DOCUMENT RESUME

ED 373 067 TM 021 788

AUTHOR Sawyer, Richard
TITLE Eliciting Utility Functions for Validating Course Placement Systems.
PUB DATE Apr 94
NOTE 77p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 5-7, 1994).
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; *Courses; Educational Assessment; Higher Education; Models; Pilot Projects; Prediction; Student Attitudes; *Student Placement; Teacher Attitudes; Teaching Methods; *Validity
IDENTIFIERS Decision Theory; *Utility Functions

ABSTRACT

Course placement systems in postsecondary education consist of an assessment component (to predict 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 course). The effectiveness of a placement system depends on students' ultimately succeeding in the standard courses. Success is usually defined in terms of course grades. In this paper a model, based on statistical decision theory, is proposed to help judge the effectiveness of the assessment and instructional components of course placement systems. Methods for eliciting students' and faculty members' preferences for different outcomes of course placement systems are also described. Finally, results of a pilot study of the feasibility of the elicitation methods involving 14 students and 12 faculty members are summarized. Respondents had few difficulties describing their value functions for earning grades and taking remedial courses, and they were able to state their equivalence probabilities. Six tables are included. Appendix A contains the study questionnaires, and Appendix B gives details concerning eliciting the utilities for course grades. (Contains 20 references.) (Author/SLD)

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, New Orleans

April 7, 1994

Abstract

Course placement systems in postsecondary education consist of an assessment component (to predict 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.

In this paper a model, based on statistical decision theory, is proposed to help judge the effectiveness of the assessment and instructional components of course placement systems. Methods for eliciting students' and faculty members' preferences for different outcomes of course placement systems are also described. Finally, results of a pilot study of the feasibility of the elicitation methods are summarized.

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.

During the past two decades, there has been a significant increase in the number of students whose academic preparation is not adequate for them to do work at a level traditionally expected of first-year college students. Surveys of remedial instruction typically indicate that about 90% of all postsecondary institutions have some form of placement and remedial instruction (Woods, 1985; Wright and Cahalan, 1985; McNabb, 1990).

Sawyer (1993) 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 also described a second model that could be used to investigate the effectiveness of remedial instruction. In all decision theory models, a numerical specification of the preferences of the decision maker (i.e., student, counselor, instructor, administrator) is essential. The principal goals of this study were to expand the remedial effectiveness model and to investigate alternative methods for eliciting the decision makers' preferences.

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

[1] I want to thank Jill Crouse, Mark Houston, Alan Nicewander, and Julie Noble for their help in collecting data and for their comments on earlier drafts of this paper. I also want to thank Mark Houston for his expert and rapid calculations of the utility functions.

for the standard course. At some institutions, the lower-level course may be given other names, such as "college-preparatory," "compensatory," "developmental," or "review"; at other 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 research, 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 have at least all of the following characteristics:

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 prediction is not an end in itself, but merely 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 even if they are provided remedial instruction, if they eventually drop out or fail in the standard course, no useful purpose will have been served 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 have selected their preferred educational programs in the first place, through appropriate counseling. This scenario illustrates that effective counseling is essential to an effective placement system. 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 all 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. Lively (1993), for example, reported on efforts in different states to eliminate remedial instruction from four-year public institutions by designating that role to two-year colleges.

Decision Theory Models

Statistical decision theory has been proposed by several writers, such as Cronbach and Gleser (1965) and Petersen and Novick (1976), as a useful means for analyzing educational selection problems. 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 previously collected. Lindley (1972) showed that an 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.

To illustrate, let us first consider the requirement that a placement test accurately identify underprepared students. Suppose that there is a cutoff score K on a placement test, and that:

* test scores are obtained for all freshmen at an institution;
* students whose test scores are less than K are classified as needing remedial instruction; and students whose test scores are greater than or equal to K are classified as not needing remedial instruction;
* the hypothetical performance of students in the standard course, without any prior remedial instruction, can be determined for all students.

Each student is classified either as being adequately prepared for the standard course (if his or her test score equals or exceeds K), or as needing remedial instruction (if the score is less than K). Because the classification for any student depends on the assumed cutoff score 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 classifications.

In the identification component of a placement system, the "outcomes" (Θ) for a group of students are their test scores and their performance (without prior remedial instruction) in the standard course. Therefore, for each student, four possible events could occur, as shown in Table 1 below:

Table 1
Events Associated With Identifying Underprepared Students

Event	Test score	Classification of student	Performance in standard course
A	≥K	Adequately prepared	Successful
B	≥K	Adequately prepared	Not successful
C	<K	Needs remedial instruction	Not successful
D	<K	Needs remedial instruction	Successful

In identifying high risk students, events A and C are preferred, because they correspond to correct classifications.

At an institution without a placement system, the events in Table 1 could be observed by requiring all students, regardless of their test scores, to enroll in the standard course, and then noting 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 A, B, C, and D. Let us suppose, temporarily, that this assumption is true; the modifications required when there is prior selection resulting from an existing placement system are described on p. 7.

Let $p_A(K)$, $p_B(K)$, etc., denote the observed proportions of students corresponding to events A, B, etc., in the entire group of students when the cutoff score is K. Then $p_A(K) + p_C(K)$ is the proportion of students who are correctly classified, and $p_B(K) + p_D(K)$ is the proportion of students who are incorrectly classified. The overall usefulness of the predictions could then be evaluated in terms of the benefits of correct classifications and the losses resulting from incorrect classifications. A function that assigns a value to outcomes such as these is called a utility function. For this simple model, in which outcomes for an entire group of students are considered, one possible utility function would be the relative frequency of correct classifications (accuracy rate); according to such a utility, every correct classification results in the same positive benefit, and every incorrect classification results in zero benefit.

A more complex utility function would assign different values to each event, and weight their sum:

$$u_1(K;\theta) = w_A p_A(K) + w_B p_B(K) + w_C p_C(K) + w_D p_D(K) \qquad (1)$$

where $0 \le w_A,\ldots,w_D \le 1$. Such a function would quantify the different benefits of the two types of correct classifications, as well as the different costs of the two types of incorrect classifications.

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. The accuracy rate, obtained by letting $w_A=w_C=1$ and $w_B=w_D=0$ in Expression (1), would seem to be a reasonable statement of the preferences of institutions in identifying underprepared students. Students themselves, on the other hand, may have different preferences (Whitney, 1989). For example, selection for and failure in the standard course (event B) might be more favorable to a student than enrollment in the remedial course when, in fact, the student could have been successful in the standard course. A utility function with weights $w_A=1$, $w_C=2/3$, $w_B=1/3$, and $w_D=0$ would correspond to students' preferences to minimize the time and cost required to pass their courses.

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 performance in the standard course) are not yet known. In (1), for example, the actual proportions $p_A(K)$, $p_B(K)$, 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,

$$u_1'(K) = E_\theta[\, u_1(K,\theta)\,] = w_A\hat{p}_A(K) + w_B\hat{p}_B(K) + w_C\hat{p}_C(K) + w_D\hat{p}_D(K) \qquad (2)$$

where $\hat{p}_A(K) = E_\theta[\, p_A(K)\,]$, $\hat{p}_B(K) = E_\theta[\, p_B(K)\,]$, etc., are estimated from a past group of students. In the Bayesian model, the estimates $\hat{p}_A(K)$, $\hat{p}_B(K)$, 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 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).

Sawyer (1993) described a simple procedure, based on sampling theory, for estimating the cell probabilities $\hat{p}_A(K)$, $\hat{p}_B(K)$, 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 Expression (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.

At an institution with an operational placement system, it is not possible to observe directly the events C and D in Table 1. The reason is that students whose test scores are below the cutoff score K 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 used to estimate the probabilities associated with these events, even though the events themselves are not directly observable. The estimation essentially involves extrapolating the logistic regression function, estimated from the data of students enrolled in the standard course, to test scores below the cutoff score K. 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 (defined below) was truncated at the 25th, 50th, and 75th percentiles, respectively, as compared to the standard error associated with no truncation.

Once estimates *a* and *b* have been obtained for the unknown parameters α and β, the conditional probability of success $\hat{P}(x)$ can be estimated by substituting *a* and *b* in (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 A in Table 1 can be estimated by:

$$\hat{p}_A(K) = \sum_{x \geq K} \hat{P}(x) * n(x) \; / \; N \qquad (4)$$

where $\hat{P}(x)$ = estimated $P\,[Y = 1 \mid X = x]$,

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

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

N = Σ n(x), the total number of students.

The proportions for B, C, and D 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). 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).[2]

[2] In 1994, ACT will begin operating a Course Placement Service (CPS). The CPS will enable postsecondary institutions to obtain accuracy rates and other placement statistics for first-year courses and placement groups they want to study.

It the expected utility function u_1' (K) has a definite peak at some score K_0 between the minimum and maximum possible scores, then K_0 is called an "optimal" cutoff score. For example, if $w_A=w_B=1/2$, then the optimal cutoff score maximizes the proportion of correct placement decisions in the placement group. (The implications of u_1' being monotonic are discussed below.) Of course, other aspects of the course placement system (such as the proportion of students placed in the remedial course) would need to be considered in actually selecting a cutoff score.

Definitions of success

In the model described in Table 1, there are only two outcomes 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 in principle describe people's preferences more accurately. 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 preliminary evidence about the appropriateness of threshold utility functions in course placement.

In some data sets, the expected utility function does not have a definite peak, but instead is monotonic. For example, when success is defined as earning a C or higher grade in the standard course, the estimated accuracy rate usually is very high (e.g, greater than or equal to .85) for all test scores, but frequently is a strictly decreasing function of test score. In this case the "optimal" cutoff score is the lowest possible test score, which is to say that every student should be placed in the standard course regardless of test score.

What a result like this really means is that even among students with low test scores, a large majority earn a C or higher. Now, this phenomenon could occur for several reasons. One possible reason is that the test scores are invalid for placement in the course (whether because of content mismatch, administration error, or whatever). A second possible reason is that there is an existing placement system, and it is succeeding wonderfully, because nearly every student placed in the standard course is successful. A third possibility is that the C-or-higher success criterion is not a valid measure of achievement in the course, because of instructors' reluctance to assign Ds and Fs. Shea (1994) suggested that "grade inflation" may be widespread, particularly in humanities courses. One effect of grade inflation is "grade compression," in which high-achieving students earn about the same grades as everyone else (Hancher, 1994).

I have analyzed dozens of data sets in which the accuracy rate curve for the C-or-higher success criterion does not have a definite peak. In every such data set, the accuracy rate curve for the B-or-higher success criterion *does* have a definite peak. Therefore, for these data sets at least, invalidity of the test scores is an unlikely explanation.

Further evidence could be obtained by eliciting the utilities of instructors for different grades. If $u(x)$ denotes an instructor's utility for grade x; if "<<" denotes "much less than"; and if the following result were observed:

$$u(F) < u(D) \ll u(C) < u(B) < u(A), \tag{5}$$

then the grade compression explanation for the placement validity results for the C-or-higher success criterion would be supported.

Modeling the Effectiveness of Remedial Instruction

A realistic decision model must consider not only the accuracy of identifying underprepared students, but also the effectiveness of remedial instruction. Let $P_R(x)$ denote the conditional probability of success in the standard course for someone whose placement test score is x, and who completes the remedial course before enrolling in the standard course. Let $P_S(x)$ be the corresponding conditional probability of success for someone who enrolls directly in the standard course. If the remedial instruction provided to underprepared students is effective, then $P_R(x)$ should exceed $P_S(x)$ for all test scores x below the cutoff.

As was noted earlier, there are statistical difficulties in estimating conditional probabilities of success. In a statistically ideal situation, students would be randomly assigned either to the remedial course, or to the standard course. Estimated conditional probabilities of success $\hat{P}_R(x)$ and $\hat{P}_S(x)$ could then be computed separately for each group, and compared. In most real applications, of course, students can not be randomly assigned to treatment groups. The most difficult situation occurs when students are assigned to remedial instruction solely on the basis of whether their test scores equal or exceed a cutoff score K. In this case, the conditional probability of success function can still be estimated separately for each group, but in order to make comparisons, $\hat{P}_S(x)$ must be extrapolated from the range of data collected ($x \geq K$), to the range of interest ($x < K$). In principle, such an extrapolated estimate is unbiased, given the assumptions of the model, but its sampling error can be significantly inflated by the truncation. Moreover, significant extrapolation makes the assumptions of the model less tenable, thus raising the possibility of bias. An example of the apparent biasing effects of extreme truncation is given by Schiel and Noble (1993).

A realistic decision model should also weigh the benefits of remedial instruction against the cost of providing it. The 2-by-2 table previously considered for identifying underprepared students can be adapted to address the cost/benefit issue. Suppose that a cutoff score K is assumed and that:

* test scores are obtained for all freshmen;
* students whose test scores are less than K are classified as needing remedial instruction, and are placed in the remedial course; and students whose test scores are greater than or equal to K are classified as being adequately prepared and are placed in the standard course;
* students' actual final performance in the standard course is known for all students (for those who were provided remedial instruction, as well as for those who were 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 2 below.

Table 2
Events Associated with the Identification and Remedial Instruction of Underprepared Students

Event	Test score	Course into which student is placed	Final performance in standard course
A	≥K	Standard	Successful
B	≥K	Standard	Unsuccessful
C	<K	Remedial	Unsuccessful
D	<K	Remedial	Successful

As before, let $p_A(K)$, $p_B(K)$, etc., denote the observed proportions of students corresponding to the events A, B, etc., in the entire group of students when the cutoff score is K. Define the function u_2 as follows:

$$u_2(K;\theta) = w_A\, p_A(K) + w_B\, p_B(K) + w_C\, p_C(K) + w_D\, p_D(K) \tag{6}$$

The weights w_A, w_B, w_C, and w_D reflect the relative benefits and costs associated with Outcomes A, B, C, and D. Although the function u_2 is mathematically the same as the function u_1 defined in (1), the outcomes and their associate weights have quite different meanings. Consider, for example, the trade-offs a student must make in their utilities. 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 $w_A, \ldots, w_D$ must balance the benefit in performance in the standard course against the extra time and money spent on taking the remedial course.

Just as in the classification problem, a richer model could be defined by considering a more complex criterion measure. 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, W, I). In this case, there would be 14 events (rather than 4) in the model.

Let $u_2'(K) = E_\theta[u_2(K;\theta)]$ be the expected utility function for the model in Table 2. If u_2' 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 placement variable, then the effectiveness of the placement system as a whole is supported. On the other hand, if u_2' attains a maximum value at the maximum possible score, the effectiveness of the placement variable is called into question -- the placement variable is not able to separate students who should enroll directly in the standard course from those who should first take the remedial course. Finally, if u_2' attains a maximum value at the minimum possible score, then the effectiveness of both the remedial instruction and the placement variable are called into question. Of course, all of these inferences depend on the validity of the success criterion variable (see previous discussion on pp. 9-10).

Assessing 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 in Table 2, 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 assessing (or, as some authors prefer to say, "eliciting") utility functions. One important concept distinguishing various utility theories is whether they are deterministic or stochastic:

* A *Bernoullian utility* measures the satisfaction of any sort of "want" without regard to uncertainty. For example, in economics, utility is the satisfaction that an individual receives from consuming

commodities (Yates, 1990). The key characteristic of Bernoullian utility is that it assigns numerical values to the subjective worth of outcomes without regard to uncertainty.

A simple example of eliciting a Bernoullian utility 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 utility values to outcomes is done outside any context of uncertainty or risk.

* A *von Neumann-Morgenstern* utility, by contrast, is explicitly defined in terms of probability. The standard assumption in von Neumann-Morgenstern 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 von Neumann-Morgenstern utility functions are defined in terms of probability; therefore, their elicitation is naturally done in reference to hypothesized probability distributions. Farquhar (1984) did a comprehensive review of different strategies for eliciting von Neumann-Morgenstern 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 von Neumann-Morgenstern 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. Bernoullian utility functions are sometimes called "value functions" to distinguish them from von Neumann-Morgenstern utilities

(Yates, 1990). I shall use the terminology "value function" and "von Neumann-Morgenstern utility function" when a distinction between the two is required; otherwise, I shall use the terminology "utility function" when referring generically to both.

One class of methods for eliciting von Neumann-Morgenstern 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.

Pilot Study

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 utility functions of students, instructors, and administrators is certainly crucial: If these decision makers are unable to provide accurate and inexpensive information that reflects their preferences, the method will be only a toy of statisticians, rather than a practical means for improving postsecondary education.

To obtain information on this issue, I conducted a pilot study. The purpose of the 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. Are the utility functions for different grades described by Inequality (5)?

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, large-scale service. 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 not be economically or practically feasible, however, to develop an interactive computer system for a course placement validity service and administer it to hundreds of institutions every year. If it were possible to elicit utilities reasonably accurately through a simple questionnaire, however, then a course placement validity service might feasibly incorporate utilities.

Question 2 also has implications for eliciting utilities in a large-scale course placement validity service. Value functions are much easier to elicit than von Neumann-Morgenstern utility functions. If elicited value functions of most decision makers closely approximate their elicited von Neumann-Morgenstern utility functions, then one need elicit only the value functions. If this did not work, one could search for a simple design for eliciting von Neumann-Morgenstern utility functions that would be satisfactory.

The motivation for investigating Question 3 was discussed on pp. 9-10.

Because students and instructors obviously have different interests, it is necessary to quantify their interests separately (Question 4). Two other groups, counselors and administrators, are also important decision makers in course placement systems, and likely have utilities that differ in important ways from those of both students and instructors. Unfortunately, there was not enough time in this study to develop and administer a questionnaire to counselors and administrators. Future studies will include them.

Definitive answers to these questions undoubtedly depend on many educational context 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) undoubtedly differ from those of students and instructors at community colleges. This pilot 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

Questionnaires were administered to students, English instructors, and mathematics instructors at a 2-year college in the midwest. To make the questionnaire items more meaningful to the respondents, I developed separate questionnaires for each of these three groups. The questionnaires are reproduced in Appendix A. (The college has been given the fictitious name "Midwest Community College").

Background information. Section 1 of each questionnaire asked about 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 utility information and whether their utility functions differ substantially from each other.

Value function. Next, a value function was elicited for the grades of B, C, and D in the standard course (see the "preliminary question" in the instruments). 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. The scale was intended to measure "satisfaction" with particular grades, with F indicating a satisfaction of 0, and A indicating a satisfaction of 100. Respondents were asked to mark the letters D, C, and B over the points on the scale that reflected their satisfaction with these grades.[3] This method of eliciting value functions is called "Stevens' magnitude estimation with modulus" (Falmagne, 1985).

Choice tasks. Section 2 of the student questionnaire elicited von Neumann-Morgenstern utility functions for the grades B, C, and D. (Recall, $u(A)=1$ and $u(F)=0$.) Students were presented a table asking

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

them to choose between different combinations of a sure grade G or a lottery in which they would receive the grade G_0 with probability p, or grade G_1 and prol bility $1-p$, where $G_0 < G < G_1$. All possible logical combinations (n=10) of G, G_0 and G_1 were presented, and were organized into blocks of choice tasks "G or $[G_0, p, G_1]$" where p took on a range of values. The values of p in the choice tasks were based on the responses of two university professors[4] to a prototype instrument.

Section 2 of the faculty questionnaires also elicited von Neumann-Morgenstern utilities, using the same table. The only difference between Section 2 of the faculty questionnaires and Section 2 of the student questionnaire is that the lotteries in the faculty questionnaire were stated in terms of grade distributions, rather than probabilities. For example, instead of being asked to indicate a preference between a student's earning a C or sure, and a lottery in which a student has a 95% probability of earning an A, and a 5% probability of earning an F (Style 1), faculty members were asked to indicate a preference between a uniform result, in which all students earned a C, and a mixed result, in which 95% of the students earned an A, and 5% earned an F (Style 2). Of course, the phenomenon described by Style 2 is the expected result of the phenomenon described by Style 1, but the items in the two styles are not, strictly speaking, asking the same question. I elected to use Style 2, because it is more natural from the perspective of instructors.[5]

Direct elicitation. The table format in Section 2 was intended to reduce the difficulty to students and faculty in selecting the equivalence probability p in the comparison $\theta_k \sim [\ \theta_i, p, \theta_j\ ]$ by presenting a series of simple choices. A respondent's value of p can be inferred from where in the table the respondent "breaks away" from choosing the lottery to choosing the sure grade. In principle, one could determine p simply by directly asking for it. In practice, decision makers find it very difficult to pick a value of p

[4] The two professors were not affiliated with the institution at which I administered the questionnaires.

[5] The previously mentioned university professors reacted to Style 1 with incomprehension, skepticism, and rejection. On the other hand, both professors understood and accepted (with some reluctance and pontification) Style 2. Both, however, strongly preferred the value function to either of the vN-M utility functions.

"out of the blue" (Farquhar, 1984). To determine whether elicitations "out of the blue" are at all feasible in course placement, respondents were asked to declare a value of p in 4 of the 10 comparisons. These questions were posed in Section 3. As in Section 2, the items in the student questionnaire were stated in terms of probability[6], and the items in the faculty questionnaires were stated in terms of grade distributions.

Remedial instruction. Section 4 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: Θ = { take remedial course & earn A in standard course, do not take remedial course & earn A in standard course, . . ., take remedial course & earn F in standard course, do not take remedial course & earn F in standard course }. Eliciting von Neumann-Morgenstern utilities for all these outcomes seemed, on its face, to be infeasible in a paper-and-pencil format. 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 the previous sections of the instrument.

The questionnaires were administered to a sample of 14 students and 12 faculty members (4 from the mathematics department and 8 from the English department) at a midwest community college. The questionnaires were administered on a "drop in" basis between 10:00 AM and 12:30 PM on March 8 and March 10, 1994. To increase the likelihood of obtaining usable data and to observe first-hand any difficulties the respondents were having, the questionnaires were administered by three ACT staff members. Different rooms were used to administer questions to students and faculty members. Respondents were allowed to take as much time as they wanted to complete the questionnaires.

Analysis

[6] Actually, the word "chances" was used, so as to seem less formal.

The following analyses were done for the student data and for the pooled faculty data:

Validity indicators. For Section 2, a validity indicator for each of the 10 comparisons (blocks of choice tasks) was computed. The values of the indicators are:

1 = Provided one or more blank response, or one or more response that was inconsistent with the instructions.

2 = Provided responses in accordance with instructions, but they were logically inconsistent.

3 = Provided logically consistent responses, but they were all in the same column.

4 = Provided logically consistent responses, and they were in both columns.

An example of a logically inconsistent result (indicator value 2) is a respondent's choosing the sure grade of B in the choice task "B or [A, .95, F]", then choosing the lottery in the choice task "B or [A, .90, F]". An indicator value of 3 would result, for example, from always choosing the sure grade of B in the choice tasks "B or [A, *p*, F]", where p=.95, .90, .85, .80, and .75. From such responses, it would be possible to tell only that $p > .95$. Therefore, the information from a block of choice tasks can be used to estimate a utility function only if the indicator value is 4.

A summary validity indicator for the responses to Section 2 was also created. The summary indicator is equal to the minimum of the validity indicators observed for the 10 comparisons. Thus a summary indicator value of 3 means that the individual's responses to all the comparisons in Section 2 were internally consistent, but for at least one comparison, there was no "break point" in the values of p.

Summary validity indicators were also assigned to the responses in Sections 1, 3, 4, and the preliminary question. For Sections 1, 4, and the preliminary question, the possible values were:

1 = Provided one or more blank response, or one or more response that was inconsistent with the instructions.

2 = Provided responses in accordance with instructions, but they were logically inconsistent.

3 = Provided logically consistent responses.

For Section 3, the possible values of the validity indicator were:

1 = Provided one or more blank response, or one or more response that was inconsistent with the instructions.

2 = Provided responses in accordance with instructions.

The validity indicators for Sections 2 and 3 did not address the logical consistency *across* comparisons. (This issue was addressed by the "validity rates" described below.) Instead, all elicited equivalence probabilities obtained from responses that were internally consistent within choice tasks were used to solve systems of linear equations.

Frequency distributions were computed for all the validity indicators. The small sample sizes obviously precluded breaking down the frequency distributions according to background and educational characteristics. I plan to do such analyses in future studies with larger samples.

Calculation of utilities. The data from each comparison in Section 2 can be represented by a linear equation. For example, the data from the comparison C ~ [B, p, D] (see Choice Tasks 31-35 in Section 2) can be represented by the linear equation:

$$u(C) = p * u(B) + (1 - p) * u(D). \quad (7)$$

The value of p in this comparison was taken to be the mid-point of the probabilities (for the higher grade) where the "break" occurred. For example, if a respondent chose the lottery in Choice Tasks 31-34 (see questionnaires in Appendix A) and the sure grade of C in Choice Task 35, p was set equal to .725.

Because there were 10 comparisons in Section 2, and because there are three "unknowns" (u(B), u(C), and u(D)), the responses to Section 2 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 have given 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. Finally, the mean of the calculated utility values that were greater than 0 and less than 1 was calculated for each respondent.

The responses to Questions 1 - 4 in Section 3 resulted in a maximum of $6 = \binom{4}{2}$ sets of estimates for u(B) and u(C). The mean of these estimates whose values were greater than 0 and less than 1 were calculated. Questions 1 - 4 did not provide estimates for u(D).

Summary and comparisons of utilities for grades. The calculated value functions, median elicited utilities from the table in Section 2, and mean elicited utilities from the questions in Section 3 were summarized over the respondents in each group (students or faculty). The minima, medians, and maxima of these distributions were calculated.

The median values of the statistics associated with the three elicitation methods were compared to each other and to Inequality (5) on p. 10. Finally, the median values for students and faculty were compared.

Calculation and summary of utilities for course placement outcomes. For each respondent, the value function elicited from Section 4 was referenced to the median utility functions elicited from the preliminary question and from Section 2. The distributions of the resulting two functions were then summarized over the respondents in each group (students or faculty).

Results

Validity indicators. Tables 3 and 4 show the distribution of the previously defined validity indicators for faculty and students, respectively. As expected, neither group had much difficulty with the educational and background questions in Section 1. Encouragingly, 11 of the 12 faculty members and all 14 students also provided consistent value function data.

Most respondents also provided consistent data in the choice tasks in Section 2. For example, 8 of the 12 faculty members had consistent responses in all 10 comparisons (summary validity indicator values of 3 or 4). Unfortunately, however, about half of the consistent responses did not result in a "break point" in the equivalence probability p; hence, they could not be used to estimate a utility function. This result

Table 3.
Distribution of Validity Indicators for Responses to Questionnaires
(Faculty; N=12)

Indicator value	Section 1 (bkgd. char.)	Prel. quest. (val. fn.)	Section 2 (von Neumann-Morgenstern utilities)											Section 3 (vN-M ut.)	Section 4 (rem. instr.)
			Cmp. 1	Cmp. 2	Cmp. 3	Cmp. 4	Cmp. 5	Cmp. 6	Cmp. 7	Cmp. 8	Cmp. 9	Cmp. 10	Summary		
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
2	1	1	3	1	3	2	2	1	3	1	0	1	4	12	0
3	10	11	3	5	5	5	5	2	3	10	9	6	8	–	11
4	--	--	6	6	4	5	5	8	6	1	3	5	0	–	–

Table 4.
Distribution of Validity Indicators for Responses to Questionnaires
(Students; N=14)

Indicator value	Section 1 (bkgd. char.)	Prel. quest. (val. fn.)	Section 2 (von Neumann-Morgenstern utilities)											Section 3 (vN-M ut.)	Section 4 (rem. instr.)
			Cmp. 1	Cmp. 2	Cmp. 3	Cmp. 4	Cmp. 5	Cmp. 6	Cmp. 7	Cmp. 8	Cmp. 9	Cmp. 10	Summary		
1	0	3	0	1	1	1	2	2	2	2	3	2	3	3	2
2	0	0	0	0	0	0	2	1	0	0	1	1	3	11	12
3	14	11	8	5	7	2	4	4	1	4	5	6	7	–	--
4	--	--	5	8	6	10	6	7	11	8	5	5	1	--	--

suggests that there were wide variations in individuals' responses to choice tasks, and that the ranges of equivalence probabilities that I provided in Section 2 were insufficient.In principle, one could develop a table with fine increments of choices of p that span the entire interval (0,1). In practice, the resulting questionnaire would be so long and tedious that respondents might refuse to complete it. A more feasible strategy would be to revise the ranges in p presented in the

choice tasks so that they covered a larger proportion of the population, by shifting the ranges of the scale values. In Comparison 8, for example, respondents had to choose between a sure grade of D and lotteries that offered either an A or an F, where the probability of A ranged from .80 to .60. Ten of the 12 faculty members (and 4 of the 14 students) always chose the lottery in preference to the sure grade. It is possible that if the probabilities were shifted downward, say from .60 to .40, more of the respondents would have had a "break point"

Alternatively, one could simply increase the increments in p among the choice tasks. For example, instead of decreasing in increments of .05, p could be made to decrease in increments of .10.[7] There is a price to be paid, however, for doing this: Larger increments in p will result in larger errors in measuring the "true" equivalence probabilities, which will, in turn, result in larger errors in the elicited utility functions.

With these two improvements, it may well be feasible to elicit utilities for course grades in a paper-and-pencil questionnaire that could be administered in a large-scale course placement validity service. The evidence from this study suggests, however, that further pilot testing will be needed.

The four questions in Section 3 asked respondents to state their equivalence probabilities directly. Surprisingly, large majorities of both faculty and students responded to these questions according to instructions.

[7] In hindsight, increments of .10 would have been a better choice. In discussing the prototype questionnaire with the two university professors, however, both had break points that were consistently very high.

Section 4 elicited a value function for taking a remedial course. As one would expect in eliciting a value function, nearly all the respondents in each group provided consistent responses to these items.

Utility function validity rates. Even though a respondent may have provided valid data for each choice task in Section 2 of the questionnaire, an elicited utility value obtained by solving a system of linear equations may still be inappropriate. Depending on the equivalence probabilities obtained from the questionnaire, a calculated utility value may be negative or greater than 1. Inappropriate values of calculated utilities result from inconsistencies *among* comparisons. (Recall, the validity indicator data shown in Tables 3 and 4 pertain to inconsistencies among choice tasks *within* comparisons.)

I determined, for each respondent, the proportion of the 108 possible utility values calculated from Section 2 data that were greater than 0 and less than 1. I also determined the proportion of the 6 possible utility values calculated from Section 3 data that were greater than 0 and less than 1. Table 5 shows the distributions of these "validity rates" (minimum, median, and maximum) separately for students and faculty. From data provided by the typical respondent, only 5% (or fewer) of the potential 108 linear equations yielded usable results. Even in the best situations, there were sufficient data to solve only 21%-27% of the equations. Clearly, there is room for improvement in applying the choice task methodology.

Interestingly, the validity rates for utilities based on directly elicited equivalence probabilities (Section 3) were much higher. Of course, because there were only 6 possible systems of linear equations using the Section 3 data, exact comparisons can not be made. Nevertheless, this result suggests that contrary to my expectations, directly eliciting utilities may be feasible.

There were no practically significant differences between the validity rates of the utilities of faculty and those of students.

Utility function values. For each respondent and for each course grade, I calculated the mean of the elicited von Neumann-Morgenstern utility function values that were between 0 and 1. Table 6 summarizes the distributions (minimum, median, maximum) of these mean values across respondents, by group and method. Table 6 also summarizes the distributions of each group's value functions.

Table 5.
Distribution of Validity Rates in Eliciting von Neumann-Morgenstern Utility Functions for Course Grades, by Group and Method

Group	Grade	Choice tasks (Sec. 2)			Direct elicitation (Sec. 3)		
		Min.	Med.	Max.	Min.	Med.	Max.
Faculty (N=12)	B	.00	.04	.24	.83	1.00	1.00
	C	.00	.03	.27	.83	1.00	1.00
	D	.00	.00	.21	---	---	---
Students (N=14)	B	.00	.05	.26	.83	1.00	1.00
	C	.00	.05	.26	.83	1.00	1.00
	D	.00	.04	.21	---	---	---

Table 6.
Distribution of Elicited Utility Functions for Course Grades, by Group and Method

Group	Grade	Value function			vN-M utility function (Choice tasks; Sec. 2)			vN-M utility function (Direct elicitation; Sec. 3)		
		Min.	Med.	Max.	Min.	Med.	Max.	Min.	Med.	Max.
Faculty (N=12)	B	.35	.80	.95	.75	.88	.95	.31	.80	.96
	C	.25	.60	.75	.58	.70	.81	.15	.65	.86
	D	.10	.25	.60	.18	.37	.63	---	---	---
Students (N=14)	B	.70	.80	.90	.81	.92	.95	.29	.98	.97
	C	.40	.60	.70	.72	.77	.89	.46	.66	.94
	D	.10	.40	.60	.43	.54	.75	---	---	---

The most striking result is the large spread in each distribution, for every group, method, and grade. For example, consider the distribution of faculty members' value functions for the grade B: One faculty member assigned u(B)=.35, while another assigned u(B)=.95! Does this result mean that utility values are wildly person-specific, or does it mean that some people had only a foggy idea of what they were doing when completed the questionnaire? In view of the distributions of validity rates in Table 5, I am more inclined to favor the latter explanation.

Another interesting result is that despite the large spread in each distribution, the medians for the different grades are consistent: For each group and method, med. u(D) << med. u(C) < med. u(B). For example, the median of faculty members' value functions for B was .80; the median for C was .60; and the median for D was .25. Note that this result supports the hypothesis in Equation (5).

In comparing results among methods and between groups, I used the median values. Let us consider .10 to be a practically significant difference in comparing results among methods and between groups. According to this standard, von Neumann-Morgenstern utilities, whether elicited directly or from choice tasks, are larger than value functions. For example, consider the median values for u(B) of students: the value function result is .80, the choice task result is .92, and the direct elicitation result is .98. In no case was a median value function larger than the corresponding median von Neumann-Morgenstern utility. If replicated in larger samples, this result would suggest that faculty and students are risk-averse (Yates, 1990, p. 270).

The median utilities for the choice tasks and the direct elicitation method were much more similar. The greatest difference occurred in eliciting u(C) for students; the median for the choice tasks was .77, and the median for direct elicitation was .66.

The median utilities of faculty and students are similar. Only for the grade of D were there differences between faculty and student utilities that exceeded .10. For example, faculty members' median value functions for B, C, and D were .80, .60, and .25, respectively. Students median value functions for these grades were .80, .60, and .40, respectively.

Remedial course taking. Large majorities of both faculty and students expressed their preference for taking a remedial course when doing so results in a higher grade in the standard course. This preference was expressed in all the choice tasks, even those in which the student would end up with a C or D after taking remedial instruction. The choice tasks for which the greatest number of respondents preferred taking the standard course directly (without prior remedial instruction) were:

* Choice Task 5 (B with, or C without prior remedial course work): 3 of 14 students, and 1 of 12 faculty.
* Choice Task 8 (C with, or D without prior remedial course work): 2 of 14 students, and 1 of 12 faculty.
* Choice Task 10 (D with, or F without prior remedial course work): 0 of 14 students, and 2 of 12 faculty.

For most people in this sample, therefore, the final grade in the standard course outweighs other considerations, such as time and cost, in considering whether to take a remedial course. If this result generalizes to larger samples, then evaluating the effectiveness of remedial instruction could be done reasonably well simply by comparing $P_R(x)$ and $P_S(x)$ for test scores x below the cutoff score (see discussion on pp. 9-10).

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 utility function of the outcomes, and by averaging the utility function with respect to a probability distribution, one can evaluate the effectiveness of the course placement system and select optimal cutoff scores. Utility functions may be categorized according to whether they are deterministic (value functions) or stochastic (von Neumann-Morgenstern utility functions).

In a pilot study at a midwestern community college, students and faculty had few difficulties in describing on a paper-and-pencil questionnaire their value functions for earning course grades and taking remedial courses. Most of them were also able to respond appropriately to choice tasks for eliciting von

Neumann-Morgenstern utilities. Furthermore (and contrary to my expectations), they also were able to state directly their equivalence probabilities. The wide range of utility values for all three methods, however, makes one question whether some of the respondents understood what they were doing.

The median von Neumann-Morgenstern utilities tended to be larger than the corresponding value functions. The median results for faculty and students were very similar.

Future research

I will revise the pilot questionnaire and test it again in summer 1994 on small samples of students and faculty. Then, in fall 1994 I will use a final version to elicit utility functions of much larger samples of students and instructors at a two-year community college and at a state university. I will also develop and administer a questionnaire for counselors and administrators. Finally, I will collect data on the placement variables and course grades of the students. By studying these data in the context of the elicited utility functions, I will obtain evidence about the perceived overall effectiveness of the placement systems at these institutions.

References

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

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 grade 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.

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.

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

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

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. L. (1989). *Validating the use of ACT Assessment scores and high school grades for remedial course placement in college.* (ACT Research Report Series No. 89-4). Iowa City, Iowa: The American College Testing Program.

Sawyer, R. (1993, April). *Decision theory models for validating course placement systems.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta.

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.

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.

Appendix A

Questionnaires Administered to Students, English Faculty, and Mathematics Faculty

Students' Preferences for Course Placement Decisions

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

Purpose and Background of Project

There are benefits, risks, and costs associated with the decision to take any college course. Part of ACT's work involves developing placement tests that can be used to help students decide which courses to take. I want to learn about your preferences in making course placement decisions.

I will ask you some questions about your academic work at Midwest, and about your preferences for different course placement decisions and grades. This questionnaire is not a test---there is no score. 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 sections. I will give you some directions and examples before you begin each section. If you finish a section early, please wait for instructions before going on to the next section. The entire questionnaire should take about 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.

Section 1: Background information

*** NOTE: All responses are confidential ***

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 taken	Grade you received	Check here if you are currently enrolled	Grade you expect to receive
Elements of Writing				
Composition I				
College Writing				
Beginning Algebra				
Intermediate Algebra				
Finite Mathematics				
Pre Calculus				
Mathematics for Decision 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)

Please continue on to the next page.

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.

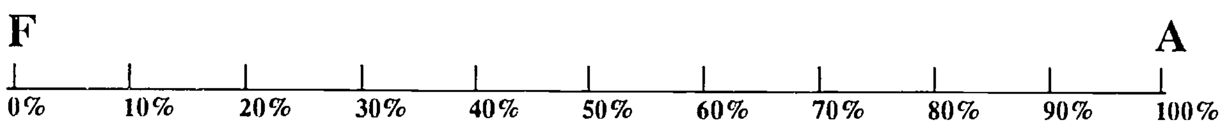
***** Please stop temporarily and wait for further instructions. *****

Students' Preferences for Course Placement Decisions

*** Preliminary Question ***

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

The line below is meant to suggest your relative preferences for the different letter grades. To simplify the discussion, I have associated the letter grade of F with 0% satisfaction, and the letter grade of A with 100% satisfaction:



Please indicate on this line your relative preferences for 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 answers should reflect your <u>satisfaction</u> with particular grades. They do not necessarily have to correspond to a percent-correct grading scale (for example, A = 90% or more correct).

Section 2: Preferences for Grades

I want to find out your preferences for grades in the typical courses you take. In answering the questions, please think of an English or mathematics course that is most typical of those you are taking or have taken.

The table below presents different imaginary situations, in which you are asked to choose between two different options:
(Col. 1) Earning a particular grade for sure in a course, or
(Col. 2) Entering a lottery in which the grade you earn depends on chance.
Although these situations are imaginary, I would like to know your opinions about them. For each situation, please check (✓) either the box in Col. (1) or the box in Col. (2), according to your preference.

	Which would you prefer ?	
	Col. (1)	Col. (2)
Situation	Earn this grade for sure ...	Enter this lottery ...
1	B	95% chance of A, 5% chance of F
2	B	90% chance of A, 10% chance of F
3	B	85% chance of A, 15% chance of F
4	B	80% chance of A, 20% chance of F
5	B	75% chance of A, 25% chance of F

*** Please continue on the next page. ***

Situation	Which would you prefer ?	
	Col. (1) Earn this grade for sure ...	Col. (2) Enter this lottery ...
6	B	90% chance of A, 10% chance of D
7	B	85% chance of A, 15% chance of D
8	B	80% chance of A, 20% chance of D
9	B	75% chance of A, 25% chance of D
10	B	70% chance of A, 30% chance of D
11	B	85% chance of A, 15% chance of C
12	B	80% chance of A, 20% chance of C
13	B	75% chance of A, 25% chance of C
14	B	70% chance of A, 30% chance of C
15	B	65% chance of A, 35% chance of C
16	C	90% chance of A, 10% chance of F
17	C	85% chance of A, 15% chance of F
18	C	80% chance of A, 20% chance of F
19	C	75% chance of A, 25% chance of F
20	C	70% chance of A, 30% chance of F

*** Please continue on to the next page. ***

Situation	Which would you prefer ? Col. (1) Earn this grade for sure ...	Col. (2) Enter this lottery ...
21	C	85% chance of A, 15% chance of D
22	C	80% chance of A, 20% chance of D
23	C	75% chance of A, 25% chance of D
24	C	70% chance of A, 30% chance of D
25	C	65% chance of A, 35% chance of D
26	C	90% chance of B, 10% chance of F
27	C	85% chance of B, 15% chance of F
28	C	80% chance of B, 20% chance of F
29	C	75% chance of B, 25% chance of F
30	C	70% chance of B, 30% chance of F
31	C	90% earn a B, 10% earn a D
32	C	85% earn a B, 15% earn a D
33	C	80% earn a B, 20% earn a D
34	C	75% earn a B, 25% earn a D
35	C	70% earn a B, 30% earn a D

*** Please continue on to the next page. ***

Situation	Which would you prefer ?	
	Col. (1) Earn this grade for sure ...	Col. (2) Enter this lottery ...
36	D	80% chance of A, 20% chance of F
37	D	75% chance of A, 25% chance of F
38	D	70% chance of A, 30% chance of F
39	D	65% chance of A, 35% chance of F
40	D	60% chance of A, 40% chance of F
41	D	75% chance of B, 25% chance of F
42	D	70% chance of B, 30% chance of F
43	D	65% chance of B, 35% chance of F
44	D	60% chance of B, 40% chance of F
45	D	55% chance of B, 45% chance of F
46	D	75% chance of C, 25% chance of F
47	D	70% chance of C, 30% chance of F
48	D	65% chance of C, 35% chance of F
49	D	60% chance of C, 40% chance of F
50	D	55% chance of C, 45% chance of F

*** Please stop temporarily and wait for further instructions. ***

Section 3: Preferences for Grades

The following questions also ask about your preferences for grades in a standard course. (A "standard course" is a for-credit course that you need to pass to satisfy the requirements of your program at Midwest.)

In answering these questions, please do not look back at your answers in Section 2.

1. Suppose I offered you the choice of:
   (i.) Earning exactly a grade of **B** for sure, or
   (ii.) Entering a lottery, where you would either earn an **A** or else you would earn a **C**.

   How large would the chances of earning an A have to be before you would prefer the lottery [Choice *(ii.)*] to the sure grade of B [Choice *(i.)*]?

   I would want the chances of earning an A to be before I chose the lottery.

2. Suppose I offered you the choice of:
   (i.) Earning exactly a grade of **B** for sure, or
   (ii.) Entering a lottery, where you would either earn an **A** or else you would earn an **F**.

   How large would the chances of earning an A have to be before you would prefer the lottery [Choice *(ii.)*] to the sure grade of B [Choice *(i.)*]?

   I would want the chances of earning an A to be before I chose the lottery.

3. Suppose I offered you the choice of:
   (i.) Earning exactly a grade of **C** for sure, or
   (ii.) Entering a lottery, where you would either earn an **A** or else you would earn an **F**.

   How large would the chances of earning an A have to be before you would prefer the lottery [Choice *(ii.)*] to the sure grade of C [Choice *(i.)*]?

   I would want the chances of earning an A to be before I chose the lottery.

4. Suppose I offered you the choice of:
   (i.) Earning exactly a grade of **C** for sure, or
   (ii.) Entering a lottery, where you would either earn a **B** or else you would earn an **F**.

   How large would the chances of earning a B have to be before you would prefer the lottery [Choice *(ii.)*] to the sure grade of C [Choice *(i.)*]?

   I would want the chances of earning a B to be before I chose the lottery.

***** Please stop temporarily and wait for further instructions. *****

Section 4:
Preferences for Course Placement Decisions

Background information.
Let a "standard course" be a for-credit course that is taken by many entering students, and that is required for a 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.*

Course placement systems usually involve placement testing. From your score on a placement test, it may be possible to estimate your chances of succeeding in the standard course. If your chances of success in the standard course are low, then you might be advised to take the developmental course.

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.

***** Please continue to the next page. *****

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

***** Please stop and wait for an announcement. *****

English Faculty Members' Preferences for Course Placement Decisions

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

Purpose of this Project

There are benefits, risks, and costs associated with the decision to take any college course. Part of ACT's work involves developing placement tests that can be used to help students decide which courses to take. I want to investigate English faculty members' preferences for the results of course placement decisions.

I will ask you some questions about your teaching responsibilities at Midwest and about your preferences for different levels of student performance. This questionnaire is not a test---there is no score. 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 sections. I will give you some directions and examples before you begin each section. If you finish a section early, please wait for instructions before going on to the next section. The entire questionnaire should take about 30 minutes to complete.

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.

Section 1: Background information

*** NOTE: All responses are confidential ***

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.

Course	Check here if you are currently teaching	Check here if you have taught	Percentage of students who earned . . .				
			A	B	C	D	F
Elements of Writing							
College Writing							
Composition I							
Composition II							

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

_______________ (month and year)

3. Do you teach part-time or full-time at Midwest? *(Check one only.)*

_______ part-time

_______ full-time

*** Please continue on to the next page. ***

4. 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):				

5. 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 percentage of As, a certain percentage of Bs, etc.

____ I grade mostly according to a fixed standard, but I may modify 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.

***** Please stop and wait for further instructions. *****

English Faculty Members' Preferences for Course Placement Decisions

*** Preliminary Question ***

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

F **A**

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

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

Please note that your answers should reflect your <u>satisfaction</u> with particular levels of students' performance. They do not necessarily have to correspond to a percent-correct grading scale (for example, A = 90% or more correct).

Section 2: Preferences for Grades

Let a "standard course" be a for-credit course that is taken by well-prepared entering students, and that is required for some program. For example, well-prepared entering students may take *Composition I*, and may need to pass it to satisfy the requirements of some programs at Midwest.

I want to find out your relative preferences for the grades that students earn in the standard courses you teach. In responding to the questions, please think of the standard course that is most typical of those you teach. If you do not teach a standard course, think of the typical course you do teach.

The table below presents different hypothetical situations in which you are asked to choose between two different results for a class that you teach:

(Col 1.) A uniform result, in which all students in the class earn the same particular grade, or
(Col 2.) A mixed result, in which some students earn one particular grade, and everyone else earns another grade.

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) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
1	B	95% earn an A, 5% earn an F
2	B	90% earn an A, 10% earn an F
3	B	85% earn an A, 15% earn an F
4	B	80% earn an A, 20% earn an F
5	B	75% earn an A, 25% earn an F

*** Please continue on to the next page. ***

Situation	Which would you prefer ?	
	Col. (1) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
6	B	90% earn an A, 10% earn a D
7	B	85% earn an A, 15% earn a D
8	B	80% earn an A, 20% earn a D
9	B	75% earn an A, 25% earn a D
10	B	70% earn an A, 30% earn a D
11	B	85% earn an A, 15% earn a C
12	B	80% earn an A, 20% earn a C
13	B	75% earn an A, 25% earn a C
14	B	70% earn an A, 30% earn a C
15	B	65% earn an A, 35% earn a C
16	C	90% earn an A, 10% earn an F
17	C	85% earn an A, 15% earn an F
18	C	80% earn an A, 20% earn an F
19	C	75% earn an A, 25% earn an F
20	C	70% earn an A, 30% earn an F

*** Please continue on to the next page. ***

Situation	Which would you prefer ? Col. (1) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
21	C	85% earn an A, 15% earn an D
22	C	80% earn an A, 20% earn an D
23	C	75% earn an A, 25% earn an D
24	C	70% earn an A, 30% earn an D
25	C	65% earn an A, 35% earn an D
26	C	90% earn an B, 10% earn an F
27	C	85% earn an B, 15% earn an F
28	C	80% earn an B, 20% earn an F
29	C	75% earn an B, 25% earn an F
30	C	70% earn an B, 30% earn an F
31	C	90% earn a B, 10% earn a D
32	C	85% earn a B, 15% earn a D
33	C	80% earn a B, 20% earn a D
34	C	75% earn a B, 25% earn a D
35	C	70% earn a B, 30% earn a D

*** Please continue on to the next page. ***

Situation	Which would you prefer ? Col. (1) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
36	D	80% earn an A, 20% earn an F
37	D	75% earn an A, 25% earn an F
38	D	70% earn an A, 30% earn an F
39	D	65% earn an A, 35% earn an F
40	D	60% earn an A, 40% earn an F
41	D	75% earn a B, 25% earn an F
42	D	70% earn a B, 30% earn an F
43	D	65% earn a B, 35% earn an F
44	D	60% earn a B, 40% earn an F
45	D	55% earn a B, 45% earn an F
46	D	75% earn a C, 25% earn an F
47	D	70% earn a C, 30% earn an F
48	D	65% earn a C, 35% earn an F
49	D	60% earn a C, 40% earn an F
50	D	55% earn a C, 45% earn an F

*** Please stop temporarily and wait for further instructions. ***

Section 3: Preferences for Grades

The following questions also ask about your preferences for students' performance in a standard course that you teach. In answering these questions, please do not review your responses in Section 2.

1. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of B, or
   (ii.) A mixed result, in which **P** percent of the students earn an A, and all the rest earn a C.

   How large would the percentage **P** of students who earn an A have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of B [Choice *(i.)*]?

   I would want P =   before selecting Choice *(ii.)* .

2. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of B, or
   (ii.) A mixed result, in which **P** percent of the students earn an A, and all the rest earn an F.

   How large would the percentage **P** of students who earn an A have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of B [Choice *(i.)*]?

   I would want P =   before selecting Choice *(ii.)* .

3. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of C or
   (ii.) A mixed result, in which **P** percent of the students earn an A, and all the rest earn an F.

   How large would the percentage **P** of students who earn an A have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of C [Choice *(i.)*]?

   I would want P =   before selecting Choice *(ii.)* .

4. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of C, or
   (ii.) A mixed result, in which **P** percent of the students earn a B, and all the rest earn an F.

   How large would the percentage **P** of students who earn an B have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of C [Choice *(i.)*]?

   I would want P =   before selecting Choice *(ii.)* .

*** Please stop temporarily and wait for further instructions. ***

Section 4:
Preferences for Course Placement Decisions

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 program. For example, well-prepared entering students may take *Composition I* and may need to pass it to satisfy the requirements of some 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", in order to acquire the skills needed to succeed in the standard course. At Midwest, for example, *Elements of Writing* would be considered a developmental course for *Composition I.*

Course placement systems usually involve placement testing. From the score on a placement test, it may be possible to estimate a student's chances of succeeding in the standard course. If the chances of success in the standard course are low, then the student is advised to take the developmental course.

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, and it costs additional money. 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.

*** **Please continue to the next page.** ***

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 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 do you prefer ?	
	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

*** Please stop and wait for an announcement. ***

Mathematics Faculty Members' Preferences for Course Placement Decisions

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

Purpose of this Project

There are benefits, risks, and costs associated with the decision to take any college course. Part of ACT's work involves developing placement tests that can be used to help students decide which courses to take. I want to investigate faculty members' and students' preferences for the results of course placement decisions.

I will ask you some questions about your teaching responsibilities at Midwest and about your preferences for different levels of student performance. This questionnaire is not a test---there is no score. 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 sections. I will give you some directions and examples before you begin each section. If you finish a section early, please wait for instructions before going on to the next section. The entire questionnaire should take about 30 minutes to complete.

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.

Section 1: Background information

*** NOTE: All responses are confidential ***

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.

Course	Check here if you are currently teaching	Check here if you have taught	Percentage of students who earned . . .				
			A	B	C	D	F
Beginning Algebra							
Intermediate Algebra							
Finite Mathematics							
Pre Calculus							
Mathematics for Decision Making							
Statistical Ideas							
Other mathematics courses (please specify):							

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

_______________ (month and year)

3. Do you teach part-time or full-time at Midwest? *(Check one only.)*

_______ part-time

_______ full-time

***** Please continue on to the next page. *****

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

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

5. 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 percentage of As, a certain percentage of Bs, etc.

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

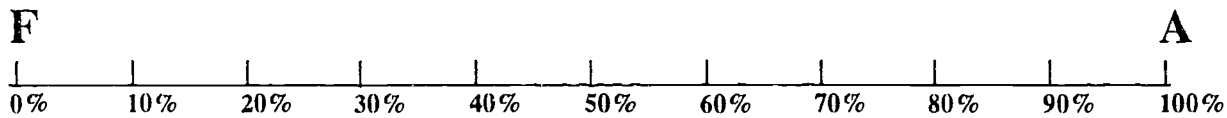
____ I grade mostly "on a curve", but I may modify some grades if students' performance merits doing so.

***** Please stop and wait for further instructions. *****

Mathematics Faculty Members' Preferences for Course Placement Decisions

*** Preliminary Question ***

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



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

Please note that your answers should reflect your <u>satisfaction</u> with particular levels of students' performance. They do not necessarily have to correspond to a percent-correct grading scale (for example, A = 90% or more correct).

Section 2: Preferences for Grades

Let a "standard course" be a for-credit course that is taken by well-prepared entering students, and that is required for some program. For example, well-prepared entering students may take *Mathematics for Decision Making*, and may need to pass it to satisfy the requirements of some programs at Midwest.

I want to find out your relative preferences for the grades that students earn in the standard courses you teach. In responding to the questions, please think of the standard course that is most typical of those you teach. If you do not teach a standard course, think of the typical course you do teach.

The table below presents different hypothetical situations in which you are asked to choose between two different results for a class that you teach:

(Col 1.) A uniform result, in which all students in the class earn the same particular grade, or

(Col 2.) A mixed result, in which some of the students earn one particular grade, and everyone else earns another grade.

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

	Which would you prefer ?	
	Col. (1) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
Situation		
1	B	95% earn an A, 5% earn an F
2	B	90% earn an A, 10% earn an F
3	B	85% earn an A, 15% earn an F
4	B	80% earn an A, 20% earn an F
5	B	75% earn an A, 25% earn an F

*** Please continue on to the next page. ***

Situation	Which would you prefer ?	
	Col. (1) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
6	B	90% earn an A, 10% earn a D
7	B	85% earn an A, 15% earn a D
8	B	80% earn an A, 20% earn a D
9	B	75% earn an A, 25% earn a D
10	B	70% earn an A, 30% earn a D
11	B	85% earn an A, 15% earn a C
12	B	80% earn an A, 20% earn a C
13	B	75% earn an A, 25% earn a C
14	B	70% earn an A, 30% earn a C
15	B	65% earn an A, 35% earn a C
16	C	90% earn an A, 10% earn an F
17	C	85% earn an A, 15% earn an F
18	C	80% earn an A, 20% earn an F
19	C	75% earn an A, 25% earn an F
20	C	70% earn an A, 30% earn an F

*** Please continue on to the next page. ***

Situation	Which would you prefer ?	
	Col. (1) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
21	C	85% earn an A, 15% earn an D
22	C	80% earn an A, 20% earn an D
23	C	75% earn an A, 25% earn an D
24	C	70% earn an A, 30% earn an D
25	C	65% earn an A, 35% earn an D
26	C	90% earn an B, 10% earn an F
27	C	85% earn an B, 15% earn an F
28	C	80% earn an B, 20% earn an F
29	C	75% earn an B, 25% earn an F
30	C	70% earn an B, 30% earn an F
31	C	90% earn a B, 10% earn a D
32	C	85% earn a B, 15% earn a D
33	C	80% earn a B, 20% earn a D
34	C	75% earn a B, 25% earn a D
35	C	70% earn a B, 30% earn a D

*** Please continue on to the next page. ***

Situation	Which would you prefer ?	
	Col. (1) All students earn a grade of ...	Col. (2) There is a mixture of grades ...
36	D	80% earn an A, 20% earn an F
37	D	75% earn an A, 25% earn an F
38	D	70% earn an A, 30% earn an F
39	D	65% earn an A, 35% earn an F
40	D	60% earn an A, 40% earn an F
41	D	75% earn a B, 25% earn an F
42	D	70% earn a B, 30% earn an F
43	D	65% earn a B, 35% earn an F
44	D	60% earn a B, 40% earn an F
45	D	55% earn a B, 45% earn an F
46	D	75% earn a C, 25% earn an F
47	D	70% earn a C, 30% earn an F
48	D	65% earn a C, 35% earn an F
49	D	60% earn a C, 40% earn an F
50	D	55% earn a C, 45% earn an F

*** Please stop temporarily and wait for further instructions. ***

Section 3: Preferences for Grades

The following questions also ask about your preferences for students' performance in a standard course that you teach. In answering these questions, please do not review your responses in Section 2.

1. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of B, or
   (ii.) A mixed result, in which P percent of the students earn an A, and all the rest earn a C.

   How large would the percentage P of students who earn an A have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of B [Choice *(i.)*]?

   I would want P = before selecting Choice *(ii.)* .

2. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of B, or
   (ii.) A mixed result, in which P percent of the students earn an A, and all the rest earn an F.

   How large would the percentage P of students who earn an A have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of B [Choice *(i.)*]?

   I would want P = before selecting Choice *(ii.)* .

3. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of C, or
   (ii.) A mixed result, in which P percent of the students earn an A, and all the rest earn an F.

   How large would the percentage P of students who earn an A have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of C [Choice *(i.)*]?

   I would want P = before selecting Choice *(ii.)* .

4. Suppose I asked you to choose between the following two results in your course:
   (i.) A uniform result, in which all students earn exactly a grade of C, or
   (ii.) A mixed result, in which P percent of the students earn a B, and all the rest earn an F.

   How large would the percentage P of students who earn an B have to be before you would prefer the mixed result [Choice *(ii.)*] to the uniform grade of C [Choice *(i.)*]?

   I would want P = before selecting Choice *(ii.)* .

***** Please stop temporarily and wait for further instructions. *****

Section 4:
Preferences for Course Placement Decisions

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 program. For example, well-prepared entering students may take *Mathematics for Decision Making* and may need to pass it to satisfy the requirements of some 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", in order to acquire the skills needed to succeed in the standard course. At Midwest, for example, *Intermediate Algebra* would be considered a developmental course for *Mathematics for Decision Making.*

Course placement systems usually involve placement testing. From the score on a placement test, it may be possible to estimate a student's chances of succeeding in the standard course. If the chances of success in the standard course are low, then the student is advised to take the developmental course.

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, and it costs additional money. 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.

*** Please continue to the next page. ***

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 and carries no program credit.

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

Situation	Which do you prefer ?	
	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

*** Please stop and wait for an announcement. ***

Appendix B

Details of 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.