Attitudes Toward Statistics and Their Relationship with Short- and Long-Term Exam Results

Stijn Vanhoof
Ana Elisa Castro Sotos
Patrick Onghena
Lieven Verschaffel
Wim Van Dooren
Wim Van den Noortgate
Katholieke Universiteit Leuven

Journal of Statistics Education Volume 14, Number 3 (2006),
ww2.amstat.org/publications/jse/v14n3/vanhoof.html

Copyright © 2006 by Stijn Vanhoof, Ana Elisa Castro Sotos, Patrick Onghena, Lieven Verschaffel, Wim Van Dooren, Wim Van den Noortgate all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Assessment; Attitudes Toward Statistics scale.

Abstract

This study uses the Attitudes Toward Statistics (ATS) scale (Wise 1985) to investigate the attitudes toward statistics and the relationship of those attitudes with short- and long-term statistics exam results for university students taking statistics courses in a five year Educational Sciences curriculum. Compared to the findings from previous studies, the results indicate that the sample of undergraduate students have relatively negative attitudes toward the use of statistics in their field of study but relatively positive attitudes toward the course of statistics in which they are enrolled. Similar to other studies, we find a relationship between the attitudes toward the course and the results on the first year statistics exam. Additionally, we investigate the relationship between the attitudes and the long-term exam results. A positive relationship is found between students’ attitudes toward the use of statistics in their field of study and the dissertation grade. This relationship does not differ systematically from the one between the first year statistics exam results and the dissertation grade in the fifth year. Thus, the affective and cognitive measures at the beginning of the curriculum are equally predictive for long-term exam results. Finally, this study reveals that the relationship between attitudes toward statistics and exam results is content-specific: We do not find a relationship between attitudes and general exam results, only between attitudes and results on statistics exams.
1. Introduction

The importance of students’ attitudes toward statistics when following an introductory statistics course is widely recognized. According to Gal, Ginsburg, and Schau (1997) such attitudes may affect the extent to which students will develop useful statistical thinking skills and apply what they have learned outside the classroom. Therefore, it is important to study thoroughly the attitudes students have toward statistics and the relationship of these attitudes with statistical performance. A first step in accomplishing this goal is to develop and evaluate instruments to assess students’ attitudes toward statistics; work that has already been initiated by a number of researchers (e.g., Roberts and Bilderback 1980; Schau, Stevens, Dauphinee, and Del Vecchio 1995; Shultz and Koshino 1998; Waters, Martelli, Zakrajsek, and Popovich 1988; Wise 1985).

A widely used instrument is the Attitudes Toward Statistics (ATS) questionnaire (Wise 1985). The ATS is a 29-item, Likert-type scale with five response possibilities ranging from ‘strongly disagree’ to ‘strongly agree’. The ATS questionnaire includes both positively and negatively formulated items. The questionnaire consists of two subscales – Field (20 items) and Course (9 items) – that respectively aim to measure attitudes toward the use of statistics in the students’ fields of study and attitudes toward the particular statistics course in which they are enrolled. Example items include:

Field

I feel that statistics will be useful to me in my profession.
Studying statistics is a waste of time.

Course

The thought of being enrolled in a statistics course makes me nervous.
I get upset at the thought of enrolling in another statistics course.

The ATS scales can be used to give a general overview of the attitudes toward statistics of a group of students. Most of the previous studies using ATS (e.g., Elmore and Lewis 1991, 1993; Waters et al. 1988; Wise 1985) include an evaluation of the internal consistency, a description of the attitudes students have toward statistics before and after taking the statistics course, and an analysis of how these attitudes are related to their first year statistics exam results (as an indication of their statistics performance). Most of these studies therefore involve two administrations, one before and one after the statistics course (but before the students know their exam results).

The present study aims at extending the existing evidence on the relationship between attitudes toward statistics and performance. This is done in three ways. First, the study provides new data and measures of reliability of the ATS by two administrations of the questionnaire in an introductory statistics course for Flemish undergraduate students in Educational Sciences. Second, while previous investigations are limited to the relationship between attitudes and first year exam results, this study examines the relationship between the attitudes students have and their exam results not only at the beginning of the curriculum, but also in later years. Third, while the previous research only addresses the relationship between students’ attitudes and their grades in a statistics course, the present study also investigates the relationship with their general exam results (short- and long-term). In this way, we examine whether the findings and trends of other studies are also valid for Flemish undergraduate students and, more importantly, we want to check whether the documented relationship between attitudes and statistics exam results also holds in the long term and for general exam results.
We are aware that some authors caution against the indiscriminate use of paper-and-pencil Likert-type scales, like the ATS, to study attitudes (Gal and Ginsburg 1994; Schau et al. 1995). For instance, it is difficult to imagine that students’ attitudes toward statistics could be captured by two global ATS scores (Gal and Ginsburg 1994). Furthermore, we have to take into account that there may be cultural differences in responding to such questionnaires, even at the level of subtle nuances in the translation and interpretation of the items. Therefore, we acknowledge that our study will only be one step toward a deeper understanding of the complex relationship between attitudes and performance.

This paper is organized in five sections. Since we will compare our data with the findings reported in previous studies, in Section 2 we start with a summary of these earlier findings. Section 3 describes the methodology of this study. In Section 4, the results are presented, followed, in Section 5, by a discussion.

2. Empirical background

Most of the previous studies use results from other investigations as a benchmark. Therefore, we will also compare the data of the current study with data from previous studies (Aldogan and Aseeri 2003; D’Andrea and Waters 2002; Elmore and Lewis 1991; Elmore, Lewis, and Bay 1993; Mvududu 2003; Rhoads and Hubele 2000; Roberts and Reese 1987; Shultz and Koshino 1998; Waters et al. 1988; Wise 1985). We first present a detailed overview of the results of these previous studies and emphasize the most important findings and trends that can be formulated based on these results. This overview will provide the reader with the necessary background to situate and interpret our new empirical data presented in Section 4.

The Appendix provides an overview of these studies with some additional information concerning the number of samples, administrations, and participants. It also includes the level of the course that is involved (undergraduate or graduate), the field of study (e.g. psychology, education, engineering) and some remarks. Most authors do not provide information on the specific content of the course (probability, descriptive statistics or inferential statistics). We acknowledge that differences in courses, fields of study, and other characteristics of the population and the specific statistics courses in the different studies can complicate the comparison. Yet, because most studies include an introductory statistics course in the field of human sciences (education, psychology), a prudent comparison seems justified.

In the following tables, we summarize the findings of these studies. Successively, we review (1) the internal consistency and test-retest reliability, (2) mean data (and standard deviations) for the Course and Field subscales (respectively for undergraduate and graduate students), and (3) the relationship with first year statistics exam results. Since not all investigations mention all measures, some tables contain only a subset of the studies involved in our comparative analysis.

Table 1 presents the observed internal consistency (Cronbach alphas). All studies yield coefficient alpha reliability estimates that are high for both subscales and for both administrations. In general, the estimates are between 0.77 and 0.93 for the Course subscale and between 0.83 and 0.96 for the Field subscale. Some studies (Elmore and Lewis 1991; Elmore et al. 1993; Roberts and Reese 1987) also mention the alpha estimate for the whole scale. Roberts and Reese (1987) find a whole scale alpha estimate of 0.91, Elmore and Lewis (1991) report for the first and the second administration an estimate of 0.92 and 0.93, respectively, and Elmore et al. (1993) 0.92 and 0.94.
Table 1: Internal Consistency (Cronbach alphas) for the two subscales of the ATS scale

| Study                        | Sample size | Course subscale | Field subscale |
|------------------------------|-------------|-----------------|---------------|
|                              |             | Adm. 1 | Adm. 2 | Adm. 1 | Adm. 2 |
| Aldogan and Aseeri, 2003      | 178         | 0.92    | 0.90     |        |        |
| Elmore and Lewis, 19911       | 58          | 0.90    | 0.82 | 0.90    | 0.92     |
| Elmore et al., 1993          | 289         | 0.90    | 0.90 | 0.90    | 0.93     |
| Rhoads and Hubele, 2000       | 63 (Adm. 1) | 0.77    | 0.85 | 0.89    | 0.90     |
| Shultz and Koshino, 1998 (sample 1) | 36 | 0.85    | 0.92 | 0.96    | 0.96     |
| Shultz and Koshino, 1998 (sample 2) | 38 | 0.93    | 0.89 | 0.90    | 0.92     |
| Waters et al., 1998          | 302         | 0.90    | 0.90 | 0.83    | 0.86     |
| Wise, 1985                   | 92          | 0.90    | 0.90 |        |        |

Note 1: ‘Adm.’ (second row) stands for ‘administration’. Most studies include two administrations, namely one before and one after the statistics course.

Note 2: Shultz and Koshino (1998) include two samples. The first sample contains undergraduate students, the second sample graduate students (see the Appendix for more information).

Some authors also investigate the test-retest reliability for the Course and Field subscales. The reported correlations are respectively 0.91 and 0.82 (Wise 1985), 0.59 and 0.72 (undergraduates, Shultz and Koshino 1998), and 0.71 and 0.76 (graduates, Shultz and Koshino 1998). For Wise (1985) there are only two weeks between the test and retest (as opposed to three months for Shultz and Koshino 1998). Obviously, the time lapse between administrations can affect the reliability. Table 2 presents the mean scores (and standard deviations) for the different studies. For all these data, if needed, item responses were reversed so that a higher score always refers to a more positive attitude. A distinction is made between undergraduate and graduate courses, since Shultz and Koshino (1998) predicted and found consistent differences in attitudes between these two groups when discussing their own and previous study results.

Because the ATS-items are scored on a Likert-type scale with five response possibilities, ‘strongly disagree’ (score 1), ‘disagree’ (score 2), ‘neutral’ (score 3), ‘agree’ (score 4) and ‘strongly agree’ (score 5), 27 indicates an average neutral position for the whole Course subscale, which contains 9 items. Similarly, because there are 20 Field subscale items, with each time ‘neutral (score 3)’ as the neutral response possibility, 60 indicates an overall neutral position for the whole Field subscale.
Table 2: Mean scores (and standard deviations) for the two subscales of the Attitude Toward Statistics scale

| Study                          | Sample size | Course subscale | Field subscale |
|-------------------------------|-------------|-----------------|----------------|
|                              |             | Adm. 1 | Adm. 2 | Adm. 1 | Adm. 2 |
| Undergraduate                 |             |         |         |         |         |
| Elmore et al., 1993           | 289         | 24.1   | 22.1   | 79.4   | 80.2   |
|                               |             | (7.8)  | (8.5)  | (9.5)  | (11.1) |
| Mvududu, 2003 (sample 1)      | 120         | 34.9   |        | 79.5   |        |
|                               |             | (6.0)  |        | (8.9)  |        |
| Mvududu, 2003 (sample 2)      | 95          | 28.9   |        | 74.0   |        |
|                               |             | (8.0)  |        | (13.1) |        |
| Shultz and Koshino, 1998 (sample 1) | 36   | 23.3   | 24.0   | 74.5   | 74.3   |
|                               |             | (6.5)  | (8.8)  | (11.8) | (11.7) |
| Waters et al., 1988           | 212         | 28.3   | 30.2   |        |        |
| Graduate                      |             |         |         |         |         |
| Elmore and Lewis, 1991        | 58          | 30.5   | 33.1   | 79.0   | 80.5   |
|                               |             | (7.4)  | (6.3)  | (9.8)  | (10.9) |
| DAndrea and Waters, 2002      | 17          | 29.1   | 35.2   | 84.9   | 86.6   |
|                               |             | (9.0)  | (5.7)  | (9.2)  | (6.7)  |
| Shultz and Koshino, 1998 (sample 2) | 38   | 29.8   | 32.5   | 81.1   | 81.3   |
|                               |             | (8.9)  | (7.1)  | (9.2)  | (9.6)  |

Note: Waters et al. (1988) do not provide standard deviations.

A comparison of the mean results for the undergraduate and graduate courses is in line with the conclusion of Shultz and Koshino (1998) that, in general, graduate students have higher scores than undergraduate students, for both the *Course* and *Field* subscale.

Table 3 shows the correlations between the attitude scores and the first year statistics exam results. In addition to the statistical significance of the correlations (which is discussed in all articles), we report effect sizes. Cohen (1988, 1992) provides a classification of effect sizes for correlations in terms of small ($r = 0.1$), medium ($r = 0.3$), and large ($r = 0.5$) effects as compared to the effects typically found in the social, educational and behavioral sciences. Except for Shultz and Koshino (1998), all studies demonstrate a statistically significant positive correlation between the first administration of the *Course* subscale scores and the exam results (first column). According to the guidelines of Cohen (1988, 1992), the corresponding correlations are small to medium. The correlations of the second administration (second column) are higher (effect sizes ranging from medium to large), and statistically significant for all studies. None of the studies shows a statistically significant correlation between the *Field* subscale.
scores and the exam results for the first administration (third column). Two studies (Shultz and Koshino 1998, first sample; Waters et al. 1988) show a statistically significant correlation for the second administration (fourth and sixth column), but for all studies in the table, the correlation at the second administration is smaller for the Field subscale as compared to the Course subscale.

Table 3: Correlations between ATS and first year exam results

| Study                         | Sample size | Course subscale | Field subscale |
|-------------------------------|-------------|-----------------|---------------|
|                               |             | Adm. 1          | Adm. 2        | Adm. 1          | Adm. 2      |
| Shultz and Koshino, 1998      | 36          | 0.06 (ns)       | 0.45 (p < 0.05) | 0.16 (ns)       | 0.43 (p < 0.05) |
| (sample 1)                    |             |                 |               |                |             |
| Shultz and Koshino, 1998      | 38          | 0.13 (ns)       | 0.34 (p < 0.05) | 0.13 (ns)       | 0.08 (ns)   |
| (sample 2)                    |             |                 |               |                |             |
| Rhoads and Hubele, 2000       | 63 (Adm. 1) | 0.29 (p < 0.05) | 0.29 (p < 0.05) | ns             | ns          |
|                               | 61 (Adm. 2) |                 |               |                |             |
| Waters et al., 1988           | 302         | 0.20 (p < 0.05) | 0.42 (p < 0.05) | 0.07 (ns)       | 0.17 (p < 0.05) |
| Wise 1985                     | 70          | 0.27 (p < 0.05) |                 | -0.04 (ns)     |             |

Note1: None of the authors report exact $p$-values. (‘ns’ stands for ‘not significant’)

Note2: Rhoads and Hubele (2000) do not provide exact correlation values for the Field subscale.

These data are in line with the conclusion of Waters et al. (1988) that there exists a consistent positive relationship between students’ attitudes toward statistics and their first year statistics exam results. They notice that especially the Course subscale scores are related to the statistics exam results, as also reported by Harvey et al. (1985, in Mvududu 2003). The latter authors suggested that a supportive atmosphere in the course can positively affect performance, regardless of the attitudes toward the field of statistics.

3. Method

3.1 Participants

Participants of the present study are 264 students (218 female, 46 male) who took an introductory undergraduate statistics course in Autumn 1996 at the Department of Educational Sciences of the Katholieke Universiteit Leuven in the Flemish speaking part of Belgium. In general, most of the students who are enrolled in this academic program follow a curriculum with a considerable amount of mathematics in secondary school (four, six or eight hours every week).
The curriculum of Educational Sciences takes five years to complete. The introductory statistics course takes place in the first semester of the first year. In general, the course deals with some introductory methodology and statistical concepts (tables, figures, and descriptive statistics), but no formal probability or statistical inference. The mathematical background required for this course is limited.

### 3.2 Instrumentation

Attitudes toward statistics are assessed with a Dutch translation of the Attitudes Toward Statistics (ATS) scale (Wise 1985). The ATS is administered twice. The first administration, in October 1996, at the beginning of the first year’s introductory statistics course ($n = 264$) and the second, in October 1997, at the start of the same students’ second year statistics course. In contrast to the studies mentioned in Section 2, the second administration takes place after the students know their exam result for the first year. About half of the students succeed in the first year. Therefore, the sample size of the second administration is much smaller ($n = 127$) and only includes students who succeed in their (overall) first year (eight of these 127 students had not been successful in their statistics exam, but nevertheless got permission to pass to the second year).

To relate the attitude scores to statistics performance, we record students’ statistics exam results and their dissertation grades at the end of the five year program. Again, this is only done for the students who completed the ATS and who completed the program successfully (see Table 4 for an overview of the sample sizes). For the statistics exam results, there are three results from obligatory statistics courses that students have to follow during their curriculum, namely in the first, the second, and the third year. For the first and the second statistics courses, the instructor is the same. For the third year’s statistics course, the same teacher as in the two previous years teaches half of the course, and another teacher teaches half of it. It is important to notice that the third year results are somewhat atypical and more difficult to interpret, because the course is evaluated through group assessment. Students do not follow any statistics courses in the fourth and fifth year, but because of the major role of methodology and statistics in a student’s dissertation, we consider this as a partial indication of long-term statistics performance.

To relate the attitude scores to general performance, we record students’ general exam results for the five years of the curriculum. For the present study, we excluded the dissertation grade from the variable ‘general exam result’, as it contributes 50% to that result.

All these measures together make it possible to relate the attitude scores of the two administrations at the beginning of the curriculum with (1) short- and long-term and (2) statistics and general exam results. As mentioned before, the conclusions of the relationship between the attitudes and long-term exam results only pertain to the students who actually pass the exams. Table 4 provides an overview of the different measures and of the sample sizes at each moment of data collection.

| Year | ATS | Statistics exam result | General exam result |
|------|-----|------------------------|---------------------|
| 96 - 97 | 1<sup>st</sup> administration (October 1996) | 1<sup>st</sup> year ($n = 234$) | 1<sup>st</sup> year ($n = 234$) |
3.3 Analyses

Reliability of the ATS is evaluated using both internal consistency (Cronbach alpha) and a test-retest reliability coefficient (correlations between ATS subscale scores on the first administration in October 1996 and the second administration in October 1997). Mean scores and standard deviations are calculated for both subscales.

The relationships of the attitude scores with the short- and long-term exam results are examined by Pearson product-moment correlation coefficients, separately for statistics and general exam results. The relationships of the attitude scores with all these exam results are compared with the relationships of first year exam results with later exam results. In other words, cognitive and affective predictors of exam results are compared, again separately for statistics and general exam results.

4. Results

In this section, we present the results of our study concerning students’ attitudes toward statistics in the three directions defined in the Introduction. First, we provide measures of reliability of the ATS and mean data about students’ attitudes toward statistics. Second, results on the relationship between attitudes toward statistics and statistics exam results (and the dissertation grade) are presented. Third, we examine the relationship between the attitudes toward statistics and general exam results.

4.1 Reliability and mean scores

The alpha estimates are high for both administrations, namely respectively 0.89 and 0.91 (Course subscale) and 0.86 and 0.86 (Field subscale). The whole scale alpha estimates are respectively 0.91 and 0.89. These results of internal consistency are similar to those mentioned in Section 2. The test-retest reliability analyses show (considerably high) correlations of 0.62 for the Field subscale and 0.76 for the Course subscale. These figures are higher than those reported by Shultz and Koshino (1998) (although for our study the time lapse between the two administrations is longer), but lower than those reported by
Wise (1985), where there were only two weeks in between the two measures.

The average Course subscale scores for the two administrations are respectively 28.5 ($s = 6.4$) and 30.7 ($s = 6.5$), indicating a rather positive attitude toward the statistics course (given that the neutral score is 27). Concerning the attitudes toward the course, our sample of undergraduate Flemish students is comparable with the (higher) graduate student scores observed in other studies. The average Field subscale scores, 66.9 ($s = 7.6$) and 68.0 ($s = 6.7$), respectively, are also positive (above the neutral score of 60), but compared to the other studies, these scores are low. Finally, the standard deviations of the Field and the Course subscale scores in our study are lower than in the other studies.

4.2 Relationship between attitudes toward statistics and statistics exam results/dissertation grade

Table 5 presents the correlations of the ATS subscale scores with all statistics exam results. In the last column of this table, we also mention the correlations between the first year statistics exam results and the other statistics exam results. Important to notice is that due to the partly different samples, comparisons must be made carefully. Therefore, we will concentrate on a comparison of correlations where the same students are involved. (The same analyses, restricted to the 72 students who have a measure on all variables, however, revealed the same trends in the data. On request, these data are available from the authors.)

| Statistics exam | 1st administration | 2nd administration | Statistics exam |
|-----------------|---------------------|---------------------|-----------------|
|                 | N | Course | Field   | N | Course | Field   | N | 1st year |
| 1st year        | 234 | 0.33 ($p < 0.001$) | 0.15 ($p = 0.02$) | 127 | 0.47 ($p < 0.001$) | 0.20 ($p = 0.03$) | 127 | 0.45 ($p < 0.001$) |
| 2nd year        | 102 | 0.23 ($p = 0.02$) | 0.14 ($p = 0.17$) | 115 | 0.31 ($p < 0.001$) | 0.20 ($p = 0.04$) | 115 | 0.26 ($p = 0.01$) |
| 3rd year        | 78 | -0.03 ($p = 0.79$) | -0.01 ($p = 0.94$) | 88 | 0.22 ($p = 0.04$) | 0.07 ($p = 0.53$) | 88 | 0.19 ($p = 0.08$) |
| 5th year        | 72 | 0.09 ($p = 0.44$) | 0.04 ($p = 0.75$) | 83 | 0.03 ($p = 0.78$) | 0.23 ($p = 0.04$) | 83 | 0.19 ($p = 0.08$) |

Note: Numbers in bold indicate significant values at the 0.001 level.

Because we are carrying out a large number of statistical tests on the same data, we have to take into account that the probability of committing at least one Type I error is substantially larger than the significance level set for each individual test. Multiple correlations are calculated and tested, the ones in Table 5 and Table 6, and additional tests are performed to compare correlated correlation coefficients.
To avoid potentially spurious results, we perform a Bonferroni correction on the overall significance level (0.05). The resulting significance level for an individual test is 0.001, which means that a p-value must be smaller than 0.001 in order to conclude that the correlation differs from zero.

For the first year, the results show statistically significant (p-values < 0.001) positive correlations between the attitudes toward the course and the statistics exam results. Although the correlations for the Field subscale are not statistically significant after the Bonferroni correction, all effect sizes (Cohen 1988, 1992) range between small and medium. The Course subscale scores show the highest correlations for both administrations, meaning that for the included sample the attitudes toward the course are a slightly better predictor of the first year exam results than attitudes toward the field. The test for comparing correlated correlation coefficients provided by Meng, Rosenthal, and Rubin (1992) shows that the difference between the correlations (Course versus Field) is statistically significant for the second administration (Z = 2.46, p = 0.01 for the first administration and Z = 3.39, p < 0.001 for the second administration). These results are a compelling replication of the findings from the earlier studies summarized in Table 3. However, recall that the students in our study already knew their exam results during the second administration of the ATS.

For the second year, the trends are similar, but differ in terms of statistical significance. The Course subscale scores are the most highly related to the second year statistics exams scores. However, only for the second administration is the correlation between Course and exam results statistically significant. The test for comparing correlated correlation coefficients shows that the difference between the correlations (Course versus Field) is not statistically significant for the second year (Z = 1.02, p = 0.31 for the first administration and Z = 1.27, p = 0.20 for the second administration). The effect sizes (Cohen 1988, 1992) of the correlations still range between small and medium.

For the third year, the attitude scores do not show statistically significant correlations with the statistics exam results. However, recall that we have to be careful with the interpretation of the data from the third year statistics exam results, because they are based on group assessment (see Section 3.2).

In the fifth year, the attitudes scores do not correlate significantly with the dissertation grade, but when we take a closer look at the results, we see that the Field subscale scores for the second administration show a substantive correlation with the dissertation grades in the fifth year (r = 0.23, p = 0.04).

Furthermore, in contrast to the correlation of the second administration with the first year statistics exam results, where Course was related highest to statistics exam results, in the long term, Field is more highly related to the dissertation grade than Course (test for comparing correlated correlation coefficients: Z = -1.93, p = 0.05).

Because this study is one of the first to explore the relation between attitudes toward statistics and long-term results, and because of the negative impact that Bonferroni corrections can have on the power of the tests, this correlation between the Field subscale and the dissertation grade – although no longer statistically significant after the Bonferroni correction – is worth mentioning.

The last column of Table 5 shows the correlations between the first year statistics exam results and all following statistics exam results (including the dissertation grade). Since these data relate to the same students as those who participated in the second administration of the ATS, the relative predictive values of affective (ATS) and cognitive (first year statistics results) characteristics in predicting later exam results can be compared for that administration.

Not surprisingly, the second year statistics exam results are more highly correlated with the first year exam scores (r = 0.45, p < 0.001) than with the ATS scores (r = 0.31, p < 0.001 and r = 0.20, p = 0.04).
respectively). The test for comparing correlated correlation coefficients shows that this difference between the correlations is most convincing (although not statistically significant after Bonferroni correction) for the *Field* subscale ($Z = -2.28, p = 0.02$).

In the long term, the observed correlation for *Field* ($r = 0.23, p = 0.04$) is higher than the correlation between the first year exam results and the dissertation grade ($r = 0.19, p = 0.08$). Thus for our sample, in the long-term, the *Field* score of the second administration is a better predictor of the dissertation grade than the first year statistics exam result. In other words, the observed affective measure shows a higher correlation with the dissertation grade than the cognitive measure, although the test for comparing correlated correlation coefficients provided by *Meng et al. (1992)* shows that this difference between the correlations is not statistically significant ($Z = 0.31, p = 0.76$).

### 4.3 Relationship between attitudes toward statistics and general exam results

Table 6 presents the correlations of the ATS subscale scores with all general exam results. An inspection of this table reveals that the important role of attitudes toward statistics is specific for statistics performance (including the dissertation grade). There is no statistically significant correlation between the ATS scores and the short- and long-term general exam results. In our sample, the total grade in the first year is more highly correlated with the following general exam results than the attitude scales.

| General exam | 1st administration | 2nd administration | General exam |
|--------------|---------------------|---------------------|--------------|
|              | N  | Course | Field  | N  | Course | Field  | N  | 1st year |
| 1st year     | 234 | 0.16   | 0.07   | 127 | 0.17   | 0.12   | 127 | 1         |
|              |     | (p = 0.02) | (p = 0.26) |     | (p = 0.06) | (p = 0.20) |     |           |
| 2nd year     | 102 | 0.01   | -0.03  | 115 | 0.09   | 0.03   | 115 | 0.42      |
|              |     | (p = 0.91) | (p = 0.79) |     | (p = 0.32) | (p = 0.76) |     | (p < 0.001) |
| 3rd year     | 78  | -0.01  | 0.07   | 88  | 0.08   | 0.16   | 88  | 0.48      |
|              |     | (p = 0.92) | (p = 0.57) |     | (p = 0.44) | (p = 0.14) |     | (p < 0.001) |
| 4th year     | 74  | 0.13   | 0.17   | 84  | 0.04   | 0.13   | 84  | -0.01     |
|              |     | (p = 0.28) | (p = 0.15) |     | (p = 0.75) | (p = 0.23) |     | (p = 0.96) |
| 5th year     | 72  | -0.01  | 0.01   | 83  | -0.04  | 0.14   | 83  | 0.46      |
| (courses)    |     | (p = 0.95) | (p = 0.95) |     | (p = 0.75) | (p = 0.23) |     | (p < 0.001) |

Note: Numbers in bold indicate significant values at the .001 level.
5. Discussion

This study provides further insight into students’ attitudes toward statistics and into the relationship between these attitudes and (short- and long-term) statistics and general exam results.

First, as in previous studies reported in the first part of this article (see Table 1), we find a high internal consistency for the Attitude Toward Statistics (ATS) scale (Wise 1985). The test-retest reliabilities are fairly high (0.62 for the Field subscale and 0.76 for the Course subscale) and within the range of the reliabilities reported by previous studies.

Second, this study provides new descriptive data concerning students’ attitudes toward statistics. These data are somewhat different from the trends mentioned in the literature. More specifically, the results on the Course subscale indicate that Flemish undergraduate students in Educational Sciences have an attitude toward the particular course in which they are enrolled that is more positive than the attitudes of undergraduate students elsewhere, but comparable to the attitudes of graduate students in other studies. However, the analysis of the Field subscale scores reveals a relatively negative attitude toward the use of statistics in the students’ field of study as compared to the scores from graduate and undergraduate students from the ‘bench-mark’ studies.

Third, the analysis of the relationship between the ATS scores and short-term statistics exam results complemented findings obtained by other authors, namely that especially attitudes toward the course are related to short-term exam results.

Fourth, an innovative element in our study is that it also yields findings concerning the analysis of the relationship between attitudes in the beginning of the curriculum and the dissertation grade. While for short-term exam results, attitudes toward the course are more highly related to statistics exam results than the attitudes toward the field, the latter are more highly related to the fifth year dissertation grade than the attitudes toward the course. Our results suggest that students who recognize the importance of statistics for their field of study (in the case of the present study: educational sciences) will tend to obtain a better dissertation grade.

Fifth, this study also investigates the relative predictive value of affective (ATS) and cognitive (exam results) measures in predicting later exam results. The data show that the relationship between the attitudes toward the field after experiencing a statistics course (affective measure) and second year statistics exam results were smaller than between first year exam results (cognitive measure) and those second year exam results. This finding is similar to the findings of Fienberg and Halperin (in Roberts and Bilderback 1980), namely that a cognitive measure predicted statistics performance with slightly higher accuracy than the measure of attitudes toward quantitative concepts. However, this difference between the affective and the cognitive measure as predictors is smaller for the relation with the long-term dissertation grade. In fact, the relationship for the affective measure is even slightly (but not significantly) higher than the relationship for the cognitive measure. These results are an important indicator of the essential role attitudes toward statistics (besides cognitive characteristics) play for the development of statistical competence.

Obviously, replications of this research on the relationship between attitudes toward statistics and long-term statistics exam results are needed. For instance, a comparison of dissertation scores and other measures that can be used as indications of long-term statistics performance, such as exam scores and/or scores on more traditional or performance-based statistical tests problems, can provide a deeper insight into this relationship. Furthermore, it would be very interesting to follow up the non-successful students, under more to compare the attitudes and statistics performance of these students with the students who
did pass the exams.

Finally, results from this study reveal that the important relationship between attitudes toward statistics and statistics performance is content-specific. Indeed, we found no relationship between the attitudes toward statistics and general exam results. Further research should investigate how the attitudes measured by the ATS differ from ‘general academic attitudes’ and how different attitude scales are related to different kinds of performance. Such research might reveal the importance of a separate assessment of attitudes toward studying specific fields of study, besides the assessment of ‘general academic attitudes’.

### Appendix: Overview of the studies

| Study       | Sample | # Admin | n   | (Under)graduate | Field of Study | Remarks |
|-------------|--------|---------|-----|-----------------|----------------|---------|
| Wise 1985   | 1      | 1       | 92  |                 | Education      | Original article ATS |
|             | 2      | 2       | 70  |                 | Education      |         |

| Roberts and Reese 1987 | 1 | 1 | 280 | Undergraduate |             | Also administration of another scale to measure attitudes toward statistics, the Statistics Attitude Survey (SAS; Roberts and Bilderback 1980. ATS is treated as one scale in this study. |

| Waters et al. 1988    | 1 | 2 | 302 | Undergraduate | Variety of majors (mainly psychology) | Also administration of SAS. Only 212 respondents were measured on both occasions. |
| Authors            | Year | Adm | N  | Sample  | Major                                      | Notes                                                                 |
|--------------------|------|-----|-----|---------|--------------------------------------------|----------------------------------------------------------------------|
| Elmore and Lewis   | 1991 | 1   | 2   | 58      | Graduate Variant of majors                 | Used to measure change in attitudes before and after a computer-integrated statistics course. |
| Elmore et al.      | 1993 | 1   | 2   | 289     | Undergraduate Variety of majors            | Used to measure change in attitudes before and after a statistics course using ‘short stories’. |
| Shultz and Koshino | 1998 | 1   | 2   | 36      | Undergraduate Psychology                   | Cross-cultural study (USA and Zimbabwe). Used to measure the relationship between attitudes toward statistics and the use of constructivist strategies. |
| Rhoads and Hubele  | 2000 | 1   | 2   | 63      | Undergraduate Engineering                 | Cross-cultural study (USA and Zimbabwe). Used to measure the relationship between attitudes toward statistics and the use of constructivist strategies. |
| D’Andrea and Waters| 2002 | 1   | 2   | 32      | Graduate Education                         | Cross-cultural study (USA and Zimbabwe). Used to measure the relationship between attitudes toward statistics and the use of constructivist strategies. |
| Aldogan and Aseeri | 2003 | 1   | 1   | 178     | Graduate Variety of majors                 | Cross-cultural study (USA and Zimbabwe). Used to measure the relationship between attitudes toward statistics and the use of constructivist strategies. |
| Mvududu            | 2003 | 1   | 2   | 95      | Undergraduate                             | Cross-cultural study (USA and Zimbabwe). Used to measure the relationship between attitudes toward statistics and the use of constructivist strategies. |

Note 1: As can be seen from the table (column Sample), some authors use two different samples.
Note 2: # Adm. stands for the number of administrations for a particular sample.
Acknowledgments

The authors would like to thank the editor and the anonymous referees for their insightful comments and suggestions to improve earlier versions of this paper.

We want to thank Brian Greer for his comments and Tine Gheysen for her assistance in data collection and initial statistical analyses.

This research was partially supported by Grant GOA 2006/01 “Developing adaptive expertise in mathematics education” from the Research Fund Katholieke Universiteit Leuven, Belgium.

References

Aldogan, A., and Aseeri, A. (2003), "Psychometric characteristics of the Attitude Towards Statistics scale," *Umm Al-Qura University Journal of Educational and Social Sciences and Humanities*, 15(2), 99-114.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences* (2nd edition), Hillsdale, NJ: Lawrence Erlbaum Associates.

--- (1992), "A power primer," *Psychological Bulletin*, 133, 155-159.

D’Andrea, L., and Waters, C. (2002), "Teaching statistics using short stories: Reducing anxiety and changing attitudes," Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.

Elmore, P. B., and Lewis, E. L. (1991), "Statistics and computer attitudes and achievement of students enrolled in applied statistics: Effect of a computer laboratory," Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Elmore, P. B., Lewis, E. L., and Bay, M. L. G. (1993), "Statistics achievement: A function of attitudes and related experience," Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Gal, I., and Ginsburg, L. (1994), "The role of beliefs and attitudes in learning statistics: Towards an assessment framework," *Journal of Statistics Education* [online], 2(2). 
[ww2.amstat.org/publications/jse/v2n2/gal.html](http://ww2.amstat.org/publications/jse/v2n2/gal.html)

Gal, I., Ginsburg, L., and Schau, C. (1997), "Monitoring attitudes and beliefs in statistics education," in *The assessment challenge in statistics education*, Eds. I. Gal, and J. B. Garfield, Netherlands: IOS Press, pp. 37-51.

Meng, X. L., Rosenthal, R., and Rubin, D. B. (1992), "Comparing correlated correlation coefficients," *Psychological Bulletin*, 111, 172-175.
Mvududu, N. (2003), "A cross-cultural study of the connection between students’ attitudes toward statistics and the use of constructivist strategies in the course," *Journal of Statistics Education* [online], 11(3). [ww2.amstat.org/publications/jse/v11n3/mvududu.html](http://ww2.amstat.org/publications/jse/v11n3/mvududu.html)

Rhoads, T. R., and Hubele, N. F. (2000), "Student attitudes toward statistics before and after a computer-integrated introductory statistics course," *IEEE Transactions on Education*, 43(2), 182-187.

Roberts, D. M., and Bilderback, E. W. (1980), "Reliability and validity of a statistics attitude survey," *Educational and Psychological Measurement*, 40, 235-238.

Roberts, D. M., and Reese, C. (1987), "A comparison of two scales measuring attitudes towards statistics," *Educational and Psychological Measurement*, 47, 759-764.

Schau, C., Stevens, J., Dauphinee, T. L., and Del Vecchio, A. (1995), "The development and validation of the Survey of Attitudes Toward Statistics," *Educational and Psychological Measurement*, 55(5), 868-875.

Shultz, K. S., and Koshino, H. (1998), "Evidence of reliability and validity for Wise’s Attitude Toward Statistics scale," *Psychological Reports*, 82, 27-31.

Waters, L. K., Martelli, T. A., Zakrjsec, T., and Popovich, P. M. (1988), "Attitudes toward statistics: An evaluation of multiple measures," *Educational and Psychological Measurement*, 48, 513-516.

Wise, S. L. (1985), "The development and validation of a scale measuring attitudes toward statistics," *Educational and Psychological Measurement*, 45, 401-405.

Stijn Vanhoof  
Department of Educational Sciences  
Centre for Methodology of Educational Research  
Katholieke Universiteit Leuven  
Leuven  
Belgium  
[stijn.vanhoof@ped.kuleuven.be](mailto:stijn.vanhoof@ped.kuleuven.be)

Ana Elisa Castro Sotos  
Department of Educational Sciences  
Centre for Methodology of Educational Research  
Katholieke Universiteit Leuven  
Leuven  
Belgium  
[anaelisa.castro.sotos@ped.kuleuven.be](mailto:anaelisa.castro.sotos@ped.kuleuven.be)

Patrick Onghena  
Department of Educational Sciences  
Centre for Methodology of Educational Research  
Katholieke Universiteit Leuven  
Leuven  
Belgium
