II							
Elementary Functions							
Quantitative Methods I							
Calculus I							

2. When did you first start teaching courses at Midwest Public University?

________________ (year)

3. Please check (✓) the appropriate boxes to indicate how important the following factors are to you in awarding grades.

Factor	Importance in awarding grades			
	Very important	Important	Somewhat important	Not important
Academic performance (as measured by tests, essays, homework, etc.)				
Attendance and participation in class				
Motivation and effort				
Other characteristics (please describe):				

4. Which of the following statements best describes your policy in assigning grades? *(Check one only.)*

____ I grade strictly according to fixed standards of student performance. Therefore, I could (at least in principle) assign all As or all Fs.

____ I grade strictly "on a curve": I always assign a certain percentag: of As, a certain percentage of Bs, etc.

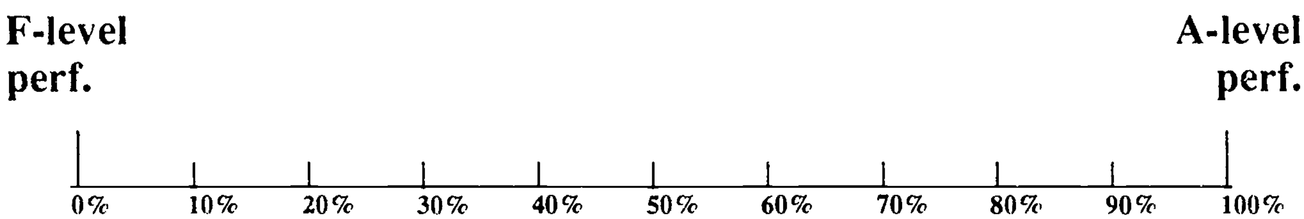
____ I grade mostly according to a fixed standard, but I may modif/ some grades so that the distribution of grades meets a target grade distribution.

____ I grade mostly "on a curve," but I may modify some grades if students' performance merits doing so.

Part 2
Course Grades

I want to find out your relative satisfaction with different levels of student performance in the courses you teach. In responding to the questions, please think of the course you now teach that is often taken by first-year students.

Instructors award grades on the basis of their students' academic achievement and other performance characteristics. The line below is meant to represent your satisfaction with student performance characteristics that would result in your assigning different letter grades. To simplify the discussion, I have associated F-level performance with 0% satisfaction, and A-level performance with 100% satisfaction:



Please indicate on this line your relative satisfaction with B-level, C-level, and D-level performance by marking the letters "B", "C", and "D" at appropriate points above the line. For example, if you feel about half as satisfied with the performance of a student who earns a B as you do with the performance of a student who earns an A, then you would write a "B" above the 50% mark.

Please note that your responses should reflect your *satisfaction* with particular levels of student performance. Your responses need not correspond to a percent-correct grading scale (where, for example, an A represents 90% or more correct).

Part 3
Course Grades (cont'd)

The following questions are also related to your satisfaction with different levels of student performance. When responding to the questions, please think of the course you now teach that is often taken by first-year students.

In each question, you are asked to choose between two hypothetical scenarios involving student performance in your course: In Scenario *(1)*, all students perform at exactly the same level (e.g., every student performs at a level to which you would assign a grade of D). In Scenario *(2)*, a certain percentage of students perform at a higher level, and all the rest perform at lower level (e.g, 75% perform at C level, and 25% perform at F level). Neither of these scenarios is realistic; but, by comparing your responses to different questions, I can estimate your *relative* satisfaction with different levels of student performance.

1. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at D level, or
   (2) A mixed result, in which **P** percent of the students perform at **C** level, and all the rest perform at **F** level.

   How large would the percentage **P** of students who perform at **C** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at **D** level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

2. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at D level, or
   (2) A mixed result, in which **P** percent of the students perform at **B** level, and all the rest perform at **F** level.

   How large would the percentage **P** of students who perform at **B** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at **D** level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

3. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at D level, or
   (2) A mixed result, in which **P** percent of the students perform at **A** level, and all the rest perform at **F** level.

   How large would the percentage **P** of students who perform at **A** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at **D** level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

4. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at C level, or
   (2) A mixed result, in which **P** percent of the students perform at **B** level, and all the rest perform at **F** level.

   How large would the percentage **P** of students who perform at **B** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at C level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

5. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at C level, or
   (2) A mixed result, in which **P** percent of the students perform at **B** level, and all the rest perform at **D** level.

   How large would the percentage **P** of students who perform at **B** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at C level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

6. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at C level, or
   (2) A mixed result, in which **P** percent of the students perform at **A** level, and all the rest perform at **F** level.

   How large would the percentage **P** of students who perform at **A** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at C level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

7. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at C level, or
   (2) A mixed result, in which **P** percent of the students perform at **A** level, and all the rest perform at **D** level.

   How large would the percentage **P** of students who perform at **A** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at C level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

8. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at **B** level, or
   (2) A mixed result, in which **P** percent of the students perform at **A** level, and all the rest perform at **F** level.

   How large would the percentage **P** of students who perform at **A** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at **B** level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

9. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at **B** level, or
   (2) A mixed result, in which **P** percent of the students perform at **A** level, and all the rest perform at **D** level.

   How large would the percentage **P** of students who perform at **A** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at **B** level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

10. Suppose I asked you to choose between the following two scenarios in your course:
   (1) A uniform result, in which all students perform at **B** level, or
   (2) A mixed result, in which **P** percent of the students perform at **A** level, and all the rest perform at **C** level.

   How large would the percentage **P** of students who perform at **A** level have to be before you would prefer the mixed result [Scenario *(2)*] to the uniform performance at **B** level [Scenario *(1)*]?

   I would want P = ______ before selecting the mixed result.

Part 4
Course Placement

Background information.
Let a "standard course" be a for-credit course that is taken by well-prepared entering students, and that is required for a major. For example, well-prepared entering students may take Quantitative Methods I and may need to pass it to satisfy the requirements of some majors at Midwest Public University.

One purpose of a course placement system is to determine whether a student is ready to take a particular standard course. If a student is not ready to take the standard course, he or she can instead enroll in a "developmental course" to acquire the skills needed to succeed in the standard course. At Midwest Public University, for example, Basic Algebra II would be considered a developmental course for the standard course Quantitative Methods I.

Taking a developmental course will tend to increase a student's chances of success in the standard course. However, taking a developmental course also has disadvantages: It increases the time required to complete a program, it costs additional money, and it may not carry credit toward a degree. Some students may be discouraged from even starting a program if they have to take developmental courses. Therefore, the decision to take a developmental course involves a trade-off: an increased chance of eventually succeeding in the standard course versus extra time and money.

I want to find out how you see these trade-offs.

The table below presents different situations in which you are asked to choose between the student's taking a developmental course *before* taking the standard course [Col. (1)], or the student's enrolling *directly* in the standard course [Col. (2)]. Assume that the developmental course [Col. (1)] is 1 semester in length.

For each situation, please check (✓) either the box in Col. (1) or the box in Col. (2), according to your preference:

	Which do you prefer ?	
Situation	**Col. (1)** **The student takes the developmental course *before* taking the standard course.** **Then, the student earns this grade in the standard course:**	**Col. (2)** **The student enrolls *directly* in the standard course, and earns this grade:**
1	A	B
2	A	C
3	A	D
4	A	F
5	B	C
6	B	D
7	B	F
8	C	D
9	C	F
10	D	F

Appendix B

Eliciting von Neumann-Morgenstern Utilities for Course Grades

Table 1.
Comparisons of Sure Events and Lotteries, and Their Associated Linear Equations

	Comparison		Linear Equation
1	B ~ [A, p_1, F]	1	$u(B) = p_1$
2	B ~ [A, p_2, D]	2	$u(B) = p_2 + (1-p_2)\,u(D)$
3	B ~ [A, p_3, C]	3	$u(B) = p_3 + (1-p_3)\,u(C)$
4	C ~ [A, p_4, F]	4	$u(C) = p_4$
5	C ~ [A, p_5, D]	5	$u(C) = p_5 + (1-p_5)\,u(D)$
6	C ~ [B, p_6, F]	6	$u(C) = p_6\,u(B)$
7	C ~ [B, p_7, D]	7	$u(C) = p_7\,u(B) + (1-p_7)u(D)$
8	D ~ [A, p_8, F]	8	$u(D) = p_8$
9	D ~ [B, p_9, F]	9	$u(D) = p_9\,u(B)$
10	D ~ [C, p_{10}, F]	10	$u(D) = p_{10}\,u(C)$

Table 2.
Systems of Full-Rank Linear Equations

System	Equations*	System	Equations*	System	Equations*
1	1, 2, 3	26	1, 7, 8	51	2, 6, 9
2	1, 2, 4	27	1, 7, 9	52	2, 6, 10
3	1, 2, 5	28	1, 7, 10	53	2, 7, 8
4	1, 2, 6	29	1, 8, 10	54	2, 7, 10
5	1, 2, 7	30	1, 9, 10	55	2, 8, 10
6	1, 2, 10	31	2, 3, 4	56	2, 9, 10
7	1, 3, 5	32	2, 3, 5	57	3, 4, 5
8	1, 3, 7	33	2, 3, 6	58	3, 4, 7
9	1, 3, 8	34	2, 3, 7	59	3, 4, 8
10	1, 3, 9	35	2, 3, 8	60	3, 4, 9
11	1, 3, 10	36	2, 3, 9	61	3, 4, 10
12	1, 4, 5	37	2, 3, 10	62	3, 5, 6
13	1, 4, 7	38	2, 4, 5	63	3, 5, 7
14	1, 4, 8	39	2, 4, 6	64	3, 5, 8
15	1, 4, 9	40	2, 4, 7	65	3, 5, 9
16	1, 4, 10	41	2, 4, 8	66	3, 5, 10
17	1, 5, 6	42	2, 4, 9	67	3, 6, 7
18	1, 5, 7	43	2, 4, 10	68	3, 6, 8
19	1, 5, 8	44	2, 5, 6	69	3, 6, 9
20	1, 5, 9	45	2, 5, 7	70	3, 6, 10
21	1, 5, 10	46	2, 5, 8	71	3, 7, 8
22	1, 6, 7	47	2, 5, 9	72	3, 7, 9
23	1, 6, 8	48	2, 5, 10	73	3, 7, 10
24	1, 6, 9	49	2, 6, 7	74	3, 8, 9
25	1, 6, 10	50	2, 6, 8	75	3, 8, 10

* Equations enumerated in Table 1.

(continued on next page)

Table 2 (cont'd.)
Systems of Full-Rank Linear Equations

System	Equations*	System	Equations*
76	3, 9, 10	101	6, 7, 10
77	4, 5, 6	102	6, 8, 9
78	4, 5, 7	103	6, 8, 10
79	4, 5, 9	104	6, 9, 10
80	4, 5, 10	105	7, 8, 9
81	4, 6, 7	106	7, 8, 10
82	4, 6, 8	107	7, 9, 10
83	4, 6, 9	108	8, 9, 10
84	4, 6, 10		
85	4, 7, 8		
86	4, 7, 9		
87	4, 7, 10		
88	4, 8, 9		
89	4, 9, 10		
90	5, 6, 7		
91	5, 6, 8		
92	5, 6, 9		
93	5, 6, 10		
94	5, 7, 8		
95	5, 7, 9		
96	5, 7, 10		
97	5, 8, 9		
98	5, 9, 10		
99	6, 7, 8		
100	6, 7, 9		

* Equations enumerated in Table 1.

Appendix C

Constructing Summary Value Functions for Course Placement Outcomes

Appendix C

Constructing Summary Value Functions for Course Placement Outcomes

Part 4 of each questionnaire asks respondents to make 10 choices. Each choice involves either taking the remedial course before taking the standard course, and earning grade G_1 in the standard course, or else taking the standard course directly, and earning grade G_2, where $G_1 > G_2$.

The result of the choices is a sequence of Rs and Ss, where:

R= Prefer to take the remedial course before taking the standard course.
S= Prefer to take the standard course directly.

There are $2^{10} = 1024$ possible sequences of response patterns, but most of them are "incoherent," because they are inconsistent with the transitivity property of preference relations. A coherent sequence is one that satisfies the following inequalities:

a. (R,A) $\succ$ (R,B) $\succ$ (R,C) $\succ$ (R,D) $\succ$ (R,F), and
b. (S,A) $\succ$ (S,B) $\succ$ (S,C) $\succ$ (S,D) $\succ$ (S,F),

where $\succ$ is a respondent's preference. Then Inequality a. implies, for example, that if (S,C) $\succ$ (R,A), then (S,C) $\succ$ (R,B), because (R,A) $\succ$ (R,B). Moreover, Inequality b. implies that if (S,C) $\succ$ (R,A), then (S,B) $\succ$ (R,A), because (S,B) $\succ$ (S,C). To simplify matters, I have also assumed that the following preferences exist:

c. (R,A) $\succ$ (S,F)
(R,B) $\succ$ (S,F)
(R,C) $\succ$ (S,F)
(R,D) $\succ$ (S,F)

The inequalities in c. imply that in Choices 4, 7, 9, and 10, the respondent must always choose taking the remedial course and earning a passing grade in preference to taking the standard course directly and receiving an F. These preferences may not actually be true of students who are very willing to take risks. Making these assumptions, however, considerably reduces the number of allowable sequences. Finally, I assume that:

d. (S,F) $\succ$ (R,F)

Inequalities a. - d. imply that every other course placement result is preferable to (R,F) (i.e., taking the remedial course and then receiving an F in the standard course). Table 1 on p. 3 shows the 14 choice sequences that satisfy these inequalities. I computed for each respondent an indicator function CHCSEQ:

CHCSEQ=1, if the respondent's sequence of choices was one of those listed in Table 1; and CHCSEQ=0, otherwise. The respondents for whom a coherent summary value function could be imputed were those for whom both CHRIND1=1 (where CHRIND1 is the coherence indicator for the grade value function) and CHCSEQ=1. These people were identified by the summary value function coherence indicator CHRIND2=CHRIND1*CHCSEQ.

Table 1.
Coherent Choice Sequences

Choice sequence	Choice number									
	1 (R,A) or (S,B)	2 (R,A) or (S,C)	3 (R,A) or (S,D)	4 (R,A) or (S,F)	5 (R,B) or (S,C)	6 (R,B) or (S,D)	7 (R,B) or (S,F)	8 (R,C) or (S,D)	9 (R,C) or (S,F)	10 (R,D) or (S,F)
1	R,A	R,A	R,A	R,A	R,B	R,B	R,B	R,C	R,C	R,D
2	R,A	R,A	R,A	R,A	R,B	R,B	R,B	S,D	R,C	R,D
3	R,A	R,A	R,A	R,A	S,C	R,B	R,B	R,C	R,C	R,D
4	R,A	R,A	R,A	R,A	S,C	R,B	R,B	S,D	R,C	R,D
5	R,A	R,A	R,A	R,A	S,C	S,D	R,B	S,D	R,C	R,D
6	S,B	R,A	R,A	R,A	R,B	R,B	R,B	R,C	R,C	R,D
7	S,B	R,A	R,A	R,A	R,B	R,B	R,B	S,D	R,C	R,D
8	S,B	R,A	R,A	R,A	S,C	R,B	R,B	R,C	R,C	R,D
9	S,B	R,A	R,A	R,A	S,C	R,B	R,B	S,D	R,C	R,D
10	S,B	R,A	R,A	R,A	S,C	S,D	R,B	S,D	R,C	R,D
11	S,B	S,C	R,A	R,A	S,C	R,B	R,B	R,C	R,C	R,D
12	S,B	S,C	R,A	R,A	S,C	R,B	R,B	S,D	R,C	R,D
13	S,B	S,C	R,A	R,A	S,C	S,D	R,B	S,D	R,C	R,D
14	S,B	S,C	S,D	R,A	S,C	S,D	R,B	S,D	R,C	R,D

Note: The shaded cells correspond to choosing to take the standard course directly.

If we use the customary grades A-F to measure achievement in the standard course, and if we neglect Withdrawal (W) grades, then there are 10 possible final outcomes of the placement system: X = {(R,A), . . ., (R,F), (S,A), . . ., (S,F)}, where R denotes taking the remedial course before taking the standard course, and S denotes taking the standard course directly. The set X, together with the set of possible placement test scores is the outcome space Θ.

In principle, one could elicit a value function for X with a diagram like that in Part 2 of the questionnaires. With such a diagram, however, the respondent would have to mark 8 outcomes (rather than the 3 outcomes A, B, and C) above the 0-100 scale. I believe that most respondents would have great difficulty doing this. Therefore, I elected to *impute* a summary value function *svf* for X, using the grade value function *gvf* elicited in Part 2 of the questionnaires as a reference. Now, there are many ways one could impute a summary value function; I chose the simplest method I could think of. Specifically, the imputed value function *svf* has the following properties:

a. *svf*(S,G) = *gvf*(G), for G = A, B, C, D
b. *svf*(R,F) = 0.
c. *svf*(S,F) = .25**gvf*(D)
d. For G = A, B, C, D, the values of *svf*(R,G) are interpolated between appropriate values of 1=*gvf*(A), *gvf*(B), *gvf*(C), *gvf*(D), and 0.

Equation a. says that the summary value function associated with taking the standard course directly and earning a particular grade G is equal to the grade value function *gvf* elicited in Part 2 of the questionnaires. Equation b. says that the worst possible result is to take the remedial course, then receive an F in the standard course. Equation c. says that taking the standard course directly, and receiving an F is slightly better than receiving an F in the standard course after taking the remedial course; I have arbitrarily assigned the value .25**gvf*(D) to this result. Property d. says that the outcomes associated with first taking the remedial course are to be assigned values according to the respondent's 10 choices in Part 4 of the questionnaires. Provided that the respondent's sequence of choices is one of the 14 coherent sequences listed in Table 1, it is possible to interpolate between values of *gvf* in a consistent way. Each of the 14 coherent choice sequences defines a separate imputed summary value function *svf*. The resulting values of the imputed summary value functions *svf* are shown in Table 2 on the following page.

Table 2.
Imputed Summary Value Functions for Course Placement Outcomes

Choice sequence	First course (Grade in standard course)									
	Std. (A)	Rem. (A)	Std. (B)	Rem. (B)	Std. (C)	Rem. (C)	Std. (D)	Rem. (D)	Std. (F)	Rem. (F)
1	1	.50 + .50*gvf(B)	gvf(B)	.50*gvf(B) + .50*gvf(C)	gvf(C)	.50*gvf(C) + .50*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
2	1	.50 + .50*gvf(B)	gvf(B)	.50*gvf(B) + .50*gvf(C)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
3	1	.50 + .50*gvf(B)	gvf(B)	.75*gvf(C) + .25*gvf(D)	gvf(C)	.50*gvf(C) + .50*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
4	1	.50 + .50*gvf(B)	gvf(B)	.75*gvf(C) + .25*gvf(D)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
5	1	.50 + .50*gvf(B)	gvf(B)	.90*gvf(D)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
6	1	.75*gvf(B) + .25*gvf(C)	gvf(B)	.50*gvf(B) + .50*gvf(C)	gvf(C)	.50*gvf(C) + .50*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
7	1	.75*gvf(B) + .25*gvf(C)	gvf(B)	.50*gvf(B) + .50*gvf(C)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
8	1	.75*gvf(B) + .25*gvf(C)	gvf(B)	.75*gvf(C) + .25*gvf(D)	gvf(C)	.50*gvf(C) + .50*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
9	1	.75*gvf(B) + .25*gvf(C)	gvf(B)	.75*gvf(C) + .25*gvf(D)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
10	1	.75*gvf(B) + .25*gvf(C)	gvf(B)	.90*gvf(D)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
11	1	.75*gvf(C) + .25*gvf(D)	gvf(B)	.50*gvf(C) + .50*gvf(D)	gvf(C)	.25*gvf(C) + .75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
12	1	.75*gvf(C) + .25*gvf(D)	gvf(B)	.50*gvf(C) + .50*gvf(D)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
13	1	.75*gvf(C) + .25*gvf(D)	gvf(B)	.90*gvf(D)	gvf(C)	.75*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0
14	1	.90*gvf(D)	gvf(B)	.75*gvf(D)	gvf(C)	.60*gvf(D)	gvf(D)	.50*gvf(D)	.25*gvf(D)	0

Note: Choice Sequence No. 1 corresponds to always choosing to take the remedial course before taking the standard course. The shaded cells for Choice Sequences Nos. 2 - 14 show the modifications in the summary value function that are associated with choosing to take the standard course directly.