Ross, J. A. & Starling, M. (2008). Self-assessment in a technology supported environment: The case of grade 9 geography. *Assessment in Education: Principles, Policies, and Practice*, 15(2), 183-199.

Self-Assessment in a Technology Supported Environment:

The Case of Grade 9 Geography

John A. Ross*

Michelle Starling

University of Toronto

*Corresponding Author

Dr. John A. Ross, Professor,
Curriculum, Teaching, & Learning,
Ontario Institute for Studies in Education,
University of Toronto,
Box 719, 1994 Fisher Dr.,
Peterborough, ON K9J 7A1
CANADA

jross@oise.utoronto.ca
tel: 705-742-9773, ext 2293
fax: 705-742-5104

Acknowledgements

An earlier version of this article was presented at the annual meeting of the American Educational Research Association, Montreal, 2005. Preparation of the manuscript was funded by the Social Sciences and Humanities Research Council. Views expressed in the article do not necessarily represent the views of the Council.

John A. Ross is Professor of Curriculum, Teaching, and Learning at the Ontario Institute for Studies of Education at the University of Toronto and head of the Institute's field center in Peterborough, Ontario. His research interests are student assessment, mathematics education, and program evaluation.

Michelle Starling was a graduate student in Curriculum, Teaching, and Learning at the Ontario Institute for Studies of Education at the University of Toronto at the time of the study. She is currently a secondary school teacher in Trillium Lakelands District School Board.

Abstract

We investigated the impact of self-assessment training on student achievement and on computer self-efficacy in a technology-supported learning environment (grade 9 students using Global Information Systems software). We found that self-assessment had a positive effect on student achievement, accounting for 25% of the variance across three measures. The treatment effect was as large for females as for males and for those with low initial self-efficacy as it was for those with higher scores. In addition, self-efficacy increased more in the control than in the treatment group. We interpreted the self-efficacy results to be a positive outcome of the treatment: teachers may have used self-assessment training to depress the inflated self-perceptions of some teenagers.

Self-Assessment in a Technology Supported Environment:

The Case of Grade 9 Geography

Most teachers include self-assessment in their repertoires (Noonan & Duncan, 2005), particularly if they are enacting assessment reform in their classrooms. Self-assessment embodies many of the key features of assessment reform (as defined by Aschbacher, 1991; Newman, 1997; Wiggins, 1993; 1998). For example, interpreting one's performance using explicit criteria meets the reform objective that assessments involve higher-level thinking and disciplined inquiry. Self-assessment that makes visible the procedures, criteria and standards of assessment meets the reform goal of transparency. Negotiating differences between student and teacher appraisals addresses the reform expectation that assessments provide students with opportunities for feedback and revision during the task. A central element of assessment reform is the expectation that assessments will contribute to improved student. However, evidence of the consequential validity of self-assessment is derived almost entirely from studies of classrooms in which technology is not in high use. In this article we will present the results of a study that investigated the consequential validity of self-assessment in a technology-supported environment (grade 9 Geography).

In this article we will first review research on self-assessment to demonstrate that self-assessment is a valid and reliable technique for assessing student performance, particularly in contexts in which self-assessments are used for formative rather than summative purposes; that self-assessment contributes to student achievement in regular classrooms; and that few studies of the effects of self-assessment in technology-rich classrooms are available and these produced mixed results. We will argue, from a theoretical model of the mechanisms that link self-

assessment to achievement, that researchers need to include student self-efficacy beliefs in examining the achievement effects of self-assessment in technology-supported learning environments. From this foundation we will report the results of a study of self-assessment involving Geography students in their first high school year.

Foundation for the Study

Self-assessment

In this article we will follow Klenowski's (1995) definition of self-assessment as "the evaluation or judgment of 'the worth' of one's performance and the identification of one's strengths and weaknesses with a view to improving one's learning outcomes" (p. 146). We use the term assessment rather than evaluation to emphasize the formative purpose of self-appraisal. We recommend that teachers assign self-assessment tasks for the purpose of improving the quality of student work; i.e., assessment for learning. Although self-assessment can be productively used for summative purposes, this is not our goal. In the province in which we conducted the research, self-assessments may not be used in calculating students' grades.¹

Previous studies have demonstrated that the psychometric properties of self-assessment are sufficient for classroom use. Self-assessments produce consistent results across tasks, across items, and over short time periods. For example, Fitzgerald, Gruppen and White (2000) found that the self-assessments of medical students were consistent over task formats (their own clinical examination of standard patients versus interpretation of test results administered by others). Reliability across items was demonstrated by Ross, Rolheiser and Hogaboam-Gray (1999-a) who reported internal consistencies on grade 4-6 self-assessments in narrative writing of $\alpha=.84$. Sung, Chang, Chiou, and Hou (2005) demonstrated consistency over short time periods in a study of 14-15 year olds assessing the quality of their web-designs over three time

periods. The evidence of the consistency of self-assessments over long periods of time is much weaker. Blatchford (1997) found little agreement in self-assessments from ages 7 to 11 with slightly greater agreement from ages 11 to 16.

The evidence in support of the validity of self-assessment is mainly based on studies of the agreement of student self-assessments with teacher appraisals of the same work. Correlations tend to be moderately positive with considerable variation among individual studies (see reviews by Boud & Falchikov, 1989; Steve Ross, 1998). Student self-assessments tend to be modestly higher than teacher judgments, particularly if the self-assessments contribute to students' grades (Boud & Falchikov, 1989). However, researchers report that discrepancies between teacher and student assessments can be reduced by increasing student understanding of the criteria for appraisal (Ross et al., 1999-a) and by informing learners that their self-assessments will be compared to peer or supervisor ratings (Fox & Dinur, 1988). Correlations between self-assessments and an external criterion (such as standardized test scores) are influenced by age (older students are more accurate) and by knowledge of the domain measured (see review by Ross, 2006).

These studies suggest that self-assessment provides a credible representation of student accomplishment, sufficient to provide students with valid and reliable information about their progress on school tasks. A key instructional question is whether students benefit from self-assessment.

Self-assessment and Achievement

In this section we review previous studies of the link between self-assessment and student achievement, first in regular classrooms and then in technology-rich environments. Asking students to assess their performance, without further training, contributes to greater intrinsic

motivation and stronger achievement in some studies (Schunk, 1996) but not in others (Fuchs, Fuchs, Karns, Hamlett, Kataroff, & Dutka, 1997). Others have examined the effects of programs that train students how to assess their work. Positive effects for self-assessment have been reported for a range of subjects (Andrade & Boulay, 2003; Arter, Spandel, Culham, & Pollard, 1994; Hillocks, 1986; McDonald & Boud, 2003; Ross et al., 1999-a; Ross, Hogaboam-Gray, & Rolheiser, 2002-b) and for non-academic outcomes such as help giving (Ross, 1995), self-direction (Henry, 1994) and self-control (Nelson, Smith, & Colvin, 1995). Only a few studies have found negative effects for self-assessment. For example, a study of grade 11 mathematics students found that self-assessment heightened the awareness of some students that they were unsuccessful on challenging mathematics tasks, leading to depressed confidence in their ability, and lower achievement (Ross, Hogaboam-Gray, & Rolheiser, 2002-a)

These self-assessment studies were conducted in classrooms in which technology had a minimal role. The effect of self-assessment in technology rich environments has rarely been investigated. Some computer-aided learning packages include a self-assessment component but the feedback students receive tends to be limited and researchers have not disentangled the effects of self-assessment from other treatment dimensions (e.g., Volland, Topping, & Evans (1999). We found only two studies that isolated the effects of self-assessment in classrooms using technology. In study 1, Schunk and Ertmer (1999) found that students in a learning goal condition who self-assessed on a single occasion in a computer studies course learned more than control group students. Self-assessment had no effect on students in a performance goal condition. Study 1 confirmed a similar finding about the interaction of self-assessment and goal orientation on achievement generated by Schunk (1996) in classrooms that were not technology rich. However, in Study 2 of Schunk and Ertmer (1999) the finding did not replicate. Students

who self-assessed on three occasions performed no better in a computer studies course than students who did not self-assess regardless of their goal orientation condition. In both these studies, Schunk and Ertmer found that self-assessment correlated with self-efficacy.

The shortage of studies on the effects of self-assessment in technology-rich environments and the discrepancy in the outcomes of the two studies relevant to the domain constitute an important gap in our understanding of the effects of self-assessment on achievement.

Consequently the first research question motivating our study was: Will self-assessment contribute to higher student achievement in technology-rich classroom environments?

Theory Linking Self-assessment to Achievement

In this section we describe a theoretical framework for linking self-assessment to achievement. Figure 1 from Ross et al. (2006) uses social cognition theory (Bandura, 1997) to make the link.

Figure 1 About Here

Self-assessment embodies three processes that self-regulating students use to observe and interpret their behavior (Schunk, 1996). First, students produce self-observations, deliberately focusing on specific aspects of their performance that correspond to their subjective standards of success. Second, students make self-judgments in which they determine how well their general and specific goals were met. Third, are self-reactions, interpretations of the degree of goal achievement that express how satisfied students are with the results of their actions.

Self-assessment contributes to self-efficacy, i.e., “beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (Bandura, 1997, p. 2). Self-efficacy is not an omnibus trait but a differentiated set of self-beliefs linked to distinct realms of human functioning; for example, self-efficacy for completing computer tasks is likely to be

weakly correlated with self-efficacy for composing a narrative. Students who perceive themselves to have been successful on the current task (i.e., who recognize it as a mastery experience) are more likely to believe that they will be successful in the future (Bandura, 1997).

Students with greater confidence in their ability to accomplish the target task are more likely to visualize success than failure. They set higher standards of performance for themselves. Student expectations about future performance also influence effort. Confident students persist. They are not depressed by failure but respond to setbacks with renewed effort. For example, students with high self-efficacy interpret a gap between aspiration and outcome as a stimulus while low self-efficacy students perceive such a gap as debilitating evidence that they are incapable of completing the task (Bandura, 1997). The combination of higher goals and increased effort contributes to higher achievement.

Positive self-assessments foster an upward cycle of learning, as demonstrated by the studies that found positive outcomes for self-assessment. But the processes in Figure 1 can generate negative outcomes, as found in Ross et al. (2002-a). In that case study of self-assessment in a grade 11 mathematics classroom, we found that student self-assessments persuaded students that they did not understand core mathematics ideas, even though they were working hard to learn them. The conclusion that many students drew was that they lacked the ability to do advanced level mathematics. Some responded with ego-protecting effort reduction that reduced achievement. A stream of negative self-assessments can lead students to select personal goals that are unrealistic, adopt learning strategies which are ineffective, exert low effort and make excuses for performance (Stipek, Recchia, & McClintic, 1992).

Computer Self-Efficacy

The theoretical model in Figure 1 highlights the importance of self-efficacy as a mechanism linking self-assessment to achievement. In addition, Schunk and Ertmer (1999) reported strong correlations of self-assessment and self-efficacy in their studies of self-assessment in computer studies. Low computer self-efficacy is frequently cited as a motivator of technology avoidance (e.g., Harrison & Rainer, 1992).

Students' past experiences are the strongest influence on self-efficacy beliefs. Those who are successful on computer tasks will be more likely to seek out similar experiences than those who are less successful. Success and experience reinforce each other, creating a strong correlation between computer experience and computer self-efficacy. Researchers have demonstrated that computer experience predicts computer self-efficacy (Coffin & McIntyre, 1999; Durndell & Haag, 2002; Ertmer, Evenbeck, Cennamo, & Lehman, 1994; Hasan, 2003). Negative prior experience depresses self-efficacy. For example, Salanova, Grau, Cifre, & Llorens (2000) found that increased experience and training both contributed to computer burnout (exhaustion and cynicism). However, the effects were moderated by self-efficacy. Those with higher computer self-efficacy were less likely to experience negative effects. These findings suggest that prior computer experience is an important criterion in determining the equivalence of treatment and control groups.

Other factors influence computer self-efficacy, either directly or indirectly through increased computer experience, as in the case of computer ownership (Levine & Donitsa-Schmidt, 1998) and internet access. Parental modeling of successful computer practice is likely to have a positive effect on the computer self-efficacy of their children, through vicarious experience and through parental persuasion—two factors that contribute to self-efficacy, although not as strongly as successful experience (Bandura, 1997).

Finally, there is strong evidence that gender influences computer self-efficacy directly and indirectly through experience. The size of gender differences in cognitions about computers and computer use is not large (e.g., Bannert & Arbinger, 1996 found that gender explained 5% of the variance in computer outcomes) and there is some evidence that gender differences are declining as computer use becomes institutionalized (Christensen, Knezek, & Overall, 2005; Schumacher & Morahan-Martin, 2001). Although not all studies found evidence of gender differences, virtually every study that did reported that males are advantaged. Males are more likely to own a computer (e.g., Durndell & Lightbody, 1993), use it more often (e.g., Robertson, Calder, Fung, Jones, O'Shea, & Lambrechts, 1996), have greater computer experience (Bannert & Arbinger, 1996), and report more support from parents for computer use (Reinen & Plomp, 1997). Males are more confident about their computer abilities than females (Durndell & Haag, 2002; Murphy, Coover, & Owen, 1988).

Prior research on computer self-efficacy influenced the direction of our study in two ways: First, Schunk and Ertmer (1999) found that student achievement in computer studies was strongly linked to self-efficacy in both their studies, even though the effects of self-assessment on student achievement were inconsistent. For this reason, we included self-efficacy as a possible outcome of self-assessment in our research question: Will self-assessment contribute to higher self-efficacy in technology rich classroom environments? Second, because self-efficacy was such a highly salient variable in computer-supported instruction, we used the antecedents of computer self-efficacy as criteria for assessing the equivalence of students in our study conditions and we included pretest self-efficacy scores as a covariate in our analysis.

Purpose and Context of Our Research

Our goal was to extend research on the effects of self-assessment training to computer-supported learning environments. The research questions of the study were:

(1) Will self-assessment contribute to higher student achievement in technology rich classroom environments?

(2) Will self-assessment contribute to higher self-efficacy in technology rich classroom environments?

The context was a grade 9 Geography course in which students solved real-life geographic problems using GIS (Global Information Systems) software: ArcView 3.2 (ESRI Canada, 2002). ArcView is an automated mapmaker containing a database and database manager that stores, retrieves, manipulates and displays geographic data. Evidence about the effects of ArcView training on geographic knowledge is positive, albeit anecdotal (Audet & Paris, 1997; Solem, 2001). GIS is used in all secondary school Geography courses in the province in which this study was conducted because it emphasizes spatial concepts that are the basis of geographic understanding (Oldakowski, 2001). Students do not encounter ArcView prior to grade 9 and it was not available for home use at the time of the study. Although associated most closely with Geography, many disciplines draw on GIS (Audet & Abegg, 1996), so that the study findings have application beyond Geography courses.

Method

Sample

We invited teachers in three secondary schools to participate in the study. Two teachers, each with two classes, from one school volunteered to serve as the treatment group and four teachers, each with one class, from the other two schools agreed to be the control, creating a quasi-experimental design (pre-post, nonequivalent groups with statistical adjustment). All

students (N=164) were from a single school district in central Ontario, Canada, representing 10% of the grade 9 students in the district. Students were almost exclusively white, less than 1% spoke a language other than English in their home, and 6% were receiving special education services².

Instruments

At the beginning of the study all students responded to survey items concerning gender (1=female, 2=male), home ownership of a computer (1=yes; 2=no), Internet access (1=yes; 2=no), hours of student computer use, and parent use of computers (1=yes; 2=no). These measures were used to establish the equivalence of the treatment and control groups.

At the beginning and end of the study, all students completed a computer self-efficacy scale in which they reported confidence in their ability to perform 25 computer application tasks (e.g., logging onto a networked computer, copying a disk). Students used a 1-6 scale anchored by very little confidence and a lot of confidence. Responses were averaged to produce a 1-6 score. The internal consistencies were high on pre- ($\alpha=.97$) and post-tests ($\alpha=.98$).

At the end of the study all students completed an individual performance assessment over two 75-minute periods that produced three scores. Students created a map using GIS and wrote a report of a problem solving task in which they described the importance of a Geographic problem (recycling in Canada), generated alternative solutions to the problem, identified advantages and disadvantages of each alternative, and provided a rationale for the solution they selected. *GIS Map* consisted of a 1-5 score based on computer application skills. Level one was given to maps with two or more substantive errors (e.g., map content errors). Level five was given to a map completed correctly with two or more meaningful additions to the basic procedure (e.g., inclusion of an inset map or use of colour/ shading to highlight an argument).

GIS Report consisted of a 1-5 score based on dimensions of problem solving (problem definition, data analysis, and recommendation), and structural conventions. The rubric for both measures is available from the authors. The third measure *GIS exam*, was a low inference test of GIS mapping skills in which each student created a map with specified features (e.g., ecozone boundaries), labeled the features, found relationships between features, and exported the map to hard copy. Students received a score from 0-50 on the GIS exam.

The GIS Maps and GIS Reports were coded by two judges who were trained to 80% agreement. Each judge independently coded each map and report and resolved differences through discussion. Inter-rater reliability before discussion, defined as agreement within one level of the scale, produced agreement of 94% for the Maps and 99% for the Reports. Cohen's (1960) Kappa, a chance-adjusted measure of agreement, for the less stringent definition was .88 for the Maps and .98 for the Reports. (Bakeman & Gottman, 1997 described Kappa scores of .75+ as excellent reliability.) The GIS exam was marked by classroom teachers using a low inference scoring sheet. Reliability could not be calculated because we had access only to the total scores.

Study Conditions

Students in both treatment and control groups worked on computer-supported units once per week for four months. The units were based on a public policy issues (e.g., developing a fire management policy for British Columbia forests). Students constructed maps, tables, and charts representing aspects of the problem (e.g., forested regions, areas of fire, causes of fire, effects of selected policies). Data were drawn from the ArcView database and from other sources. Students completed individual assignments but worked in pairs (which produces higher learning in computer tasks than working alone: Berg, Brandstorm, & Sutter, 1995; Hooper, Temiyakarn, &

Williams, 1993; Mevarech, Silber, & Fine, 1991). Students were paired on the basis of their pretest self-efficacy (highs with lows). Students in all conditions used the same set of student tasks, learning materials, hardware and software.

Students in the treatment, but not the controls, were trained in self-assessment procedures based on the four stages in Rolheiser (1996): (i) Involve students in defining assessment criteria; i.e., with teacher assistance, students constructed a rubric that expressed performance expectations and student criteria for the GIS project in language meaningful to students. The rubrics, unique to each class but sharing common features, were displayed in the classroom and were used by students and teachers to assess projects. (ii) Teach students how to apply the criteria; i.e., the teachers modeled application of the rubric by assessing examples of student work in whole class debriefings and when intervening with student pairs. (iii) Give students feedback on their self-assessments; i.e., teachers engaged students in rubric-based discussions of the differences between their self-assessments and assessments by the teacher. (iv) Help students use assessment data to develop action plans; i.e., students completed short instruments assessing project progress which they shared with their partners to set immediate goals. We gave teachers in the treatment condition detailed strategies for applying these stages in the context of the GIS project and supplied a manual of strategies and assessment forms (Rolheiser, 1996). Students in the control condition were evaluated solely by the teacher with no formal self-assessment.

Teachers in both treatment and control conditions attended three half-day in-services, two before the start of the semester and a third near its end. At sessions 1 and 2 all teachers experimented with GIS software, reviewed GIS exercises, and developed strategies for assisting students through the activities. The third session focused on solving problems. At each session, the treatment teachers met separately for one hour on strategies for teaching self-assessment.

During this hour control group teachers continued to focus on the classroom use of GIS software. At the end of the study we gave the control group teachers the self-assessment materials that the treatment group teachers received during the study.

Implementation of Study Conditions

There was no direct observation of teaching. At the end of the study each teacher was interviewed about implementation of the GIS unit and self-assessment training. All teachers in both conditions stated they completed the six sections of the GIS unit, including student production of reports in the final four sections. Both treatment teachers focused self-assessment training activities on stage 1 (involve students in developing criteria) and stage 2 (show students how to apply the criteria to their work). Each teacher reported following the procedures recommended in the in-service with a high degree of fidelity. However, they gave much less attention to stage 3 (give students feedback on their self-assessments) and stage 4 (help students set action plans based on their self-assessments). One of the treatment teachers systematically compared her assessments to the self-assessments of students using the same rubric; the other treatment teacher incorporated student self-assessment into his lesson planning but did not provide direct feedback to each student on the accuracy of their self-assessments. Both treatment teachers reported describing how to use self-assessment to set learning goals but neither had students overtly do so. The four control teachers reported they did not use the self-assessment training materials given to the treatment teachers nor did they ask students to self-assess during the unit.

Analysis Procedures

After determining the reliability of study variables and the pretest equivalence of the groups, data were entered into a multivariate analysis using GLM (General Linear Modeling) in

SPSS. The dependent variables were GIS map, GIS report, GIS exam, and computer self-efficacy. The independent variable was experimental condition, and the covariate was pretest computer self-efficacy.

Results

Table 1 displays the means, standard deviations, and t-tests for treatment and control groups on the pretest variables. The table shows there were no significant differences between the treatment and control groups on any of the pretest measures, suggesting that despite nonrandom assignment of students to conditions, the groups were equivalent.

Table 1 Here

Table 2 shows the means and standard deviations on the posttest variables for both groups. The multivariate analysis showed that self-assessment had a significant effect on the outcomes, after controlling for the effects of pretest self-efficacy. The multivariate model explained 84% of the variance in outcome scores. Both self-assessment [$F(4,137)=18.221$, $p=.001$] and pretest self-efficacy [$F(4,137)=9.469$, $p=.001$] had significant effects. The treatment accounted for 22% of the variance across outcomes, a large effect according to Cohen's (1988) criteria for interpreting effect sizes.

Table 2 Here

Examination of the univariate results revealed main effects for pretest self-efficacy on two of the outcome measures. Students with higher self-efficacy on entry to the program had higher self-efficacy on the posttest [$F(1,143)=73.835$, $p=.001$, $\eta^2=.345$]. In addition, students who had greater initial confidence in their ability to complete computer tasks had higher scores on the GIS Map [$F(1,143)=4.000$, $p=.047$, $\eta^2=.028$].

Research question 1: Will self-assessment contribute to higher student achievement in technology rich classroom environments? After controlling for the effects of pretest self-efficacy, students in the self-assessment group scored higher than students in the control group on all the achievement measures: GIS Map [$F(1,143)=4.452, p=.021, \eta^2=.045$], GIS Report [$F(1,143)=24.606, p=.001, \eta^2=.141$], and GIS exam [$F(1,143)=5.367, p=.006, \eta^2=.062$]. Self-assessment had a greater impact on each of the achievement outcomes than did pretest computer self-efficacy; (i.e., the partial η^2 for pretest self-efficacy on the achievement measures were .006, .006, and .028 respectively).

Research question 2: Will self-assessment contribute to higher self-efficacy in technology-rich classroom environments? Self-assessment had a significant effect on posttest computer self-efficacy, after controlling for pretest effects. Students in the self-assessment group scored lower than control students on posttest self-efficacy [$F(1,143)=5.068, p=.032, \eta^2=.038$].

To investigate this unexpected result, we constructed a new variable, self-efficacy change, by subtracting pretest self-efficacy from posttest self-efficacy. The mean for self-assessment students was lower than the mean for control students (.09 to .25), although the results on a t-test were not statistically significant [$F(154)=.214, p=.072$]. (T-test is a less powerful procedure than GLM.) Table 3 displays the frequencies. The table shows that the number of students who declined in computer self-efficacy was virtually identical in both groups. In contrast, more control than treatment students increased their self-efficacy scores, particularly in the extreme score range where there were ten control students with large self-efficacy increases compared to only one in the self-assessment group. Table 3 suggests that student confidence in their computer skills increased as they worked through the GIS program and that self-assessment had a moderating effect on self-efficacy increases.

In summary, the effects of self-assessment in this study ranged from small (computer self-efficacy, GIS map, and GIS exam) to medium (GIS Report). We also investigated interaction effects in a multivariate GLM involving the four dependent variables of the study. The analysis (not shown) indicated that the interaction of self-assessment with pretest self-efficacy and the interaction of self-assessment with gender were not statistically significant on any of the outcome measures. We also conducted a three-way (gender, pretest self-efficacy, study condition) MANOVA with posttest self-efficacy as the dependent variable. We did so by splitting pretest self-efficacy (a continuous variable) at the median: the results were not statistically significant [$F(4,130)=1.598, p=.179$]. These analyses demonstrated that the treatment was equally effective regardless of students' gender or computer self-efficacy on entry to the program.

Discussion

The study makes three contributions to academic knowledge. First, the study demonstrates that teaching students how to self-assess makes a positive contribution to their achievement in computer-supported learning environments. The effects were consistent across three domains of learning: spatial reasoning (represented by the computer-supported mapping score), solving authentic problems (represented by the GIS report), and basic geographic knowledge and skill (represented by the GIS exam). The only previous studies of self-assessment training in computer classrooms known to us (Schunk & Ertmer, 1999) found limited (Study 1) or no (Study 2) achievement differences. Self-assessment in that study may have had a weak effect because (i) it was done infrequently (once or three times), (ii) students in the other experimental conditions may have self-assessed even though it was not required, and (iii) students were given no explicit training in how to evaluate beyond the instruction to do so. Our

study suggests that if students are trained in self-assessment procedures and apply those procedures at least once per week over a four-month period, the effects on achievement will be as strong in technology-rich settings as have been reported for regular classrooms (Andrade & Boulay, 2003; Arter et al., 1994; Hillocks, 1986; McDonald & Boud, 2003; Ross et al., 1999-a; Ross et al., 2002-b).

The second contribution of our study is that it demonstrates that students' expectations about their ability to accomplish computer-supported tasks significantly influence their learning. Although the meta-analyses of Multon, Brown and Lent (1991) and Pajares (1996) demonstrate the influence of self-efficacy on learning in traditional settings, few studies have demonstrated the same link in computer-supported contexts. Our study provides evidence that student beliefs about their abilities in computer domains influence how much they learn, even after the effects of the treatment have been controlled.

The third contribution of our study is the finding that although both groups developed more positive expectations about their ability to perform computer tasks, the self-efficacy of control students increased to a greater degree than the self-efficacy of treatment students. In previous research we found that self-assessment training increased the accuracy of self-assessments; i.e., the gap between teacher judgments and student self-appraisals narrowed (Ross et al., 1999-a), a finding also reported by Sung et al. (2005). Increased accuracy can be viewed as a change in the underlying metric that treatment group students used when assessing their abilities. This recalibration or response-bias shift during a treatment was demonstrated by Howard (e.g., Howard & Dailey, 1979) and more recently by Cantrell (2003). Changes in self-assessment could occur through each of the processes displayed in Figure 1: joint construction of a rubric for appraising student work may influence self-observation by focusing student attention

on particular aspects of performance. The rubric might also influence self-judgments by redefining the standards students use to determine whether or not they were successful. Self-reactions might be influenced by teacher feedback to reinforce positive reactions when students accurately recognize successful performance.

The effects of self-assessment training on students might be magnified by teacher interventions. Lindsley, Brass, & Thomas (1995) argued that upward and downward spirals of self-efficacy beliefs are both bad, leading to over- or under-confidence. Each cycle depresses subsequent performance. Lindsley et al., writing in an organizational context, argued that it is the task of the leader to prevent upward and downward spirals. We believe that a similar process was operating in our study. Previous research has demonstrated that practice on computer tasks contributes to higher computer self-efficacy (Coffin & McIntyre, 1999; Durndell & Haag, 2002; Ertmer et al., 1994; Hasan, 2003). Our study confirmed this finding, showing that the self-efficacy increased from pre- to post-test (as shown in Tables 1 and 2). Table 3 showed that the effects of computer practice on computer self-efficacy were moderated by self-assessment. Students in the treatment group had more modest increases and the number of students with large increases was ten times higher in the control group. We speculate that teachers used the feedback stage of self-assessment training to lower self-efficacy beliefs of teenagers who held inflated views of their capacity to use computers. Although our design did not enable us to test this speculation directly, we revised Figure 2 to include the influence of self-assessment training and teacher interventions on the relationship between self-assessment and student learning.

Figure 2 Here

Limitations of the Study

The credibility of our findings is weakened by the use of a quasi-experimental rather than true experimental design: The teachers in this study volunteered for their roles. The most serious threat to internal validity in this design, pre-existing group differences, was mitigated by the demonstration that the treatment and control groups were equivalent on variables associated with outcomes in computer-supported learning environments in previous research (i.e., self-efficacy, gender, computer ownership, computer use, Internet access, and parental use of computers). We also used the most powerful of these factors, pretest self-efficacy, as a covariate in the analysis.

We did not directly observe implementation of study conditions. Although we interviewed the six teachers at the end of the study about their implementation of the GIS unit and fidelity to the self-assessment treatment, our implementation data are entirely self-report. It is also possible that students in the control group self-assessed, even though there was no formal requirement that they do so, and students may have discussed their self-assessments with their peers without the knowledge of their teachers or the researchers.

In addition, we can make only modest claims about the external validity of our study because our sample was not representative of a known population. However, our results confirm previously reported effects of self-assessment in classrooms in which instructional use of computers was not a focus (Arter et al., 1994; Hillocks, 1986; Klenowski, 1995; Nelson et al., 1995; Ross, 1995; Ross et al., 1999-a; 1999-b; 2002-b; Schunk, 1996; Sparks, 1991). The triangulation of results between technology aided and traditional settings suggest that our finding may be sufficiently robust to generalize to other computer supported learning environments.

Directions for Future Research

We suggest two directions. First, we recommend qualitative investigations of how teachers influence student self-assessments. What strategies do they use? Do they rely on

structures such as the procedures and forms in Rolheiser (1996)? Do teachers consciously or unconsciously uplift the under-confident and depress the self-inflated? Of special interest are strategies used when teachers intervene in small groups because teachers vary substantially in their intervention techniques used (Webb & Ing, 2006) and intervention messages may conflict with whole class debriefings (Ross, 1995). Qualitative studies could also address how students respond to teacher attempts to influence their self-assessments and whether student responses vary by identifiable student characteristics.

The second recommendation is for randomized control studies of the effects of self-assessment in technology-rich environments. Of particular interest is whether the self-efficacy results of the current study generalize. In addition, quantitative studies could examine whether self-assessment training results in students becoming more accurate in their self-appraisals (i.e., whether self-assessments converge with appraisals by independent assessors). The latter inquiry might build on the results of qualitative studies by examining the paths among self-assessment, teacher interventions, self-efficacy, and achievement. Quantitative research might also examine whether self-assessment training contributes to self-efficacy directly. For example, classroom discussions of exemplars of student work may provide examples of successful experience by students' peers that contribute to self-efficacy through vicarious experience. In addition, the willingness of teachers to share control of assessment constitutes an "inviting message"; i.e., information that the teacher perceives students to be able and responsible, an important source of positive efficacy information (Usher & Pajares, 2005).

Conclusion

The most important theoretical outcome of our study is that it provides additional evidence for the consequential validity of self-assessment and more generally strengthens the claims that assessment reform has important benefits for students.

The most important practical outcome of our study is the finding that self-assessment training contributes to student learning in computer-supported learning environments. The treatment accounted for 25% of the variance in achievement across measures, compared to 4% of the variance explained by pretest computer self-efficacy. In addition the treatment effects were as large for females as for males and for students with low expectations about their ability to perform computer tasks as they were for students with higher expectations.

Fuchs et al. (1997) found that simply providing students with the opportunity to self-evaluate without additional training had negative effects. Although not explicitly tested in this study, a treatment that provided only self-assessment opportunities could reinforce the belief of students with low expectations that they are technologically illiterate. To be an effective part of an instructional design the teacher needs to identify instructional goals, clarify appraisal criteria, give students feedback on the accuracy of their self-appraisals, and guide students in how to use self-assessment data to set learning priorities. We are optimistic that teachers who incorporate student self-assessment into their instructional design will be rewarded with higher student achievement.

References

- Andrade, H. G., & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research, 97*(1), 21-34.
- Arter, J., Spandel, V., Culham, R., & Pollard, J. *The impact of training students to be self-assessors of writing*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Aschbacher, P. R. (1991). Performance assessment: State, activity, interest, and concerns. *Applied Measurement in Education, 4*, 275-288.
- Audet, R., & Abegg, G. (1996). Geographic Information Systems: Implications for problem solving. *Journal of Research in Science Teaching, 33*, 21-45.
- Audet, R., & Paris, J. (1997). GIS implementation model for schools: Assessing the critical concerns. *Journal of Geography, 96*, 293-300.
- Bakeman, R., & Gottman, J. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge: Cambridge University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Bannert, M., & Arbinger, P. R. (1996). Gender-related differences in exposure to and use of computers: Results of a survey of secondary school students. *European Journal of Psychology of Education, 21*(3), 269-282.
- Berg, C., Brandstrom, J., & Sutter, J. *One computer per small group versus one computer per class: How two different formats affect the quality and quantity of student-student interactions while using a computer simulation*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher

- education: A critical analysis of findings. *Higher Education*, 18, 529-549.
- Cantrell, P. (2003). Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers. *School Science and Mathematics*, 103(4), 177-185.
- Christensen, R., Knezek, G., & Overall, T. (2005). Transition points for the gender gap in computer enjoyment. *Journal of Research on Technology in Education*, 38(1), 23-37.
- Coffin, R. J., & MacIntyre, P. D. (1999). Motivational influences on computer-related affective states. *Computers in Human Behavior*, 15(5), 549-569.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd edition. Hillsdale, NJ: Erlbaum.
- Durndell, A., & Haag, Z. (2002). Computer self efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample. *Computers in Human Behavior*, 18(5), 521-535.
- Durndell, A., & Lightbody, P. (1993). Gender and computing: Change over time? *Computers in Education*, 21(4), 331-336.
- Ertmer, P. A., Evenbeck, E., Cennamo, K. S., & Lehman, J. D. (1994). Enhancing self-efficacy for computer technologies through the use of positive classroom experiences. *Educational Technology, Research and Development*, 42(3), 45-62.
- ESRI (Environmental Systems Research Institute) Canada. (2002). ArcView (Version 3.2) Toronto: author.
- Fitzgerald, J. T., Gruppen, L. D., & White, C. B. (2000). The influence of task formats on the accuracy of medical students' self-assessments. *Academic Medicine*, 75(7), 737-741.
- Fox, S., & Dinur, Y. (1988). Validity of self-assessment: A field evaluation. *Personnel Psychology*, 41, 581-592.

- Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C., Katzaroff, M., & Dutka, S. (1997). Effects of task-focused goals on low-achieving students with and without learning disabilities. *American Educational Research Journal*, 34(3), 513-543.
- Harrison, A. W., & Rainer, R. K. (1992). An examination of the factor structures and concurrent validates for the computer attitude scale, the computer anxiety rating scale, and the computer self-efficacy scale. *Educational and Psychological Measurement*, 52, 735-744.
- Hasan, B. (2003). The influence of specific computer experiences on computer self-efficacy beliefs. *Computers in Human Behavior*, 19(4), 443-450.
- Henry, D. (1994). *Whole language students with low self-direction: A self-assessment tool*. Charlottesville: University of Virginia. (ERIC Document Reproduction Service No. ED 372359)
- Hillocks, G. (1986). *Research on written composition: New directions for teaching*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills.
- Hooper, S., Temiyakarn, C., & Williams, M. (1993). The effects of cooperative learning on high- and average-ability students. *Educational Technology, Research and Development*, 41(2), 5-18.
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination in self-report measures. *Journal of Applied Psychology*, 64, 144-150.
- Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education*, 2(2), 145-163.
- Levine, T., & Donitsa-Schmidt, S. (1998). Computer use, confidence, attitudes, and knowledge: A causal study. *Computers in Human Behavior*, 14(1), 125-146.
- Lindsley, D. H., Brass, D. J., & Thomas, J. B. (1995). Efficacy-performance spirals: A multilevel

- perspective. *The Academy of Management Review*, 20(3), 645-678.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10(2), 209-220.
- Mevarech, Z., Silber, O., & Fine, D. (1991). Learning with computers in small groups: Cognitive and affective outcomes. *Journal of Educational Computing Research*, 7(2), 233-247.
- Moss, P. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, 17(2), 6-12.
- Multon, K., Brown, S., & Lent, R. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counselling Psychology*, 38, 30-38.
- Murphy, C. A., Coover, D., & Owen, S. V. *Assessment of computer self-efficacy: Instrument development and validation*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.
- Nelson, J. R., Smith, D. J., & Colvin, G. (1995). The effects of a peer-mediated self-evaluation procedure on the recess behavior of students with behavioral problems. *Remedial and Special Education*, 16(2), 117-126.
- Newman, F. (1997). Authentic assessment in social studies: Standards and examples. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, adjustment, and achievement*. San Diego: Academic Press.
- Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools . *Practical Assessment, Research & Evaluation*, 10 (17).
- Oldakowski, R. K. (2001). Activities to develop a spatial perspective among students in introductory geography courses. *Journal of Geography*, 100, 243-250.

- Ontario Ministry of Education and Training. (2000) *Guide to the provincial report card, grades 1-8* [Web Page]. URL
<http://www.edu.gov.on.ca/eng/document/forms/report/1998/repgde.pdf>. [2006, February 22].
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Reinen, I. J., & Plomp, T. (1997). Information technology and gender equality: A contradiction in terms? *Computers in Education*, 28(2), 65-78.
- Robertson, S. I., Calder, J., Fung, P., Jones, A., O'Shea, T., & Lambrechts, G. (1996). Pupils, teachers and palmtop computers. *Journal of Computer Assisted Learning*, 12, 194-204.
- Rolheiser, C. (Ed.). (1996). *Self-evaluation...Helping kids get better at it: A teacher's resource book*. Toronto: OISE/UT.
- Ross, J. A. (1995). Effects of feedback on student behavior in cooperative learning groups in a grade 7 math class. *Elementary School Journal*, 96(2), 125-143.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, 11(10). Available online:
<http://pareonline.net/getvn.asp?v=11&n=10>
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002-a). Self-Evaluation in grade 11 mathematics: Effects on achievement and student beliefs about ability. In D. McDougall (Ed.), *OISE Papers on Mathematical Education*. Toronto: University of Toronto.
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002-b). Student self-evaluation in grade 5-6 mathematics: Effects on problem solving achievement. *Educational Assessment*, 8(1), 43-58.

- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999-a). Effects of self-evaluation training on narrative writing. *Assessing Writing*, 6(1), 107-132.
- Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999-b). Teaching students how to self-evaluate when learning cooperatively: The effects of collaborative action research on teacher practice. *Elementary School Journal*, 99(3), 255-274.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1-20.
- Salanova, M., Grau, R. M., Cifre, E., & Llorens, S. (2000). Computer training, frequency of usage and burnout: the moderating role of computer self-efficacy. *Computers in Education*, 16(6), 575-590.
- Schumacher, P., & Morahan-Martin, J. (2001). Gender, Internet and computer attitudes and experiences. *Computers in Education*, 17(1), 95-110.
- Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal*, 33(2), 359-382.
- Schunk, D. A., & Ertmer, P. (1999). Self-regulatory processes during computer skill acquisition: Goal and self-evaluative influences. *Journal of Educational Psychology*, 91(2), 251-260.
- Solem, M. N. (2001). A scoring guide for assessing issues-based geographic inquiry on the World Wide Web. *Journal of Geography*, 100, 87-94.
- Sparks, G. E. (1991). *The effect of self-evaluation on musical achievement, attentiveness and attitudes of elementary school instrument students*. Unpublished doctoral dissertation, Louisiana State University and Agricultural and Mechanical College.
- Stipek, D., Recchia, S., & McClintic, S. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development*, 57, 1-84.

- Sung, Y.-T., Chang, K.-E., Chiou, S.-K., & Hou, H.-T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers and Education*, 45(2), 187-202.
- Usher, E. L., & Pajares, F. (2005, April). *Inviting confidence in school: Sources and effects of academic and self-regulatory efficacy beliefs*. Paper presented at the meeting of the American Educational Research Association, Montreal.
- Vollands, S. R., Topping, K. J., & Evans, R. M. (1999). Computerized self-assessment of reading comprehension with the accelerated reader: Action research. *Reading and Writing Quarterly: Overcoming Learning Difficulties*, 15(3), 197-211.
- Webb, N., & Ing, M. (2006, April). *Student discourse and learning in elementary school mathematics classrooms*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Wiggins, G. (1993). *Assessing student performance: Explore the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.
- Wiggins, G. (1998). *Educative Assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass.

Table 1 Evidence of Pretest Group Equivalence: Means, SDs and Results of t-tests

Variable	Treatment			Control			t-tests
	N	Mean	SD	N	Mean	SD	
Pre Self-efficacy	68	5.03	.568	88	5.05	.619	$t(154)=-.670, p=.867$
Gender	68	1.57	.498	86	1.53	.502	$t(152)=.476, p=.813$
Own computer	68	1.03	.170	86	1.02	.152	$t(152)=.234, p=.815$
Internet access	66	1.05	.210	84	1.01	.109	$t(92.27)=1.179, p=.241$
Hours of use	66	2.36	1.002	84	2.29	.964	$t(148)=.483, p=.630$
Parents' use	65	1.15	.364	84	1.15	.364	$t(147)=-.015, p=.988$
Age	68	14.12	.406	86	14.09	.292	$t(117.51)=.421, p=.674$

Table 2 Means and Standard Deviations of Outcome Variables by Experimental Condition

Outcome	Treatment	Mean	SD	N
GIS Map	Treatment	3.42	1.39	64
	Control	3.03	1.24	79
	Total	3.20	1.21	143
GIS Report	Treatment	2.87	.94	64
	Control	2.18	.75	79
	Total	2.49	.90	143
GIS Exam	Treatment	36.47	9.49	64
	Control	32.97	8.95	79
	Total	34.53	9.33	143
Self-Efficacy 2	Treatment	5.11	.62	64
	Control	5.34	.57	79
	Total	5.24	.60	143

Table 3 Changes in Computer Self-Efficacy by Study Condition

Change in Self-Efficacy	Number of Treatment Students (<i>N</i> =68)	Number of Control Students (<i>N</i> =88)
Decreases		
-1.0 or less	3	3
-.50 to -.99	4	2
0 to -.49	19	19
Increases		
0 to .49	26	37
.50 to .99	15	19
1.0 or more	1	10

Figure 1: How Self-Assessment Contributes to Learning (From Author, 2006)

Figure 2: How Self-Assessment, Self-Assessment Training, and Teacher Interventions Contribute to Learning.

Endnotes

---

¹ “Self Assessment is an important skill by which students objectively examine the quality of their own work and/or their learning skills. It should **not** be used to inform their report card grade or mark” (Ontario Ministry of Education, 2000, p. 463, emphasis in the original).

² In Ontario only grouped data on student demographics are available.