Assessment of Students' Feedback Behavior in A Game-Based Automated Feedback System – A Cross-Cultural Replication Study

Annika SILVERVARG [a*], Kristen BLAIR[b], Maria CUTUMISU[c] and Agneta GULZ[d]
[a]*Dept of Computer and Information Science, Linköping University, Linköping, 58183, Sweden*
[b]*Graduate School of Education, Stanford University, Stanford, CA, USA*
[c]*Dept of Educational Psychology, University of Alberta, Edmonton, Canada*
[d]*Lund University Cognitive Science, Lund, 22100, Sweden*
*annika.silvervarg@liu.se

Abstract: In this paper, we argue for the importance of conducting replication studies over various schools and countries when addressing topics about learning and instruction and propose educational technology to be a tool for this endeavor. We present an example of a cross-cultural replication study that makes use of educational technology in the form of a digital game-based automated feedback system. The study addresses feedback related behavior in 11–15-year-old students in US and Swedish classrooms, investigating students' choices to seek confirmatory (i.e., positive) or critical (i.e., negative) feedback, as well as their subsequent choices to revise their work based on this feedback. Comparisons of the data collected at several schools in the US and Sweden showed similar patterns of relationships among students' feedback-seeking behavior, their tendency to revise their work, and their learning outcomes in and outside the assessment environment. Overall, the findings revealed that this assessment approach seems to be generalizable from a North American to a European population. However, the findings showed both a significant difference between Sweden and the US regarding the preference for critical feedback and between different schools within each country. Thus, it is possible that the difference between countries reflects school differences rather than cultural differences.

Keywords: feedback, self-regulated learning, assessment, educational technology, cross-cultural replication study

1. Introduction

Although research on instruction and learning using educational technology has gained momentum around the world, it is not clear how relevant, applicable, and transferrable the results are across countries. Thus, there is a need for research examining the cross-cultural validity of these findings. In psychology, researchers debated the applicability and generalizability of research conducted in WEIRD (Western, Educated, Industrialized, Rich and Democratic) populations/countries/cultures to other non-WEIRD populations/countries/cultures (Arnett, 2008; Henrich et al., 2010). A study reviewing papers published in artificial intelligence in education (AIED) and intelligent tutoring systems (ITS) conferences during 2002-2103 found that most of first authors and samples for empirical studies originated from the US (50% and 62%) followed by English-speaking countries (23% and 22%) and European countries (17% and 11%). However, non-WEIRD continents (Asia, Latin America, Africa) only accounted for 10% of the authors and 6% of the samples. Moreover, there are noted cultural differences between the American and other WEIRD societies, particularly with regards to individualism (Henrich et al., 2010).

Thus, there is a need for replication or semi-replication studies across countries to enable discoveries of patterns and contextual factors (e.g., culture) that influence study outcomes. However, such studies are rarely being conducted within educational science and AIED. One exception is Ogan et al.'s (2014) study that investigated help-seeking behavior across cultures in the US, Philippines, and Costa Rica. The researchers found that their models for effective help seeking based on data logging in a digital learning environment transferred to some degree between the US and Philippines, but not between these countries and Costa Rica. They discuss this finding in relation to cultural dimensions (Hofstede et al., 2010) such as Individualism and Power Distance, and how differences in these were found in the interaction between students and the teacher in the classroom. Differences in Power Distance have also been raised by teachers in Sweden regarding the (limited) applicability of research done in the UK to Swedish schools (Buljubasic, 2018). In Sweden, the governmental organization (Skolverket) recommends international research as a basis for teachers' work with formative evaluation and feedback, such as work by Dylan Wiliam (cf. Black & Wiliam, 2005) and Carol Dweck (cf. Dweck 2012). However, in a blog post, the Swedish teacher Johan Kant (Kant, 2021) discusses how the grading system and relation between the teacher and the student differs between countries and how that affects the way the results from studies on feedback and assessment conducted in the UK can or cannot be applied in Swedish schools.

In this paper we present a cross-cultural study on feedback behavior in an educational game used by 11–15-year-old students in US and Swedish classrooms, which is a replication of previous studies conducted in the US. We investigate students' choices of positive (i.e., confirmatory) versus negative (i.e., critical) feedback, as well as their inclination to revise the task on which they have received feedback. Then, we examine the correlations between (1) these behaviors and learning outcomes within the educational game used in the study and (2) these behaviors and an independent measure of student success (e.g., standardized test results). All studies were based on the curriculum-independent tool, Posterlet (Cutumisu, Blair, Chin, & Schwartz, 2015), a digital assessment game in which students design posters and learn graphical design principles. We pose the following research questions:

- Do learning choices (seeking critical feedback and revising posters) vary with country and/or school?
- Do the relations between learning choices and learning outcomes within the learning environment (i.e., Posterlet) vary with country?
- Do the relations between learning choices and broader learning outcomes outside the learning environment (i.e., academic achievement, such as grades or standardized tests) vary with country?

2. Background

In this section we provide background on the value of critical feedback, and how feedback-related behavior can differ between countries and cultures. We also describe the educational game used as the research instrument in our studies.

2.1 Critical Feedback

Feedback is crucial for effective learning as evidenced by a large body of scientific literature on feedback within the learning sciences (Fyfe & Rittle-Johnson, 2015). Despite this, little is known about how learners in different contexts pay attention to and process critical informative feedback. Evidence suggests that critical informative feedback is a form of feedback that can be especially beneficial for learning (Kluger & DeNisi, 1998). One explanation for why critical or negative feedback can be more effective than positive feedback for continued performance is that positive feedback indicates that one has done enough, whereas negative feedback indicates the need for a change (Cutumisu et. al, 2015). On the other hand, negative feedback runs the risk of triggering ego threat issues that lead people to ignore or neglect the feedback (Hattie & Timperley, 2007, Cutumisu et al., 2015; Tärning et al., 2020).

2.2 Culture and feedback related behavior

To our knowledge, no studies have directly addressed cultural differences in feedback behavior in educational contexts regarding either of: (i) students' inclination to seek critical feedback, (ii) their preferences when given the choice between critical or confirmative feedback, and (iii) their inclination to revise a task on which they have received critical feedback. However, Suzuki et al. (2008) conducted a study with multi-ethnic youths, investigating their responses to praise and criticism. The participating 36 teenagers (16-year-olds on average) were girls in 4 sports teams at US schools. All sports teams had members from three or four ethnic groups: African Americans, Asian Americans, European Americans, or Latinas. The study focused on whether there would be cultural differences in the participants' reactions to (i) being praised (for doing a job well) and (ii) being criticized (for making a mistake). The hypotheses that potential differences in responses to critical versus confirmative feedback would correspond to the degree of collectivism versus individualism in the different cultural backgrounds of the participants were confirmed by the data. Participants from more interdependent (collectivistic) cultures (Asian Americans and Latinos) were more inclined to accept negative feedback or criticism and more inclined to use it to self-correct. A reason behind this may be that members in these cultures often are socialized to receive negative feedback and use it to self-correct in order to achieve normative behavior (Greenfield et al., 2000; Kitayama et al., 1995). On the other hand, participants from the more individualistic cultures (European and African Americans) were more inclined to deny responsibility for a mistake, and more likely to assert their own point of view despite having someone else bring criticism or critical feedback.

On a broader level feedback behavior is an important aspect of self-regulated learning (SRL). Purdie et al. (1996) found both differences and similarities in beliefs about learning and use of SRL strategies for Japanese and Australian students. Both groups used the "environmental structuring" and "self-evaluating" strategies, where the latter typically includes checking, revising, and redoing one's work, using self-questioning through quizzes, or being asked questions by other people. However, the strategy least used was "reviewing tests and other work", which is closely related to getting and responding to feedback. The authors speculate that this may be due to the students not getting informative feedback on test or work. As stated earlier, positive feedback does not necessary motivate students, and negative feedback that comes without an indication of what and how work should be improved may not help students learn either.

McInerney (2008) also explores how motivation and SRL can be related to cultural differences and cultural identity and concludes that even though the construct of self-regulation appears to be universal, how it is actually realized in different cultures seems to vary (e.g., with regards to help-seeking, fear of failure, and parents' expectations). Thus, given that the way students seek and use feedback relates to several SRL constructs, it is reasonable to assume that these behaviors or choices are influenced by culture.

2.3 The research instrument – The educational game Posterlet

Posterlet is a game-based assessment that enables students to design three posters for their fictitious school's Fun Fair (Cutumisu, Blair, Chin, & Schwartz, 2015). For each poster, the environment offers students the choice of either confirmatory (e.g., "It's good you told them what day the fair is.") or critical (e.g., "People need to be able to read it. Some of your words are too small.") feedback from three animal characters (Step 4, Fig. 1) to help them learn about graphical design principles. It also measures whether students choose to revise their work after feedback (Step5-6). Posterlet is designed so that positive ("It's nice that the poster says how much the booth costs") and negative ("You didn't say how much the booth was") feedback provide equivalent informational value. The game assesses 21 graphical design principles (Cutumisu et al., 2015).

Posterlet has many similarities with other digital educational resources that collect and process data. However, there are some important differences. One is that the topic addressed by the game is curriculum independent. It is not expected that students will have encountered the specific graphical design principles previously, but rather that they will learn them through feedback while interacting with the game. The focus of measurement or assessment is feedback-related behavior, behavior that can occur and be relevant in all subjects (as well as in contexts outside of school). It is designed to be a stand-alone tool and, given the many demands on classroom time, a complete session can be as short as 10-15 minutes. A main advantage of conducting a cross-cultural replication study based on Posterlet is that it is a curriculum independent, stand-alone tool, since curricula - in all subjects - tend to vary strongly between different countries.

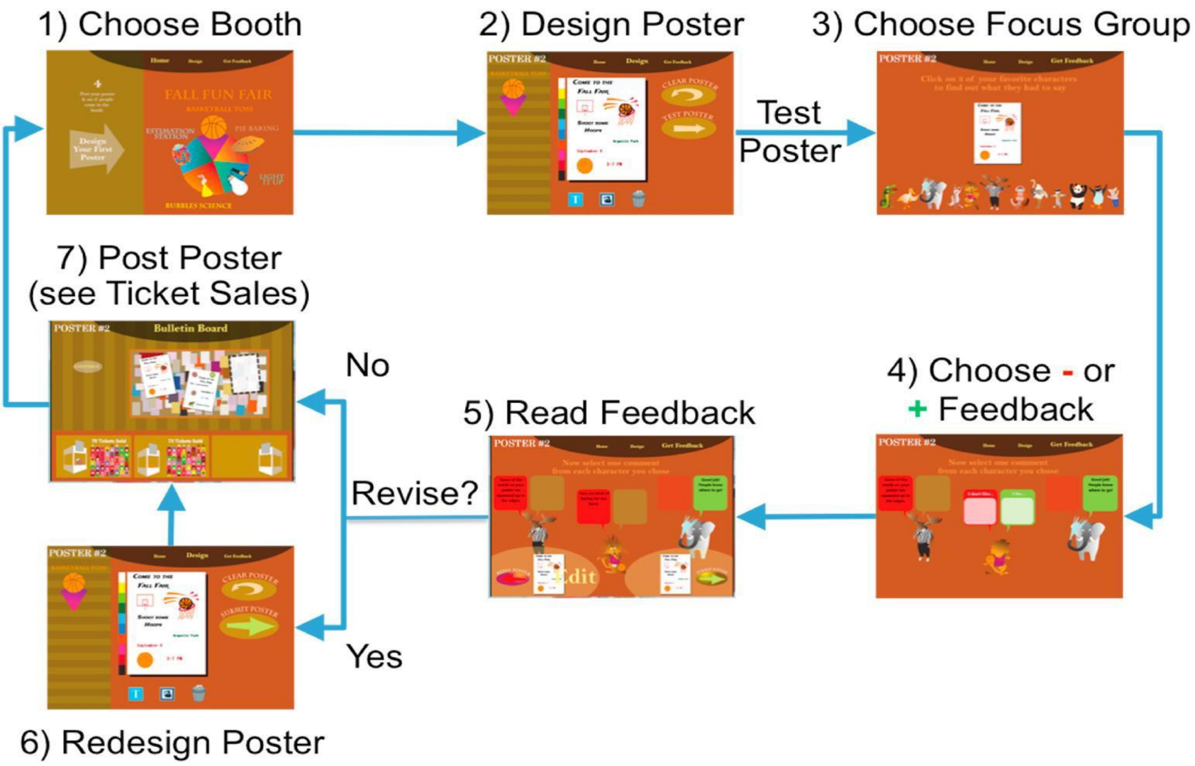

Figure 1. The steps involved in playing the Posterlet game. (Reprinted from Cutumisu et al., 2015 under Creative Commons License CC BY-NC-ND 3.0.).

2.4 Previous studies

Multiple studies using Posterlet in the US (Chin et al., 2019; Cutumisu et al., 2015, 2017; Cutumisu & Schwartz, 2021) have found consistent results across various data sets regarding the (i) relations between feedback and revising behavior (choice of negative feedback and choice to revise) and learning outcome measured in the application and knowledge of the design principles, and (ii) relations between feedback choices in the game and broader learning outcomes measured via standardized tests. These recurring patterns are that the more negative feedback students chose, the better they performed on the post-test and on the overall poster quality, even though both positive and negative feedback were equally informative. There is also a significant positive correlation between the choice to revise and post-test scores, which do not depend on the quality of the posters that students produced (Cutumisu et al., 2015, Chin et al., 2019). Secondly, the tendency to seek critical feedback also correlated with learning out-comes outside the game, such as standardized test in Mathematics and ELA (Cutumisu et al., 2015), and grades in science and math (Chin et al., 2019).

3. Method

As described above multiple studies using Posterlet in the US between 2013 and 2021 have found consistent results across various data sets. Moreover, the studies have included a variety of populations with both high and low socio-economic status, from schools with different profiles and from different geographical areas. There have also been comparisons between different ages of the students (Cutumisu & Schwartz, 2021). Thus, for our replication study we aimed for the same type of variation in recruitment of participating schools, except for age, where we limited our comparison to students in Grades 6-9 (i.e., 11–15-year-old). We strived to keep the instrument and procedure as similar as possible, described in more details below.

3.1 Participants and procedure

This study samples $N = 291$ Swedish students from five public middle/high schools in three different cities, with variation in socio-economic status. The data was analyzed and compared to data from three public middle schools from the states of California, Illinois, and New York ($N = 764$) previously reported (Chin et al., 2019; Cutumisu et al., 2015, 2017). Table 1 shows the school and participant information for the students included in the study.

Data collection took place in the participants' classrooms. Researchers provided instructions and students played the Posterlet game, followed by an online posttest on graphical design principles. Data from the game and the posttest were collected automatically online. The time spent on playing Posterlet was comparable over all studies. Swedish students played Posterlet for an average of $M = 14.8$ minutes ($SD = 4.1$), whereas USSchool2 students played Posterlet for an average of $M = 12.2$ minutes ($SD = 5.9$), USSchool1 students played Posterlet for an average of $M = 14.9$ minutes ($SD = 6.2$), and USSchool3 students played Posterlet for an average of $M = 14.9$ minutes ($SD = 4.07$).

Not all students completed the Posttest and learning outcomes such as grades and standardized test were not collected for all schools in the US (Table 1). To maximize the available data for each analysis, we used the subset of students who had complete data for each specific research question.

Table1. *Participant information*

School/ Country	**SES**	***N***	**Grade**	**Age Mean (Interval)**	**Posttest *N***	**Stand Test/ Grade *N***
SweSchool1	Low	39	7	13.23 (13-14)	39	39
SweSchool2	High	45	7	13.22 (13-14)	26	45
SweSchool3	Mid	34	8	15.18 (15-16)	22	34
SweSchool4	Mid	116	7	13.47 (13-15)	71	116
SweSchool5	Mid	52	7	13.9 (13-14)	52	52
Swe All		**286**	**7-8**	**13.68 (13-14)**	**200**	**286**
USSchool1	Low	172	6-8	12.2 (11-14)	163	65
USSchool2	Mid	272	6-9	12.1 (11-15)	226	116
USSchool3	High	89	7-8	(13-14)	-	75/57
US All		**533**	**6-9**	**11-15**	**389**	**256/57**

3.2 Measures

Three types of measures were employed in the study: (1) learning choices within the education game: choosing critical feedback and choosing to revise; (2) learning outcomes within the game: poster quality and posttest of graphical design principles; and (3) academic achievement/learning outcomes outside the digital learning environment (grades or standardized test results).

3

3.2.1 Learning choices

4

Critical Feedback (CF) measures the number of times students choose critical feedback (Step 4 of Fig. 1), and ranges from 0 to 9 (three posters and three feedback choices per poster). Revision is the number of times students chose to revise their posters, and ranges from 0 to 3 (one opportunity per poster).

3.2.2 Learning outcomes

5

Poster Quality is an in-game performance measure that scores how well the student used the 21 graphical design principles embedded in the game (-1 was assigned for an incorrectly applied principle; 1 was assigned for a correctly applied principle). Thus, a score ranging from -21 to 21 is possible for each poster, and the total measure for three posters can range from -63 to 63. Posttest measures the in-game learning of the design principles through four questions. The first is an open-response questions where students are asked to list mistakes beginners who design a poster might make. For the other three questions students are provided a model poster that they are asked to write feedback on, either in free form (question 2) or by ticking of good or bad things about the poster relating to the design principles (question 3 and 4). The two first questions have maximum score of 21 (since there are 21 design principles) and on the last two one can score at most 5 since there are five correct answers. The total Posttest score is computed by summarizing the normalized scores (Z-scores) for each of the four questions.

1.1.3 Academic achievement

For a broader measure of learning outcomes outside the digital learning environment, data on students results on state standardized tests for Mathematics and English Language Art (for all US schools) or Math grades (for all Swedish schools and one school in US) and Swedish Language grades (for Sweden) were gathered when available.

4. Results

4.1 Do learning choices (seeking critical feedback and revising posters) vary with country and/or school?

6

We investigated whether students seek critical feedback and choose to revise to a similar or different extent. Table 2 shows the mean for the measures Critical Feedback (CF) and Revision for Sweden and for US overall, as well as for the different schools in each country.

Table 2. *Mean and standard deviation M(SD) for learning choices by Country and School*

School/Country	Critical Feedback	Revision
SweSchool1	3.6 (2.4)	0.9 (1.2)
SweSchool2	4.2 (2.2)	1.1 (1.1)
SweSchool3	4.4 (2.1)	1.4 (1.0)
SweSchool4	4.8 (2.0)	1.3 (1.2)
SweSchool5	5.1 (1.9)	1.5 (1.1)
Swe All	**4.6 (2.1)**	**1.3 (1.2)**
USSchool1	3.2 (2.1)	1.2 (1.1)
USSchool2	4.0 (2.5)	1.1 (1.1)
USSchool3	5.8 (2.1)	2.0 (1.0)
US All	**4.0 (2.5)**	**1.3 (1.1)**

A t-test comparing Sweden and US revealed a significant difference ($t(817) = 3.4566$, $p < .001$) in CF, with a small effect size (Cohen's $d = .2599$). No significant difference was found for Revision.

As previously reported (Cutumisu et al., 2015), a t-test for USSchool1 and USS-chool2 showed a significant difference ($t(402) = -3.2$, $p < .01$) for CF, but not a significant difference for Revision.

A one-way ANOVA revealed that there was a statistically significant difference in CF between at least two Swedish schools ($F(4, 285) = 3.515$, $p = .008$). Bonferroni post-hoc tests showed that the mean value of CF was significantly different ($p = .011$) between SweSchool1 ($M = 3.6$) and SweSchool5 ($M = 5.1$); it was also significantly different ($p = .025$) between SweSchool1 ($M = 3.6$) and SweSchool4 ($M = 4.8$); but it showed no significant differences between the other schools. A one-way ANOVA revealed that there was not a statistically significant difference in Revision between the Swedish schools ($F(4, 285) = 1.952$, $p = .102$).

Thus, we found differences between US and Sweden in students' critical feedback-seeking behavior. However, it is worth noting that the US range of critical feedback-seeking (3.2 to 5.8) is greater than the range of Sweden (3.6 to 5.1). Thus, even though there are significant differences between the mean of Sweden and the US, this may relate to relative sample sizes across the different schools rather than culture, which is supported by the fact that there are significant differences between the schools within the countries.

No differences were found regarding students' choice to revise after they received feedback between the US and Sweden, nor were there differences between the Swedish schools or the USSchool1 and USSchool2.

4.2 Do the relations between learning choices and learning outcomes within the learning environment (i.e., Posterlet) vary with country?

7

To answer this question, correlations for the Swedish data were similar to those reported in the US studies (Table 3 and Table 4), such as the correlations between critical feedback-seeking (CF), Revision, Poster Quality and Posttest. The correlation coefficients for the Swedish study are presented in Table 5.

Table 3. *Correlations for USschool1 and USSchool2 study*

	N	**Revision**	**Poster Quality**	**Posttest**
CF	473	.47**	.28**	.23**
Revision	473		.34**	.24**
Poster Quality	473			.39**

Table 4. *Correlations for USSchool3 study*

	N	**Revision**	**Poster Quality**
CF	89	.40**	.37**
Revision	89		.26**

Table 5. *Correlations for Swedish study*

	N	**Revision**	**Poster Quality**	**Posttest**
CF	286	.48**	.22**	.19**
Revision	286		.21**	.23**
Poster Quality	286			.41**

To compare the correlations between the Swedish and US studies, a Fisher's *Z* test was performed. Table 6 includes *z*-scores and *p*-values for the Swedish and the US studies in USSchool2and USSchool1. It shows that the only significant difference in the correlations is between Revision and Poster Quality, which is higher for the US study. Overall, the relations between learning choices and learning outcomes were similar, regardless of culture.

Table 6. *Comparison of correlations between Swedish and US studies for USSchool1+2*

	CF-Rev	**CF-Poster quality**	**CF-Posttest**	**Rev-Poster Quality**	**Rev-Posttest**	**Poster Quality-Posttest**
Sweden	.48	.22	.19	.21*	.23	.41
US (School1+2)	.47	.28	.23	.34*	.24	.39
z-score, p	.21, .42	-.81, .21	-.05, .30	-1.82, .04	-.16, .44	.22, .41

The strongest correlation was found between CF and Revision. Thus, multiple regressions were performed to determine if CF and Revision were independent predictors of learning from the game. One regression used the Posttest as the dependent measure and one used Poster Quality.

As reported previously (Cutumisu et al., 2015), for USSchool1+2 both CF ($t(470) = 3.2$, $p = .002$, $\beta = .15$) and Revision ($t(470) = 5.4$, $p < .001$, $\beta = .25$) were predictors ($F(2,470) = 36.07$, $p < .001$, $R^2 = .13$) of Poster Quality score. Both CF ($t(411) = 2.7$, $p = .006$, $\beta = .15$) and Revision ($t(411) = 2.8$, $p = .005$, $\beta = .16$) were significant predictors ($F(2,411) = 16.11$, $p < .001$, $R^2 = .07$) of Posttest score. For USSchool3, CF ($t(88) = 3.40$, $p =.001$, $\beta = .37$) was a significant predictor of Poster Quality ($F(2,86) = 9.13$, $p < .001$, $R^2 = .17$; Chin et al., 2019).

The same regressions were conducted on the Swedish data, which showed that both CF ($t(283) = 2.39$, $p =.018$, $\beta = .16$) and Revision ($t(283) = 2.11$, $p = .036$, $\beta = .14$) were significant predictors ($F(2,283) = 9.74$, $p < .0001$, $R^2 = .064$) of Poster Quality. Revision ($t(197) = 2.01$, $p = .02$, $\beta = .18$) was a significant predictor for Posttest score ($F(2,197) = 6.27$, $p = .002$, $R^2 = .05$)

Thus, CF was a predictor of Poster Quality for all studies in both countries, and Revision predicted both Poster Quality and Posttest scores for both Sweden and the US study in USSschool1 and USSchool2.

4.3 Do the relations between learning choices and broader learning outcomes outside the learning environment (i.e., academic achievement, such as grades or standardized tests) vary with country?

We calculated the correlations between CF and Revision with Math and Swedish language grades for the Swedish students and compared these with similar correlations for the US studies. For USSchool1 and USSchool2 results for Standardized English Language Arts and Mathematics achievement tests were used, while both standardized tests (Math-CST and ELA-CST) as well as math grade were used in the USSchool3 study.

Table 7. *Correlation between learning choices and learning outcomes in Math and English/Swedish (grades or standardized test scores)*

	Math	**CF**	**Revision**	**Language**	**CF**	**Revision**
Sweden	Math grade	0.20**	0.24**	Swedish grade	0.16**	0.20**
USSchool1	ISAT	0.33**	0.21	ISTA	0.41**	0,31*
USSchool2	NYSTP Math	0.39**	0.28**	NYSTP ELA	0.33**	0.08
USSchool3	Math grade	0.21	0.30*	ELA-CST	0.23	0.18
USSchool3	Math-CST	0.29*	0.19			

As shown in Table 7, there were significant positive correlations in the Swedish data set between CF as well as Revision and grades in Math as well as Language. These are consistent with the previous results from the US studies, with several significant correlations between learning behaviors in the educational game and outcomes outside the educational game, which are positive and weak to moderate.

5. Discussion, Limitations, and Conclusions

We set out to conduct a cross-cultural replication study by replicating US studies in Sweden using the same method and instrument, hoping to make two different contributions: on the topic of feedback behavior and on the topic of conducting replication studies in another country.

Regarding feedback behavior, we found that a correlation between seeking out critical feedback and revising ones’ work after feedback is present in both countries, with negative feedback but not positive feedback used as an incentive to improve. This confirms previous results showing that praise or positive feedback is not used to revise nor necessarily leads to improved learning.

We also found significant correlations between seeking critical feedback and learning outcomes for all studies, which confirms previous theories that negative feedback overall is more fruitful than positive feedback (Kluger & DeNisi, 1998). This is especially interesting in this study because positive and negative feedback in Posterlet is designed to be equally informative.

Finally, even though the correlation between the amount of critical feedback the students chose, and their academic achievement was modest in our study, seeking critical feedback seems to be a productive behavior in academic settings in both countries.

The question remains whether this pattern of results is generalizable to other countries and cultures that are more different, for example more collectivistic cultures. Even though there are similarities between the US and Sweden, they are both part of WEIRD societies. Thus, it would be useful to expand the replication study to a greater diversity of countries, particularly those that fall outside the WEIRD categorization.

Regarding cross-cultural replication studies, one lesson learned is that replicating studies in different countries can inform us of what results are generalizable, but that it is hard to draw conclusions about differences. For example, we found a significant difference between Sweden and the US regarding the tendency to choose critical feedback, but we also found significant differences between different schools within each country, indicating that the difference between countries in our case likely reflects the sampling for different schools rather than country or culture. Furthermore, socio-economic status of schools is likely a relevant factor to consider for this kind of comparisons (Ewijk & Sleegers, 2010). We did not include this in the analysis, but there is a trend that students overall seek less critical feedback and revise to a lower degree in Schools with low socio-economic status (Table 2).

Also, school profiles and classroom practices – sometimes referred to as 'classroom cultures' – likely play a role (Roll & Wylie, 2016) and need to be taken into account. This is especially important to consider when using a digital tool for data collection that is considered "neutral" but can be integrated in the classroom in different ways. Teachers can be anywhere from absent from the classroom to collaborating and helping students. Ogan et al. (2014) noted that even if students were instructed to work silently and alone by their computer, in some classrooms, students may choose to collaborate informally, which influences how much help they seek in the digital learning environment. Such contextual differences are not captured by the data logs. Thus, if using educational technology for replication studies, the conditions in the classroom must either be carefully replicated or considered during analysis. As Roll and Wylie (2016) point out, it is important that such information also is included in articles.

Overall, when conducting replication studies and comparisons, using an instrument such as Posterlet is a strength in that it is curriculum independent and digitally delivered. Thus, some potential confounders in how a data collection of this kind would be contextualized and performed in different countries, are reduced.

References

Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63(7)

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5-31.

Blanchard, E. G. (2012, June). On the WEIRD nature of ITS/AIED conferences. *International Conference on Intelligent Tutoring Systems* (pp. 280-285). Springer, Berlin, Heidelberg.

Buljubasic, A. (2018). Formativ bedömning i praktiken – En kvalitativ intervjustudie om lärares användning av formativ bedömning. (eng. Formative assessment in practice – A qualitative interview study about teachers) Master Thesis. Göteborgs Universitet.

Chin, D. B., Blair, K. P., Wolf, R. C., Conlin, L. D., Cutumisu, M., Pfaffman, J., & Schwartz, D. L. (2019). Educating and measuring choice: A test of the transfer of design thinking in problem solving and learning. *Journal of the Learning Sciences*, 28(3), 337-380.

Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2015). Posterlet: A game-based assessment of children's choices to seek feedback and to revise. *Journal of Learning Analytics*, 2(1), 49-71.

Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2017). Assessing whether students seek constructive criticism: The design of an automated feedback system for a graphic design task. *International Journal of Artificial Intelligence in Education*, 27(3), 419-447.

Cutumisu, M., & Schwartz, D. L. (2021). Feedback choices and their relations to learning are age-invariant starting in middle school: A secondary data analysis. *Computers & Education*, 171, 104215.

Dweck, C. (2012). *Mindset: How You Can Fulfil Your Potential*. New York: Constable & Robinson.

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational research review*, 5(2), 134-150.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology*, 108(1), 82.

Greenfield, P. M., Quiroz, B., & Raeff, C. (2000). Cross-cultural conflict and harmony in the social construction of the child. *New directions for child and adolescent development*, 2000(87), 93-108.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29-29.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). Cultures and organizations: Software of the mind, third edition (3rd ed.). McGraw-Hill Professional.

Kant, J. "Några invändningar mot Dylan William". Johan Kants blogg (blog), September 28, 2011, https://johankant.wordpress.com/2011/09/28/nagra-invandningar-mot-dylan-wiliam/

Kitayama, S., Markus, H. R., & Matsumoto, H. (1995). Culture, self, and emotion: A cultural perspective on “self-conscious” emotions.

Kluger, A. *N*., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current directions in psychological science*, 7(3), 67-72.

McInerney, D. M. (2008). The motivational roles of cultural differences and cultural identity in self-regulated learning. *Motivation and self-regulated learning: Theory, research, and applications*, 369-400.

Ogan, A., Walker, E., Baker, R., Rodrigo, M., Mercedes, T., Soriano, J. C., & Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International Journal of Artificial Intelligence in Education*, 25(2), 229-248.

Purdie, *N*., Hattie, J., & Douglas, G. (1996). Student conceptions of learning and their use of self-regulated learning strategies: A cross-cultural comparison. *Journal of educational psychology*, 88(1), 87.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582-599.

Suzuki, L. K., Davis, H. M., & Greenfield, P. M. (2008). Self-enhancement and self-effacement in reaction to praise and criticism: The case of multiethnic youth. *Ethos*, 36(1), 78-97.

Tärning, B., Lee, Y. J., Andersson, R., Månsson, K., Gulz, A., & Haake, M. (2020). Assessing the black box of feedback neglect in a digital educational game for elementary school. *Journal of the Learning Sciences*, 29(4-5), 511-549.