On the Average Perception of Perceived Averages

Vanessa C. Harvey, Patrice Farand, Robert Legros and Jason R. Tavares*
Department of Chemical Engineering, Polytechnique Montreal
P.O. Box 6079, station Centre-Ville, Montreal, Quebec (Canada), H3C 3A7

E-mail: jason.tavares@polymtl.ca
Telephone: 1-514-340-4711 ext. 2326
Fax: 1-514-340-4159
*Corresponding author

Abstract. The Chemical Engineering curriculum at Polytechnique Montreal is structured to gradually provide more and more autonomy to the students. The third-year Unit Operations is taught using an outcomes-based approach and represents a turning point in the undergraduate curriculum where rubrics-based assessments overtake normative assessments. This begs the question: is it really necessary to divulge the average to students following assessments? Those from a more industrial background see the average as an unnecessary crutch for students, while the more academically inclined see it as a useful pedagogical tool to provide feedback and help students determine if they have attained their learning objectives. To settle this debate, we set into motion a yearlong study during which the average results to tests were withheld. Students were asked to predict their grade and the class average, and provide feedback on the assessment process. Results show that students are able to predict the average, but have difficulty predicting their individual performance (especially before a test, where more than 50% of students are off by a factor of more than 10%). Students award more importance to their personal sense of learning satisfaction than their position with respect to the average, and do not systematically use the average to plan study time (despite preferring to know it). Thus, it may be possible to substitute alternate frames of reference to the class average in an outcomes-based course, but this is not necessarily desirable and should at the very least be the subject of a more open discussion.

Keywords: averages, outcomes-based approach, chemical engineering, unit operations, gender, perception

1. Introduction and context

1.1 Polytechnique Montreal's undergraduate chemical engineering curriculum

The 120-credit Chemical Engineering undergraduate curriculum at Polytechnique Montreal is a 4-year program constructed as a learning program-based approach (Pr gent et al., 2009). This has allowed for a high level of synergy between the various core classes in the curriculum. Ten years ago, this synergy allowed for the creation of several outcomes-based classes. The undergraduate curriculum committee accomplished this by first outlining 12 criteria that define the competence of a chemical engineering graduate, and then formulating a pedagogical design for course sequencing. Therefore, it is possible to monitor the evolution of a future engineer's competence over a series of classes, specifically through the use of rubrics (Stevens and Levi, 2005; Huba and Freed, 2000). The curriculum is structured to promote student autonomy and develop certain skills that are specific to chemical engineers. Overall, the implemented strategy aims to change the type of support available to students throughout their academic career. In the beginning considerable support is offered; this gradually changes from support to mentoring as the student advances in his/her career. This evolution is linked not only to progression of the course content, but the assessment frequency and structure used, transitioning from a normalized approach (comparing students to each other) toward a criterion-based approach (comparison of students based on pre-established criteria).

1.1.1 First year

Most students registered in the first year of the Chemical Engineering bachelor program have completed college or a similar level of schooling (K+13). The transition from college to university is an important step that requires students to make many adjustments. To make this easier for new students, we conduct courses with a traditional framework that is largely similar to what they already know. In the beginning of the program, many courses have several assessments (graded tutorials) that account for a small percentage of the final course grade. This strategy forces the students to work steadily, thus preparing for their exams more easily. Autonomy is therefore limited at first to promote their integration into the academic world and contribute to the development of good working habits.

Evaluations are corrected using traditional normative methods, which means scaling the grades so they are well distributed. In general, the grades are distributed along a normal curve. At the end of the assessment, the average, the standard deviation and a histogram showing the distribution of grades are presented to the students. These statistics help guide the professors when giving out final grades (A*, A, B+, B, C+, C, D+, D, F) for the course. In addition, using these statistics, students can determine how well they did in the course based on the average. At the end of their first year, the students complete an integrative project. This is the first course in which we begin assessing their skills using set rubrics.

1.1.2 Second year

In the second year, the professors give the students more autonomy. The courses have fewer assessments of lesser weight and more recommended sets of problems , pushing the students to study on their own. In addition, the nature of the assessments helps incorporate the concepts learned in the first year. For certain assessments, set rubrics are used. The schemes are provided to students before any evaluation, allowing them to clearly understand the criteria used to formulate their final grades. Since the grader makes use of rubrics, he/she is able to compare student performance in terms of these set criteria, thus greatly diminishing correction bias and forming a clearer picture of competence development over a course sequence (Tardif, 2006). Following the evaluation, each student receives his or her completed assessment grid. The average can be calculated based on the various grades obtained, but it has no direct impact on the students' final grade. Situating their performance on the assessment grid scale (clearly defined criteria) and not as a comparison to the group average is what is most important. However, the average can help the students to see where they stand compared to the group and therefore to make more effort, if needed.

1.1.3 Third year

In the third year, professors push the students to further develop their autonomy, with very little formal support, especially in situations they have already encountered. When students have questions about a course, they are encouraged to find the answer themselves or with the help of their fellow classmates. As a last resort, they are invited to consult their professors. The majority of assessments evaluate skills using set rubrics and some more traditional exams are also employed. As in the 2nd year, the professors give out the class average for information purposes, but do not use it to determine the various final grades for the course. The main course for third year students involves designing unit operations in chemical engineering and is worth 13 credits out of 30 for the year. In this course, students carry out a project worth 6 credits in which the assessment is based exclusively on a set rubric. Using this format, it is possible to evaluate the students' skills based on precise criteria and not by comparing students with one another. The assessment grids for this particular class have been optimized over the past ten years and the clarity of the criteria used makes it possible to unambiguously define our expectations from the beginning. The students are therefore able to produce better quality work than with a more traditional method of assessment. When giving out the final letter grades for the course, the professors never use an average. The student's grade is compared to the pre-determined expectations without taking into consideration the performance of the other students. Nonetheless, for the few remaining normative evaluations (quizzes namely), averages are still disclosed to the students.

1.1.4 Fourth year

For fourth year students, we conduct courses where the students are very autonomous in the majority of decisions they make. The support offered by the professors is minimal and focuses mainly on orienting the students to useful resources because the required expertise may be outside of the competency of the professor involved in the different industrial projects. The assessment criteria are clearly defined from the beginning and directly related to industrial standards. Professors and the department's industrial partners have jointly developed these criteria and the assessment focuses only on the students' skills. In the course-project (capstone design project), both the professors and the industrial partners grade the students without taking into consideration the performance of other students. In this case, there is no average calculated.

<u>1.2 Debate on the necessity of divulging class averages</u>

This evolution from mostly normative to essentially outcomes-based evaluation brings about a pertinent question: is it still necessary to divulge the class average to students following a given evaluation in an outcomes-based class environment, specifically in the third year of the above-described curriculum? Indeed, in a traditional, normative (curve-based) class, the class average may be seen as a useful tool allowing the student to situate himself/herself with respect to his/her colleagues. However, in an outcomes-based context, the average has no bearing on the student's final performance. In fact, since students are compared to pre-established criteria, it is possible that a large number of students receive grades below or above expectations. Moreover, some faculty members have brought up the point that, once on the job market, a young engineer's performance (and remuneration) will typically not be based on a comparison with colleagues, but rather on his/her ability to meet or exceed goals set forth by management. This philosophy is reflected in the fourth-year teaching approach. On the other hand, companies may compare one individual to another in a hiring context, though hiring may also be based on set criteria for teamwork, communication or specific technical skills. As stated previously, a transition from normative to outcomes-based evaluations takes place in the second year of the Polytechnique curriculum, though averages remained divulged for a small set of evaluations in the third year (despite having no bearing on the final grade). The present work aims to study the pertinence of divulging class averages in a competence-based approach, both through a literature survey and an empirical investigation conducted in Polytechnique Chemical Engineering's flagship class for outcomes-based education in the third year: Unit Operations.

1.3 Unit Operations

The Unit Operations class (GCH3100) is an innovative course taken in the 3rd year of the undergraduate curriculum. It is structured around four principal chemical engineering themes: applied fluid mechanics, applied heat transfer, separation processes and process control. The class aims for development of discipline-specific competences in parallel with professional, organisational, reflective, personal and human attributes, through authentic evaluation mechanisms (Wiggins, 1998; Huba and Freed, 2000; Roegiers, 2000; Scallon, 2004; Tardif, 2006; Pr gent et al., 2009). The course is divided into two parts:

- Theory (7 credits) – developing knowledge and know-how for unit operations;
- Practice (6 credits) – competence development through project-based learning and practice.

This 13-credit workload represents almost the entire load for the semester in question (credit loads average 15 per semester). Students spend approximately 20 hours per week in class (roughly 50% of the total workload) and are taught by a team of 12 instructors. Class sizes range from 25 to 75 students. A specially designed classroom is dedicated to this course, as it is adapted for both traditional lectures and team projects. The integrated approach to teaching allows students to form stronger links between the various chemical engineering sub-disciplines and promotes a more autonomous learning experience. The evaluation mechanisms used to evaluate the development of knowledge and know-how are summarized in Table 1. The practice section of the course is entirely evaluated using rubrics, while the theory section is evaluated through traditional grading (though the final grade relies on a set rubric). It is pertinent to note that rubrics are used throughout the curriculum: students are therefore accustomed to knowing the evaluation criteria beforehand to better understand the instructor's requirements.

Table 1 – Evaluation mechanisms in Unit Operations

Theory section (7 credits)			**Practice section (6 credits)**		
Evaluation tools	Number	Value	Evaluation tools	Number	Value
Guided tutorials	4	15%	Learning modules (mini-projects)	2	15%
Tests (weeks 4 and 8)	2	30%	Project	1	25%
Final exam (week 13)	1	55%	Oral defence	1	25%
			Laboratory report	1	20%
			Team work evaluation	1	2.5%
			Journal article review	1	2.5%
			Document research quiz	1	5%
			Project management homework	1	5%

1.4 Literature survey

1.4.1 Importance for students

For students, the class average for an exam or a quiz is generally regarded as a welcome piece of information. It is used to compare one's own performance to that of the group, and thus serve as a ranking tool. While it is appreciated, students may not need it to form a reference framework. Skaalvik and Skaalvik (2002) put forth that students are able to instinctively formulate both internal and external markers for comparison. External markers include their school's ranking or reputation, the class average (from previous years), or comments from instructors or from particular students in or out of their class, such as students that often answer questions in lectures or those who take the longest to complete a test. Internal markers will vary from student to student and may include their own absolute performance in other classes or fields of study, their relative performance with respect to specific learning objectives, their perception of a given field of study over time, and the amount of time and effort attributed to a given course. The authors even speculate that divulging the average may be detrimental to certain students' self esteem (note: averages are not typically divulged in Scandinavian countries, where the Skaalvik and Skaalvik (2002) study was conducted).

1.4.2 Gender differences

Students generally tend to overestimate their performance (Wesp et al., 1996). However, looking more closely, several studies have determined that success is indeed perceived differently between males and females: whereas the males tend to overestimate their degree of success, females underestimate their performance. While it is quite clear that general intelligence is not a gender trait (Colom et al., 2000), the reason behind the perception bias remains unclear. Some studies explain this through differences in personalities in a given population (Soh and Jacobs, 2013; Oren, 2012): self-reported introvert students are generally better able to predict their performance on a given task, whereas their more extrovert counterparts would overestimate performance for the same task. The personality traits related to extrovert behavior tend to be more prevalent in males than females.

Others evoke the remnants of ancient stereotypes that may influence how men and women self-evaluate their position in a group (Syzmanowicz and Furnham, 2011; Beloff, 1992). The Syzmanowicz and Furnham (2011) analysis examined self-evaluation results on a wide range of topics, including general intelligence, mathematics and logic, spatial reasoning and verbal intelligence. Aside from verbal intelligence, men tend to overestimate their abilities in all other categories. However, this study did not account for the individuals' age, which may play a part in the prevalence of certain stereotypes. On the other hand, one recent study conducted in an engineering design course demonstrated that there is no significant difference between the self-evaluation of male and female students (Van Tyne, 2011). This is further corroborated by a more wide ranging study conducted by Tariq and Durrani (2011). Van Tyne hypothesizes that this result is directly correlated to the fact that the students polled were in science, technology, engineering and mathematics (STEM) fields; in other words, if a female student enrolls in a traditionally male-oriented field, she is likely to have higher self-confidence and, thus, a better handle on her own abilities. It may also be that engineering students tend to be more realistic or quantitative than the general population.

1.4.3 Optimism and evolving perceptions

Students that are optimistic with respect to their studies generally have a more positive outlook on evaluations and their ability to control outcomes (Hall et al., 2006; Ruthig et al., 2009; Perry, 1991; Ackerman and Wolman, 2007). This positive outlook is also related to their ability to predict academic success, and more generally, to attain actual academic success (Chemers et al., 2001). Moreover, Ruthig et al. (2009) demonstrated that students' optimism and perception of control did not appear to evolve over the course of a semester, regardless of the feedback received after evaluations. However, there does appear to be an evolution of this optimism before and after a given exam: students overestimate their result before a quiz, but their predictions tend to be closer to reality immediately afterwards (Shepperd et al., 1996). Wolf (2007) would argue that the aim of providing feedback through evaluations is to ensure that students are able to gradually improve their performance over the course of a semester. As a result, if their performance perception mechanisms remain constant, one would expect their predictions to evolve over time.

1.5 Study objectives

Through the present study, we will learn more about how undergraduate Chemical Engineering students at Polytechnique Montreal perceive averages and performance. Specifically, we will determine if:

- Divulging the average in an outcomes-based course is a useful pedagogical feedback tool;
- Students are able to adequately evaluate and predict their performance in a core chemical engineering class.

2. Method

The study took place over the course of 2 semesters (Fall 2012 and Winter 2013) in the third-year unit operations course and involved a total of 56 students (26 males, 30 females). Students received a detailed presentation about the study, were informed that the class average would not be divulged by any instructor over the course of the semester, and were required to sign an informed consent sheet to participate. The participants were then required to fill out seven questionnaires over the course of the semester:

- A prediction sheet during the first week of class, in which they are asked to determine which grade they expect to receive for the class, as well as their opinion on the importance of class averages as a learning tool;

- Three pre-exam prediction forms handed to them in the week preceding a test (2 tests, 1 final exam), in which they are asked to predict their grade and the class average, as well as their general feelings about the test;
- Three post-exam prediction forms handed to them immediately after completing a test (2 tests, 1 final exam), identical to the pre-exam forms (with one exception – the final questionnaire included one extra question pertaining to their appreciation of the experience).

It is pertinent to note that, despite the total sample size of 56 students, students did not systematically answer all questionnaires for a variety of reasons (absent from class, time restrictions after tests, partially filled forms, etc.). The retained sample size is indicated accordingly in the results section.

3. Results and discussion

3.1 Average as a pedagogical tool

Third-year students were asked to provide their opinion with respect to several statements pertaining to evaluations and, more specifically, how they feel about averages during the first week of classes (Table 2). While students tend to award a fair level of importance to the average (#1, 88% agreement), there is a split concerning its use as a pedagogical tool: 43% of students will not use this data to plan study time allotments for subsequent test (#2). Students generally seem to feel that their grade is typically representative of their level of understanding (#3, 73% in agreement) and that an above-average grade (#4), specifically a grade significantly exceeding the average (#5) is indicative of their performance. Statement #6 indicates that another, more personal metric (“feeling that I’ve learned something”) could provide an alternative to the average as a pedagogical tool, and it seems to be less rooted in comparisons to others (though 64% still agree with statement #7).

Table 2 – Number of students in agreement with statements pertaining to evaluations and averages (*N = 56*). Top two answers are shaded, most popular answer is darkest.

Statement	Totally disagree	Disagree somewhat	Agree somewhat	Totally agree
1. Knowing the average following a written evaluation is important to me.	2	5	26	23
2. My position with respect to the average for a given test is a decisive factor when allotting study time for subsequent tests	2	22	24	8
3. My grade following a written evaluation is typically representative of how much I understood from a course	1	14	32	9
4. In a written evaluation, an above-average grade is a strong indicator of my success	1	11	26	18
5. The difference between my grade and the average is a strong indicator of my performance in a written evaluation	1	10	30	15
6. Feeling that I've learned something is a strong indicator that I've succeeded	0	2	23	31
7. Feeling that I've learned more than my colleagues is a strong indicator that I've succeeded	3	17	20	16

3.2 Student predictions

Figure 1 compares the average of students' individual pre- and post-test predictions to their actual results, for all the evaluations made during the semester. The error bars shown represent the 95% confidence interval, thus providing information with respect to the spread of the data. Furthermore, although the data is not shown, the students' predictions of the class average show no statistical difference with respect to the actual computed average for all three evaluations (Student t-test with $\alpha = 0.05$). In a sense, this is expected, given that the class averages for this particular course has remained relatively unchanged over the past decade, and this cohort of students has studied together for three years. The Figure 1 results could be interpreted as evidence that students are able to adequately predict their performance in an evaluation, and remain consistent in that ability throughout the course of the semester (except in the case of the pre-test prediction for the final exam, which appear overestimated). If this is indeed the case, it implies that their alternative frames of reference, possibly formed through the first two years of the undergraduate curriculum, are sufficient and do not degrade as a result of not receiving class averages.

[Figure 1 about here]

However, plotting individual actual results against predictions for the final exam (test with the largest amount of complete prediction data) nuances this conclusion (Figure 2). Indeed, Figures 2a and 2b show that approximately two thirds of both male and female students are well outside of the +/- 10% margin to have their pre-test predictions considered as adequate (13 out of 19 males and 16 out of 23 females are incorrect – N=42). Moreover, both male and female students seem to share a pre-test optimism: most over-predict their performance by a wide margin. This level of over-prediction diminishes slightly when asked after the test (Figures 2c and 2d): 9 out of 19 males and 10 out of 23 females fall within the +/- 10% range. The data spread is also noticeably reduced (especially on the male side) and more symmetrical for post-test results. These individual results re-orient the conclusion stemming from Figure 1: students are capable of predicting the class average, but have greater difficulty predicting their own grade. In other words, students do not appear to be able to correctly self-assess their performance. If they were, some of them could use this information as a replacement for the average as a feedback tool and thus adjust their study patterns for the next assessment (as noted in Table 2, statement #2, over 40% of students would not see this as a useful study tool).

[Figure 2 about here]

In order to assess any gender-based effects, average performances by gender are plotted in Figure 3. Although upon first inspection, gender split results for the final exam seem to show a slight trend in agreement with the thesis brought forward by Syzmanowicz and Furnham (2011) whereby men tend to overestimate their capabilities, statistical analysis clearly demonstrates overlap in the confidence interval (Student t-test with $\alpha = 0.05$). In other words, the results are more in line with the observations made by Van Tyne et al. (2011): women enrolled in an engineering degree show no significant difference in estimating their capabilities compared to men. This is confirmed in part by the results from Figure 2 – there are not noticeable differences between male and female students when inspecting individual prediction results.

[Figure 3 about here]

Finally, student appreciation is quantified in Figure 4: when asked to rate the experience of not receiving the class average after a test over the course of a semester, the overwhelming majority indicated that they disliked the experience. This serves to indicate that students are quite set in their ways: they are not comfortable with loosing a quantitative frame of reference they have received throughout their studies (typically from high school onwards). It is pertinent to observe that 28% of students surveyed did appreciate the experience – these students correlate almost perfectly with those who indicated for statement #1 in Table 2 that knowing the class average was not particularly important to them, and show fair agreement (approximately 2/3) with the students who disagreed that their grade reflects their level of understanding (statement #3). In other words, students who normally rely on internal signals for motivation are not affected by external signals. It should further be noted that no negative pedagogical bias was observed: class averages and grade distributions were well in line with historical values for the course (data not reported). Therefore, not receiving the average following evaluations does not appear to have negatively (or positively) impacted learning.

[Figure 4 about here]

4. Conclusions

This study has explored the necessity of divulging class averages for evaluations conducted in an outcomes-based class. The literature survey indicated that, for the most part, students form multiple frames of reference and should not be affected by the loss of a single one, such as the class average. An experimental investigation conducted over a full year has shown partial agreement with these results: students are able to adequately predict their peers' performance (class average), though they have greater difficulty in predicting their individual performance. Further questioning reveals that, while students do not necessarily use the class average as a pedagogical feedback to allot study time and rather rely on their personal level of satisfaction with learning as a performance indicator, they are significantly dissatisfied when this information is withheld.

Thus, based on this limited study, instructors teaching within a competence-based framework may likely safely withhold class averages, as it does not appear to be a necessary pedagogical tool. However, this should be done in accordance with a clear curriculum progression (gradually moving away from divulging class statistics, moving towards individual indicators) and through open discussion with the students to clarify why the information is being withheld (aligning with the industrial reality where salary and other employment considerations are based upon individual performance, not a comparison to averages). This will help avoid overwhelming dissatisfaction and contributing to additional stress to the students.

Acknowledgments

The authors acknowledge the moral support of Polytechnique Montreal’s Department of Chemical Engineering and, specifically, the professors involved in teaching the Unit Operations class (Profs. Charles Dubois, Danilo Klvana, Michel Perrier, Abdellah Ajji and Fran ois Bertrand), as well as instructors and participating students. Further, we acknowledge the support of the Polytechnique Montreal ethics in research committee for approval of this study conducted on human subjects (certificate number C R-11/12-28).

References

Ackerman, P., Wolman, S., 2007. Determinants and validity of self-estimates of abilities and self-concept measures. Journal of Experimental Psychology, 57-78.

Beloff, H., 1992. Mother, father, and me: Our IQ. The Psychologist, 309-311.

Chemers, M., Hu, L., Garcia, B., 2001. Academic self-efficacy and first-year college student performance and adjustment. Journal of Educational Psychology, 55-64.

Colom, R., Juan-Espinosa, M., Abad, F., Garcia, L., 2000. Negligible sex differences in general intelligence. Intelligence, 57-68.

Hall, N.C., Perry, R.P., Chipperfield, J.G., Clifton, R.A., Haynes, T.L., 2006. Enhancing primary and secondary control in achievement settings through writing-based attributional retaining. Journal of Social and Clinical Psychology, 361-391.

Huba, M.E., Freed, J.E., 2000. Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning. Allen & Bacon, Boston.

Oren, F.S., 2012. The effects of gender and previous experience on the approach of self and peer assessment: a case from Turkey. Innovations in Education and Teaching International, 123-133.

Perry, R., 1991. Perceived control in college students: Implications for instruction in higher education, in: Smart, J. (Ed.) Higher Education: Handbook of theory and research, 7, 1-56.

Pr gent, R., Bernard, H., Koznitis, A., 2009. Enseigner à l'université dans une approche-programme. Presses Internationales Polytechnique, Montreal.

Roegiers, X. (2000). Une pédagogie de l'intégration. Compétences et intégration des acquis dans l'enseignement. De Boeck Universit , Bruxelles.

Ruthig, J., Hanson, B., Marino, J., 2009. A three-phase examination of academic comparative optimism and perceived academic control. Learning and Individual Differences, 435-439.

Scallon, G., 2004. L'évaluation des apprentissages dans une approche par compétences. Montreal: ERPI.

Shepperd, J.A., Ouellette, J.A., Fernandez, J.K., 1996. Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. Journal of Personality and Social Psychology, 70(4), 844-855.

Skaalvik, E.M., Skaalvik, S., 2002. Internal and external frames of reference for academic self-concept. Educational Psychologist, 37(4), 233-244.

Soh, L., Jacobs, K.E., 2013. The biasing effect of personality on self-estimates of cognitive abilities in males and females. Personality and Individual Differences, 24, 141-146

Stevens, D.D., Levi, A.J., 2005. Introduction to Rubrics: An Assessment Tool to save grading time, convey effective feedback and promote student learning. 2nd edition, 232 pp., Stylus Publishing LLC, Sterling, Virginia.

Syzmanowicz, A., Furnham, A., 2011. Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta-analyses. Learning and Individual Differences, 21(5), 493-504.

Tardif, J. (2006). L'évaluation des compétences : Documenter le parcours de la formation. Cheneli re ducation, Montreal.

Tariq, V.N., Durrani, N., 2012. Factors influencing undergraduates' self- evaluation of numerical competence. International Journal of Mathematical Education in Science and Technology, 43(3), 337-356.

Van Tyne, N., 2011. Gender differences in individual and teammate performance evaluations by students on engineering design teams. American Society for Engineering Education, AC 2011-220.

Wesp, R., Montgomery, K., Jacobson, C., 1996. Students tend to overestimate their course performance. Psychological Reports, 79(3), 864-866.

Wiggins, G., 1998. Educative assessment: Designing assessments to inform and improve student performance. Jossey-Bass Publishers, San Francisco.

Wolf, P.J., 2007. Academic improvement through regular assessment. Peabody Journal of Education, 82(4), 690-702.

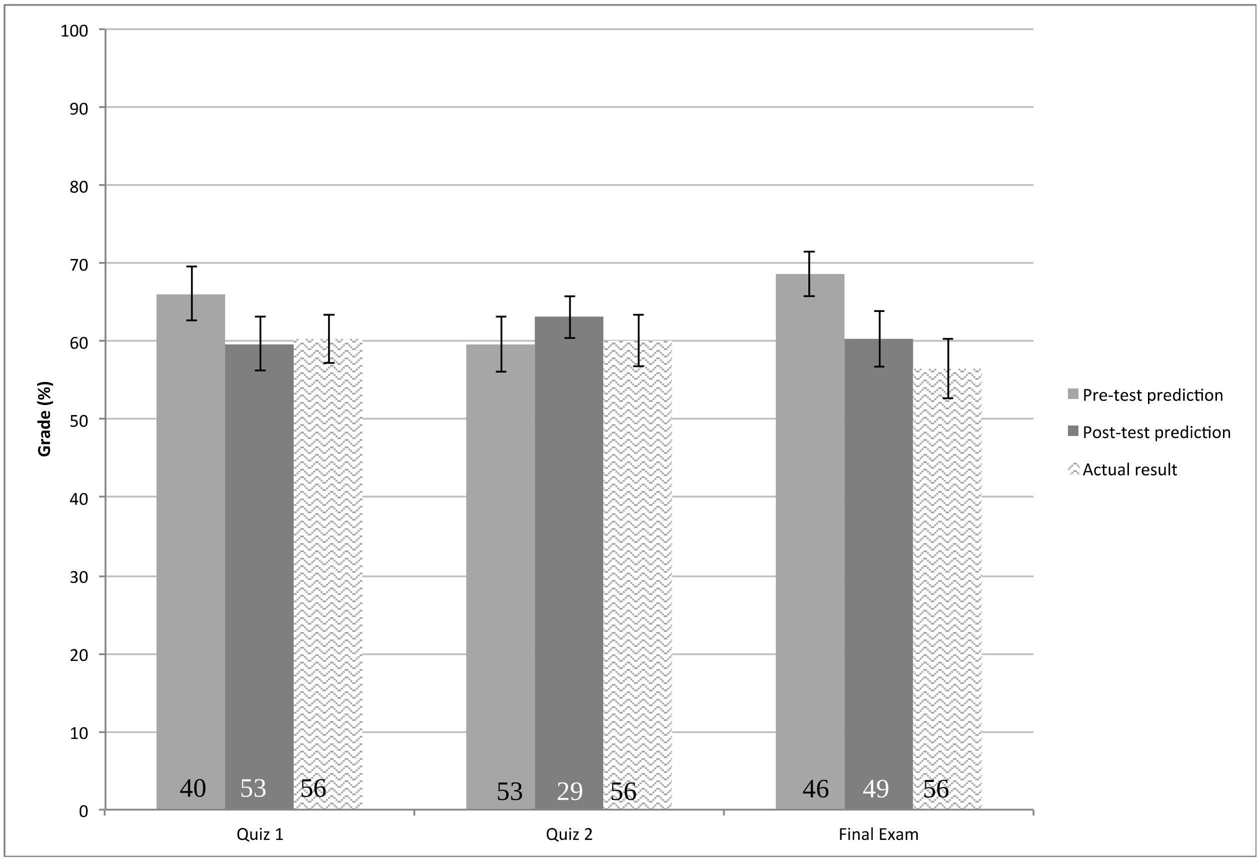
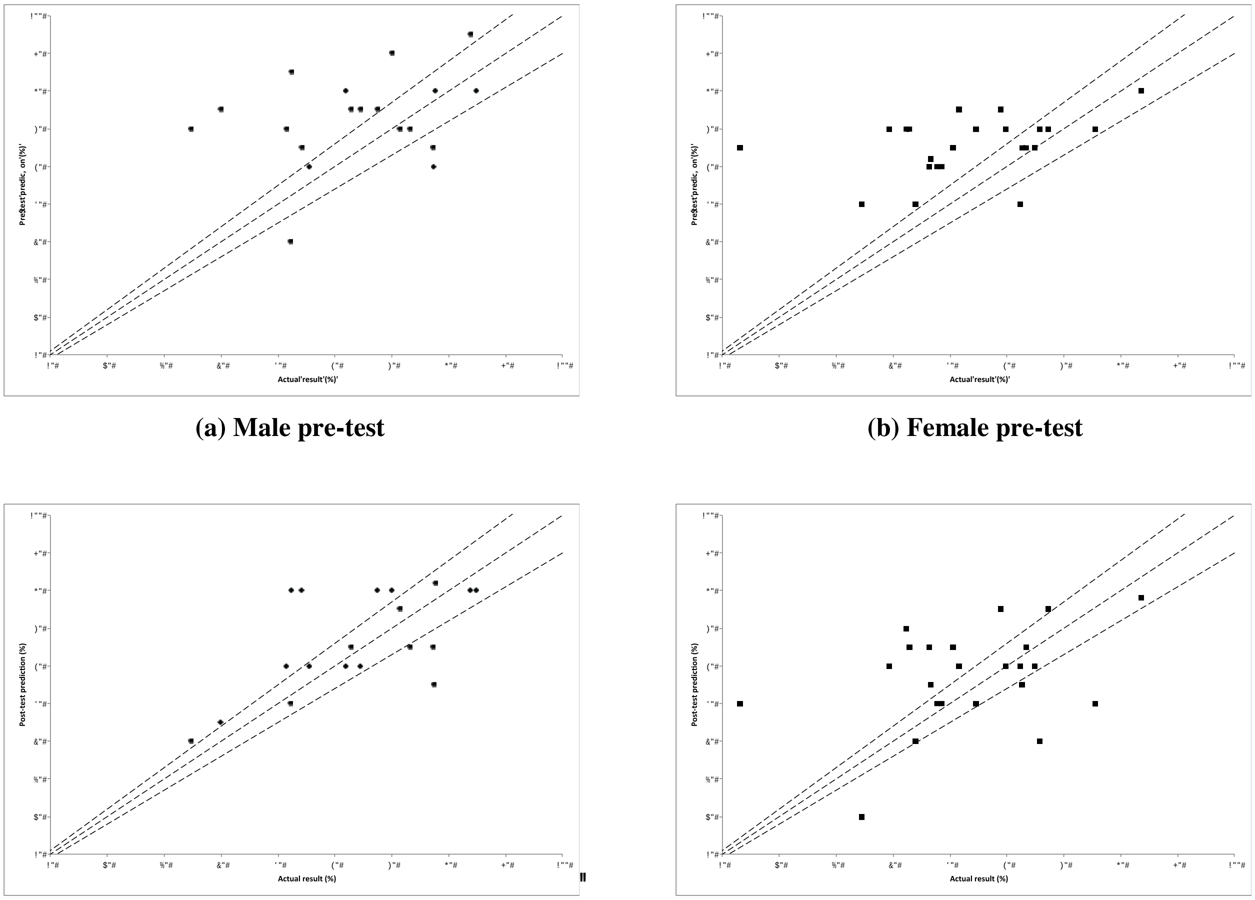


Figure 1: Averaged individual student performance predictions compared to actual test results for each written examination in the semester (errors bars represent the 95% confidence interval, numbers at the bottom of the bars represent the sample size).



(a) Male pre-test **(b) Female pre-test**

(c) Male post-test **(d) Female post-test**

Figure 2: Visual comparison of performance predictions vs actual results for the final exam - (a) pre-test predictions, males; (b) pre-test predictions, females; (c) post-test predictions, males; (d) post-test predictions, females. Dashed lines show slopes of 0.9, 1 and 1.1 respectively. (N = 42).

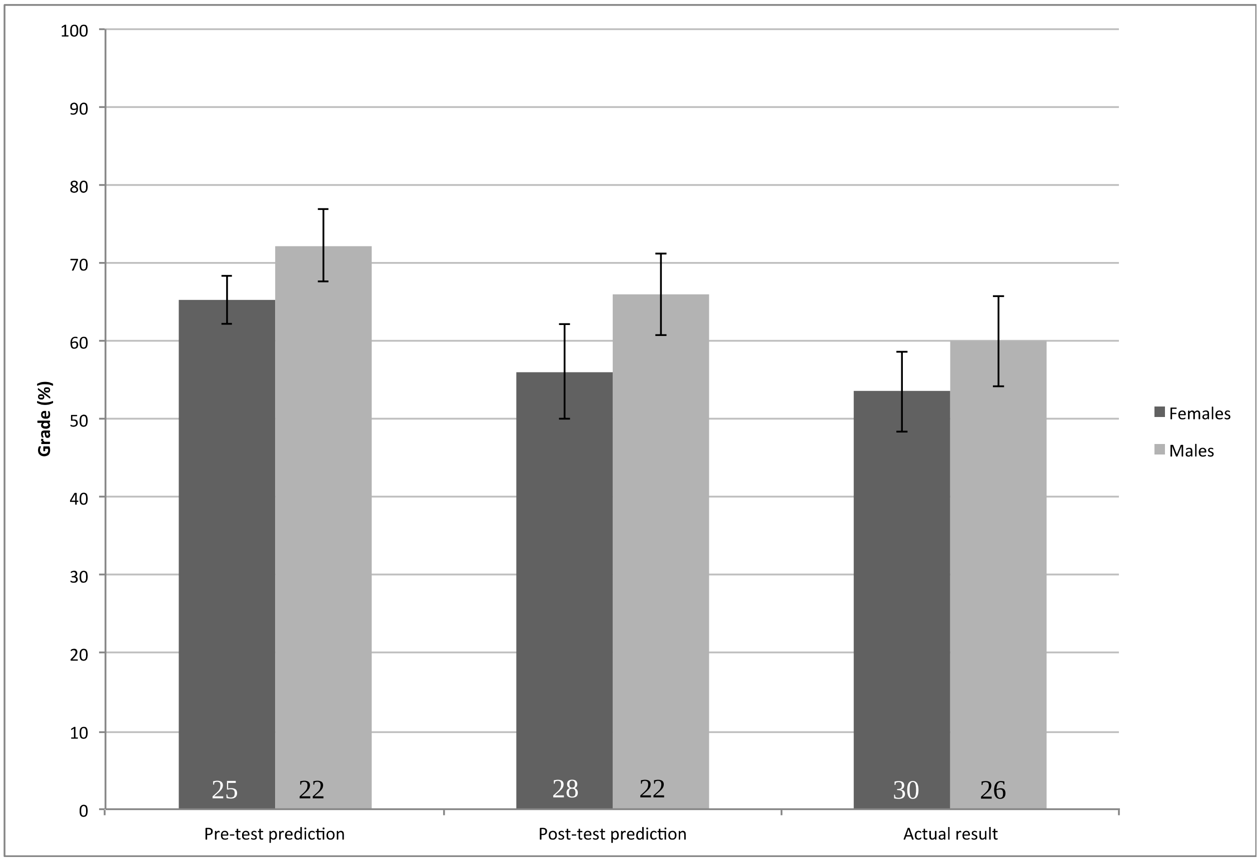


Figure 3: Average student performance predictions split by gender for the final exam (errors bars represent the 95% confidence interval, numbers at the bottom of the bars represent the sample size).

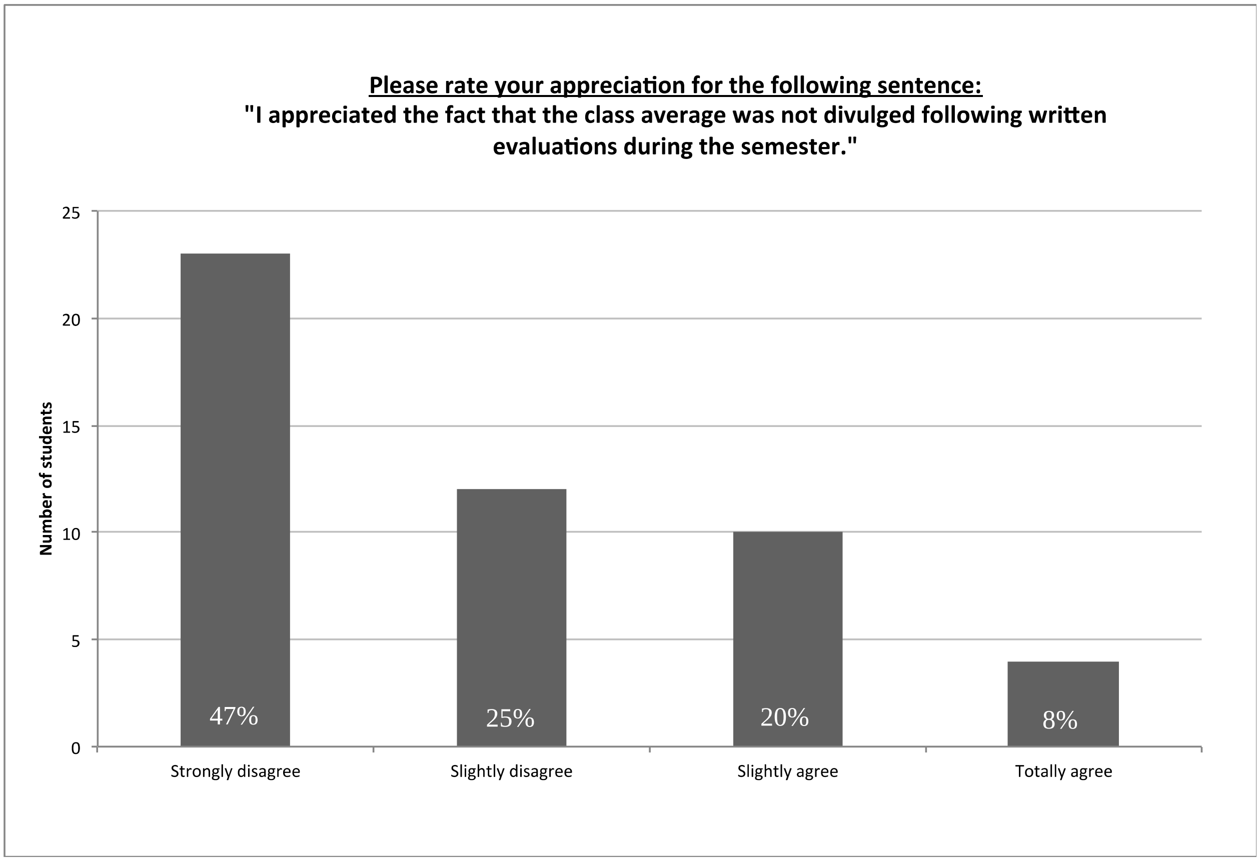


Figure 4: Students' appreciation concerning the non-disclosure of class averages over the course of the semester (N = 49).