New rules, new tools
Niessen, Anna Susanna Maria
Chapter 7
The utility of selective admission in Dutch higher education

A version of this chapter was published (in Dutch) as:
Wanneer heeft selectie in het hoger onderwijs zin?
De stand van zaken anno 2015.
Tijdschrift voor Hoger Onderwijs, 33, 4-19.
Abstract
In the past few years, the number of programs with selective admission in Dutch higher education has increased. The aim of these selective admissions is to improve graduation rates, reduce time to graduation, and improve academic performance. However, accessibility should be guaranteed and rejection rates of students who would have been successful should be minimal. Most studies on selective admission procedures do not provide information about such effects, but only report correlations between scores obtained in admission procedures and academic performance. Several utility models are available to assess the added value of selective admission procedures based on those correlations, while taking relevant context factors such as the base rate and the selection ratio into account. In this paper we discuss the Taylor-Russell model and the Naylor-Shine model in the context of admission procedures to Dutch higher education. The utility models show that selective admission only has utility when the base rate and the selection ratio are low, even when the predictive validity of admission procedures is relatively high. We illustrate this with some empirical examples. Given the high selection ratios of most Dutch selective academic programs, we show that the recent increase in selective admission will have little added value in terms of increased academic performance in the student population.
7.1 Introduction
In the past few years, the number of selective programs in Dutch higher education has increased (Education Council, 2014). About one-third of all students apply to a selective program (Ministry of Education, Culture, and Science, 2014). In addition, the possibilities for colleges to select their students have also increased; college-controlled selection will have replaced the weighted lottery system completely by 2017. Replacing lottery by selective admission should yield a better match between the students and the study programs, leading to improved academic performance and reduced dropout (Korthals, 2007).

Given these developments, a lot of attention is being paid to the effects of selective admission and several research projects have been conducted to study those effects. Some researchers performed studies in which they compared the performance of students who were admitted through different procedures. For instance, they compared the performance of students who were admitted through lottery with students who were admitted through selective admission (Schripsema, van Trigt, Borleffs, & Cohen-Schotanus, 2014; Urlings-Strop, Stijnen, Themmen, & Splinter, 2009; Visser, van der Maas, Engels-Freeke, & Vorst, 2012; Wouters, 2017). However, the students in these studies were not randomly assigned to those groups, but, instead, students chose to participate in the selection procedure or to take their chances in the lottery. Whereas relevant variables like gender and high school grades were often controlled for in these studies, other relevant variables like motivation could have influenced the choice to participate in the selection- or the lottery procedure (Schripsema et al., 2014), which makes it difficult to draw conclusions about the effects of completely replacing lottery admission by selective admission.

Other studies investigated the prediction of academic performance by studying the relationships between scores on instruments used in admission procedures and academic performance through correlations and regression analyses (van der Heijden, Hessen, & Wubbels, 2012; Niessen, Meijer, & Tendeiro, 2016). All these studies answer important questions, but they do not show the added value of selective admission in terms of academic outcomes. Several utility models can be used to estimate that added value. Utility models for selection procedures were developed within the context of organizational psychology and personnel selection, but they can also be applied to other selection contexts. However, they are rarely applied to educational selection. Utility models use information about the predictive validity of selection procedures (the correlation between scores obtained in an admission procedure and academic performance), and information...
about the context in which the selection takes place. These contextual factors have a large influence on the utility of selection procedures.

In this chapter, we extensively discuss two utility models: the Taylor-Russell model (1939) and the Naylor-Shine model (1965). We also illustrate the use of these models with empirical examples obtained from a selective admission procedure for a psychology program. We discuss the predictive validity of existing selection procedures for higher education, relevant contextual factors in Dutch higher education, and the implications of these factors for the utility of selective admission.

### 7.1.1 Admission Decisions

There are several good predictors of academic performance. Examples are previous educational performance (van der Heijden et al., 2012), personality traits, motivation, behavioral tendencies (Richardson, Abrahams, & Bond, 2012), and curriculum-sampling tests (Niessen et al., 2016; Visser et al., 2012). However, the relationships between a predictor or a combination of predictors and criterion measures are far from perfect. In addition, student's traits and skills cannot be observed directly and are measured by instruments and procedures with imperfect reliability. As a result, selection decisions are always made with a certain amount of uncertainty (Mellenbergh, 1993). In this context, Mellenbergh (1993) described different components of selection decisions. The first component is the decision maker, a person or an organization that is responsible for the decisions. Second, there is an action, in this case, accepting or rejecting a student. That action is based on certain expected criterion behavior (in our context, expected academic performance). In turn, the expected criterion behavior is based on information (in our context, performance during an admission procedure). The information is used in a decision rule that dictates what action is required given the information. Finally, the action has a certain utility; the effectiveness of the action, given the criterion behavior.

### 7.2 Utility Models

As discussed above, an important component of the utility of selection decisions is the relationship (correlation) between “the information” and the criterion behavior. However, it is important to realize that when there is no selection, a substantial percentage of students will perform successfully. Also, given that selection procedures are imperfect, it is inevitable that some unsuccessful students will get selected, and that some successful students will be rejected. How many of such incorrect decisions are made depends on the predictive validity of the procedure and on contextual factors. These contextual factors are rarely
incorporated in selection research. As Drenth (1995; 1998) and Mellenbergh (1995) have previously argued, the utility of selection procedures should be judged on the basis of a comparison with the results obtained when selection procedures are not used.

There are several utility models to estimate the added value of selection procedures in terms of increased success; the Taylor-Russell model (1939) and the Naylor-Shine model (1965) will be discussed in detail below. The utility based on these models can be estimated using tables provided in the articles cited above, or using an online calculator (McLellan, 1996). Other utility models that we do not discuss in detail are the model by Lawshe, Bolda, Brune, and Auclair (1958), Brogden, (1946), and Cronbach and Gleser (1965). In the model by Lawshe et al. (1958), all applicants are divided into groups based on their admission test scores (for example: high, average, low). The percentage of successful applicants per group can be computed using Lawshe’s utility model. This is an extension of the Taylor-Russell model that allows the distinction of more than the two groups (accept and reject) that are defined by the Taylor-Russell model. The models by Brogden (1946) and Cronbach and Gleser (1965) estimate utility in terms of financial gains (also see Mellenbergh, 1995). However, estimating financial utility is difficult in practice, because estimates of the mean and the variance of the costs of the selection procedure are needed.

7.2 Utility Models
7.2.1 The Taylor-Russell Model
The Taylor-Russell model is the most often used and the simplest utility model. In the Taylor-Russell model, the criterion behavior is dichotomized into ‘successful’ and ‘unsuccessful’ behavior, based on a cutoff value of the criterion measure. Utility is defined as the increase in the success ratio (the proportion of successful students within all selected students), compared to having no selection procedure (the base rate), or compared to the success ratio based on an alternative selection procedure. Figure 7.1 shows this graphically. Section A contains all wrongfully rejected students (rejected, but would have been successful), section B contains all rightfully accepted students (accepted and successful), section C contains all rightfully rejected students (rejected and would not have been successful), and section D contains all students who were wrongfully accepted (accepted, but were not successful).

The increase of the success ratio depends on three factors. First, the predictive validity of the selection procedure, expressed as the correlation between the scores obtained in the selection procedure and the scores on criterion behavior. Second, the base rate, that is, the proportion of successful students without a
selection procedure, or the proportion of successful students in the applicant pool. Third, the selection ratio: the proportion of students selected from the applicant pool, or, the degree of selectivity. In Figure 7.1, the base rate is the ratio of sections A and B to sections A, B, C, and D. The selection ratio is the ratio of sections B and D to section A, B, C, and D. The predictive validity is shown by the oval, which incorporates all applicants and represents the relationship between the predictor and the criterion. The stronger this relationship, the fewer persons there are in sections A and D (incorrect decisions). The success ratio is the ratio between section B to section B and D. Based on information of the predictive validity of a selection procedure, the base rate, and the selection ratio, the corresponding success ratio can be determined. The utility of a selection procedure compared to having no selection procedure equals the success ratio minus the base rate. The utility of a selection procedure compared to another selection procedure equals the difference between the success ratios.

Figure 7.1. Schematic display of a selection procedure with cut-off scores for the predictor and the criterion. Adapted from Taylor & Russell (1939). Note. A: Wrongfully rejected, B: Rightfully accepted, C: Rightfully rejected, D: Wrongfully accepted.

Success ratio = \( \frac{B}{B+D} \)

Base rate = \( \frac{A+B}{A+B+C+D} \)

Selection ratio = \( \frac{B+D}{A+B+C+D} \)
The message of the Taylor and Russell’s paper (1939) was that the utility of a selection procedure does not only depend on the predictive validity of the procedure. When the base rate is high (many students are successful even without selection), even selection procedures with very high predictive validity have little utility. Similarly, when the selection ratio is high (almost all applicants get accepted), a selection procedure will also have little utility, regardless of the predictive validity. Conversely, a selection procedure with relatively low predictive validity can have considerable utility when the base rate and the selection ratio are low.

Assumptions under the Taylor-Russell model
The Taylor-Russell model is easy to use and can provide useful information about the utility of selection procedures. However, the model does rely on some assumptions that can be difficult to meet. A first assumption is that the base rate is known. Universities that have some tradition in any form of selective admission usually do not have that information because the success ratio under an existing selection procedure with non-zero validity does not equal the base rate. Using the success ratio under a current selection procedure with non-zero validity as an estimate for the base rate can lead to an overestimation of the utility of an alternative selection procedure (Smith, 1948). Smith (1948) also describes a procedure to estimate the base rate based using the validity coefficient and the selection ratio.

A second assumption is that the validity coefficient is based on the entire applicant pool. Validity studies are usually based on data obtained in operational selection procedures in which some applicants will be rejected, which leads to range restriction. Range restriction leads to an underestimation of the validity coefficient. Hence, using a validity coefficient that was affected by range restriction for utility estimation leads to an underestimation of the utility. There are, however, methods to correct validity coefficients for range restriction (e.g., Hunter & Schmidt, 2000).

The Taylor-Russell model also assumes a bivariate-normal, linear, and homoscedastic relationship between the predictor and the criterion measure. The dichotomization of the criterion measure into “successful” and “unsuccessful” should be based on an underlying normally distributed variable. If the criterion measure is truly dichotomous (for example, dropping out of a study program), the
Taylor-Russell model does not apply. Abrahams, Alfe, and Wolfe (1971) provided tables that are suitable for truly dichotomous criterion variables. While the Taylor-Russell model assumes that the dichotomization of successful and unsuccessful students is based on an underlying normally distributed variable, all successful students are considered equally successful based on this dichotomization. As a result, the Taylor-Russell model only answers the question of the percentage of successful students given a certain selection procedure, but it does not provide insight into how much more successful the students will be. Cascio (1980) pointed out that this is a major shortcoming of the Taylor-Russell model and states that this may lead to an underestimation of the effects that selection procedures can have on criterion behavior.

### 7.2.2 The Naylor-Shine Model

The Naylor-Shine model (Naylor & Shine, 1965) overcomes Cascio’s (1980) objections; it does not require a dichotomization of the criterion behavior and provides insight into the increase in the degree of success. In the Naylor-Shine model, utility is not defined in terms of the increase in the success ratio, but in terms of the increase in mean criterion performance (for example, final GPA), given the predictive validity of the selection procedure and the selection ratio. The Naylor-Shine model assumes that there is a linear relationship between predictive validity and utility. In addition, a low selection ratio increases the utility. Similar to the Taylor-Russell model, it is assumed that the relationship between the predictor and the criterion is bivariate-normal, linear, and homoscedastic, and that the validity coefficient is free of range restriction. The Naylor-Shine model can be used for two purposes. The first is to estimate the increase in mean criterion performance based on a certain selection procedure (for example, the increase in mean GPA). The second is to determine a cutoff value for the predictor to meet a desired level of mean criterion performance in the selected group (for example, to reach a desired mean GPA).

### 7.3 Empirical Examples

Two empirical examples of utility estimation will be provided below, based on data obtained from first-year psychology students at the University of Groningen. First, the utility of the selection procedure used for admission to the psychology program is demonstrated for predicting retention (not dropping out) after the first year. Because retention is a truly dichotomous variable, the adapted Taylor-Russell tables by Abrahams et al. (1971) were used. Second, the Naylor-Shine model is demonstrated to estimate the increase in the mean first year-GPA as a result of the selection procedure. Results are shown for two hypothetical selection ratios: A selection ratio of .30, representing strict selection (30% of all applicants are
admitted), and a selection ratio of .80, representing mild selection (80% of all applicants are admitted).

We used data obtained from 638 psychology students who participated in a selection procedure before starting the psychology program. For the selection procedure, they received two chapters of a book used in the *Introduction to Psychology* course that they had to study independently. At selection day, they took an exam about the material consisting of 40 multiple-choice items (for a complete description, see chapter 2 of this thesis). None of the students were rejected, and all applicants who withdrew from enrollment did so on a voluntary basis. The correlations between the exam score and the criterion variables were rounded for simplicity (utility tables only provide whole decimal values of the base rate and the selection ratio, and correlation coefficient with increments of .05). The rounded correlation between the score on the curriculum-sampling test and retention was $r = .30$ (a moderate relationship), and the rounded correlation between the score on the curriculum-sampling test and the mean grade in the first year was $r = .50$ (a strong relationship)$^{12}$. Since there was no actual selection, we used the outcomes of the students who started the program to estimate the base rate and the mean criterion performance without selection. Because some self-selection did occur, this may result in a slight overestimation. The proportion of students who were successful (did not drop out of the program) was .80, which serves an estimate of the base rate. The mean criterion performance for the mean grade in the first year was $M = 6.6, SD = 1.4$.

### 7.3.1 Utility for Predicting Dropout

When we use the adapted Taylor-Russell tables by Abrahams et al. (1971) to find the success ratio for a validity coefficient of $r = .30$, a base rate of .80, and a selection ratio of .80 (mild selectivity), we get a success ratio of .85. So, mild selectivity based on this admission procedure would reduce dropout by 5%, compared to having no selective admission. In addition, it can be derived that 68% of the applicants are rightfully accepted and 8% are rightfully rejected, but 12% are wrongfully rejected and 12% are wrongfully accepted. Without selective admission, 80% is rightfully accepted and 20% is wrongfully accepted, but, rightfully or wrongfully, no one is rejected. When we use a more strict selection ratio of .30, the success ratio equals .92. In this situation 28% of all applicants are rightfully accepted and 18% are rightfully rejected, but 52% are wrongfully rejected and 2% are wrongfully accepted. So, the success ratio increases with

---

12 Even though no actual selection took place, there was some range restriction due to self-selection. Therefore, we applied a correction for range restriction, but that did not change the validity coefficients to such an extent that the rounded coefficients changed. For details, see chapter 3.
stricter selectivity, but the proportion of students that are wrongfully rejected also increases.

### 7.3.2 Utility for Predicting the Mean First-year Grade
When we apply the Naylor-Shine model to this data to estimate the increase in mean criterion performance as a result of implementing the selection procedure with \( r = .5 \) and a mild selection ratio of .80, the mean grade in the first year would be \( M_{\text{new}} = 6.8 \). The first-year mean grade without selection was \( M_{\text{old}} = 6.6 \), so implementing the selection procedure with mild selectivity would yield a 0.2 increase in the criterion performance. If we would adopt a more strict selection ratio of .30, the mean criterion performance would equal \( M_{\text{new}} = 7.4 \), corresponding to a 0.8 increase in mean criterion performance. So, a very selective procedure leads to a considerable increase in performance, whereas a mildly selective procedure only yields a modest increase in performance.

### 7.4 Predictive Validity and Contextual Factors in Dutch Higher Education
The effect of implementing selective admission on the academic performance of students depends on the predictive validity of the instruments used in selective admission procedures, the base rate, and the selection ratios. Below, we discuss these three concepts on the basis of some realistic numbers in the context of Dutch higher Education.

#### 7.4.1 Predictive Validity
The utility tables show validity coefficients ranging from zero to one. However, not all values within that range are equally realistic. The predictive validity of methods and instruments used to predict performance in higher education will never be even close to one (Dawes, 1979). Depending on the criterion of interest (e.g., success defined after one year, after three years, success in jobs) the predictive validity of most methods will most likely be somewhere between \( r = .30 \) and \( r = .60 \) at best. Criterion performance that is further away in time is usually more difficult to predict. So, roughly half of the utility tables consist of unrealistic scenarios for educational selection.

#### 7.4.2 Base Rate
The base rate depends on the criterion and it can vary between study programs. The base rate for not dropping out of a program in the first year can be reasonably high, while the base rate for obtaining a bachelor’s degree within three years is usually a lot lower (32% on average; Statistics Netherlands, 2014). The utility of selective admission procedures is much higher for criterion measure with lower base rates.
7.4.3 Selection Ratio
Inspecting the utility tables shows us that even if the base rate and the predictive validity are high, the utility of selection procedures will be low, unless the selection ratio is low. This is illustrated by data collected in the United States by ACT (2014). This data shows that the more selective colleges have lower dropout rates and better performing students. To obtain substantial utility, a considerable amount of applicants needs to be rejected in selective admission procedures. Whereas the number of Dutch study programs with a numerus fixus (a fixed number of available places, which allows selectivity) has recently increased (Education Council, 2014), that does not necessarily mean that access to Dutch higher education has become more restricted.

To get an overview of the degree of selectivity of Dutch higher education programs, we computed the selection ratio for each program using publicly available data (Dienst Uitvoering Onderwijs, 2014) by dividing the maximum number of admissible students by the number of applicants. In research universities, there is a distinction between programs with a “program restriction” (nation-wide restriction to the program, for example: medicine), and a “university restriction” (a restriction for a program at a specific university, for example: Law at the University of Amsterdam). Two of the seven programs that had a program restriction had a selection ratio of 1: thus all students who applied were admitted. Three of the program-restricted programs had a selection ratio lower than 1, but higher than .80, and the three remaining programs had a selection ratio lower than .80. There were 25 programs with a university-restriction, of which 19 (76%) had a selection ratio of 1. Five of these programs (20%) had a selection ratio between 1 and .80, and only one program had a selection ratio below .80.

Universities of applied sciences only have programs with university restrictions. Of the 88 numerus fixus programs at universities of applied sciences, 62 (71%) had a selection ratio of 1. Eleven programs (13%) had a selection ratio between 1 and .80, and 15 programs (17%) had a selection ratio below .80. Only two research university programs (dentistry and veterinary medicine) and four applied sciences programs (dental care at the Utrecht University of Applied Sciences, forensic research at the Amsterdam University of Applied Sciences, and medical aid at the Utrecht University of applies sciences and the Rotterdam University of Applied Sciences) had a selection ratio below .50. So, very little actual selection took place at ‘selective’ programs. The majority of the programs admitted all applicants, and if

---

13 Applicants who withdrew or that did not meet the minimum admission requirements were excluded.
there was selective admission, only a small proportion of the applicants was usually rejected.

7.5 Discussion
To estimate the utility of selective admission procedures we should not only take the predictive validity into account, but also look at the base rate and the selection ratio. On the basis of the Taylor-Russell framework (1939) it can be argued that selection procedures with relatively low predictive validity can still be very useful if the base rate and the selection ratio are low. Therefore, some authors have argued that it is not justified to dismiss instruments and selection procedures due to suboptimal predictive validity. On the other hand, and that is our take home message: even instruments and procedures with relatively high predictive validity will have little to no utility when the base rate and/or the selection ratio are high, as is often the case in Dutch selective admission for higher education programs. A large amount of these programs do not seem to need a numerus fixus to keep the number of enrolling students under a given threshold, as our analysis showed. If universities want to use selective admission procedures to increase academic achievement and to reduce dropout, they will need to drastically decrease their selection ratios. In addition, they should take into account that selective admission is inevitably accompanied by wrongfully rejecting at least some applicants. Programs should consider if the often small increase in academic achievement outweighs those wrongful rejections.