A GAME-BASED ASSESSMENT OF STUDENTS' CHOICES TO SEEK FEEDBACK AND TO REVISE

Maria Cutumisu, Doris B. Chin and Daniel L. Schwartz
Stanford University, Wallenberg Hall, 450 Serra Mall, Stanford, CA 94305

ABSTRACT

We introduce an educational game-based assessment that measuresthe choices students make while learning. We present Posterlet, a game designed to assess students' choices to seek negative feedback and to revise, in which students learn graphical design principles while creating posters. We validate our game-based assessment approach with three research studies, in whichcollege and middle-school students play Posterlet and then complete a posttest. Results showed that the game helped students learn: students who played Posterlet before completing the posttest learned more graphical design principles than students who only completed the posttest. Moreover, the choices to seek negative feedback and to revise can predict learning and be used as valid outcome measures for learning. We present a first-of-kind examination of students' choices to seek feedback and to revise, as well as of students' learning outcomes based on thesechoices, which can be used to develop and evaluate models of instruction that help students make informed learning choices.

KEYWORDS

Choice, assessment, game, learning, feedback, revision

1. INTRODUCTION

Jean Piaget stated, "The principal goal of education is to create men [and women] who are capable of doing new things, not simply of repeating what other generations have done…" (Elkind, 1968, p. SM80).Our educational goal is to prepare students for autonomy after they leave school, because the ability to self-govern, to learn, and to adapt in a changing world on one's own is crucial to a fulfilling life. Two major steps in operationalizing our goal are a) identifying the behaviors that capture students' potential to learn autonomously and b) employing tools that measure these behaviors. Researchers have identified promising behaviors for learning, such as 21st-century dispositions and attitudes (e.g., tolerance for ambiguity), but their relevance for improving learning has not yet been demonstrated, mostly because of a lack of assessment tools that can measure such behaviors.Here, we hypothesize that the choices students make when presented with a challenge constitute important behaviors for learning. Also, we introduce *choice-basedassessments* that measure not only students' knowledge,but even more importantly their choices about what, when, and how to learn. We validate these game-based assessment environments with three research studies showing that choices can both predict learning and be used as valid outcome measures for it.

2. THEORETICAL FRAMEWORK

We ground our research in the theoretical framework of *constructivist assessments* (Schwartz et al., 2009) and in *choice-based assessments* specifically (Schwartz and Arena, 2009; 2013). The core of our approach isusing choices, or learning-related behaviors, as first-order learning outcomes and measurement constructs. We need assessment environments that enable us to track students' choices and learning outcomes, so that we can examine the impact ofchoices on learning trajectories and measure students' potential to learn independently. Thus, our ideal assessment environments must satisfy the following principles.

2.1 Principles of Choice-Based Assessment Environments

Typical Performance. *Assessments need to capture every-day learning behaviours*. Students display "maximal performance" during assessment (Kleheand Anderson, 2007), which may not be congruent with their outside-school behaviors and performance. We need environments in whichstudents feel comfortable displaying their typical learning behaviors, to help us evaluate their true trajectories for lifelong learning.

Preparation for Future Learning.*Assessments need to offer learning opportunities*. Traditional assessments are retrospective, measuring students' knowledge at the end of instruction. Thus, they offer a snapshot of a current state of students' accumulated knowledge, but provide little indication regarding future learning trajectories. According to Vygotsky (1934), measuring student knowledge at the end of instruction instead of measuring learning processes progressively does not yield a holistic view of students' growth. If our goal is to measure behaviors that may be conducive to learning, we need to be able to embed learning in assessment environments. Schwartz and Bransford (1998) advocated *preparation for future learning*(PFL) assessments, which afford students opportunities to learn during the evaluation.

Choice.*Assessments need to gather information about unforced student choices*. We need to emulate the choice-rich environment in which students will learn after they leave school, because lifelong learning is based on free choices. Moreover,students must be able to safely experiment with choices before trying them in the real world. Some assessment environments blend learning and assessment, such as intelligent tutoring systems (Koedinger et al., 1997) but, in contrast to our approach that emphasizes the value of students' choices,they assume a certain sequence of steps that students take while learning, giving little choice to the students. Assessments cannot direct or influencestudents' choices. For instance, in a game-based assessment, students should be able to level up, no matter what learning choices they make in the game.

2.2 The Measurement Constructs

The Posterlet game is an instance of a choice-based PFL assessment designed to measure two behaviors important for learning: 1) the choice to seek negative feedback and 2) the choice to revise.Different learning contexts may afford many different learning choices that can be measured.We focused on the choice to seek negative feedback, because negative feedback tends to be more effective for continued learning than positive feedback (Klugerand DeNisi, 1998). At the same time, negative feedback runs the risk of triggering an ego threat that leads people to shut down rather than revise (Hattie and Timperley, 2007). While attitudes towards feedback are important for learning, there is no evidence whether the choice to seek feedback is important. In previous research, students did not exercise choice regarding feedback (but see Roll et al., 2011).Revisingmayalso bean important aspect of learning: choosing negative feedback would be of little use if students did not act on it. Although revising may be an important behavior for learning, there is no evidence in prior research whether the choice to revise is important. Thus, we investigate the impact of the *choicestoseek negative feedback*and*to revise* on learning.

3. THE ASSESSMENT ENVIRONMENT: POSTERLET

In Posterlet(Figure 1), players design posters for different booths at a funfair. On each level, after they complete their initial poster design, players choose three characters to provide positive feedback (e.g., "Your poster has big letters. Really easy to read.") or negative feedback(e.g., "People need to be able to read it. Some of your words are too small."), which carry equivalent information. Then, players choose whether to revise or submit their poster. There is one round of feedback and revision for each poster. The game has three rounds, with nine feedback and three revision choices.Finally, the game displays students' poster score as the number of tickets sold at each booth. The game's graphical analysis system tracks the posters for the use of 21 graphical design principles and generates poster-specific feedback. The principles governing choice-based assessments (typical performance, PFL, and choice) are naturally built into the game mechanics. Thegame provides an environment that encourages students' *typical behaviors*,offering themopportunities to *learn graphical design principles* while they are exercising their*feedback and revision choices*.

4. CONSTRUCT VALIDATION

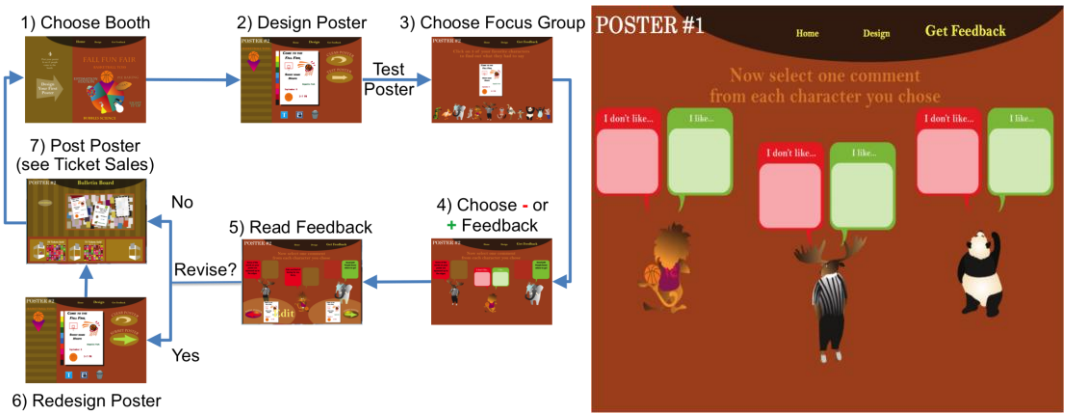

Figure 1.The Posterletgame flow (left) and the selection of positive or negative feedback (right)

Unlike assessments of knowledge, where the correct answer for 2+2 is 4 and not 5, choice-based assessments face a special challenge. To validate choices, we need to establish whether some choices (e.g., seeking negative feedback or revising) are better for learning than others. Since there is no literature available on this topic, the major goal of the current research is to show that some choices are better than others for learning.We will employ different forms of evidence to examine two aspects of validity: 1) *Internal*: Do choices to seek negative feedback and revise correlate with *in-game*learningoutcomes? We employed several subsets of learning measures across three different studies.2) *External*: Do students who seek negative feedback and revise exhibit better *in-school*learning? We employed academic achievement scores in Study 3.

Internal Learning Outcomes

Performancerepresents an in-game category of measures for students' poster design skills.Coding: We employed one measure, *Poster Quality*, to operationalize performance. The Posterlet game's graphical analysis system evaluates each poster by scoring each of the 21 design principles with 1 if used correctly, 0 if not applicable, and -1 if used incorrectly on that poster, producing an individualposter score.*Poster Quality*sums the individual poster scores of the last poster version on each game level (either the revised poster or the initial design, if the poster was not revised).*Poster Quality* ranges from -63 to 63.

Critiquerepresents a posttest category of measures for students' ability to judge posters. A posttest included two open-text questions. The first question asked students to provide some common mistakes a poster design novice might make. The second question asked students to provide written feedback on a sample poster. Coding: We employed two measures, *Common Mistakes*and*Written Feedback*, to operationalize critique. We scored each answerby counting the number of graphical design principles included in it.Each answer score (*Common Mistakes* and *Written Feedback*) ranges from 0 to 21.

Principle Selectionrepresents a posttest category of measures for students' understanding of graphical design principles. A posttest included two multiple-choice questions that asked students to choose, from a checklist of graphical design principles, the things that were good and bad, respectively, about a sample poster. Coding: We employed two measures, *Good Features*and*Bad Features*, to operationalize principle selection. We scored answers by assigning 1 point for each correctly checked answer and subtracting 1 point for each incorrectly checked answer. Each answer score (*Good Features*and*Bad Features*) ranges from -5 to 5, because each question has five correct and five incorrect answers.

Recognitionrepresents a posttest category of measures for students' ability to recognize at a glancemisused graphical design principles. A posttest included four sets of questions. For each set, students were shown successivelythree images: a poster, then a distractor Moiré pattern image, and then a second poster, each imagebeing displayed for 5 seconds. Students had to first decide whether the second poster was the same/better/worse compared to the first poster and then they had to provide a brief written explanation for their decision. We inserted a distractor image between the two posters, since humans display an exceptional memory for images (Standing, 1973). This posttest targets the following design rules: the text should not be on the poster's edge, the poster should have images, the images should be relevant to the poster's theme, and the text-background color contrast should be high. The posttest evaluates students on two dimensions: 1)

their ability to judge posters (i.e.,to decide whether the second poster was the same/better/worse compared to the first poster)and 2)their ability to justify their poster appraisal decision.Coding: We employed two measures, *Poster Ranking*and*Justify Ranking*, to operationalize recognition. *Poster Ranking* measures students' ability to judge posters.*Justify Ranking*measures students' ability to justify their decisions using graphical design principles, differentiating between novice (control) and expert(treatment) students in their abilities to perceive subtleties of design principles. Thus, students not only decide whether the posters are different, but they can also justify their decision using a new graphical design principle language they learned from the game's feedback. We scoredtheanswer that comparedthe first and the second poster ("same, better, or worse") in each of the four setsof questions by assigning 1for a correct and -1 for an incorrect answer (measured by *Text not on Edge*, *Image Present*, *Graphics Relevant*, and *Contrast High*,respectively).However, we scored*Poster Ranking* as the count of the correctanswers for each of the four poster comparisons (0-4), to be on the same scale as the measure for the justification of the answer.We scored eachjustification with 1, if the correct graphical design principle targeted by the question was included in the response, and with 0, otherwise. *Justify Ranking*is the sum of the scores across all poster comparisons (0-4).

External Learning Outcomes

Academic achievementrepresents an in-school category of measures for middle-school students'English Language Arts (*ELA*), Mathematics (*Math*), and Science(*Science*) standardized test achievement.

5. EXPERIMENTAL OVERVIEW

We describe three studies using Posterlet to gather evidence on the validity of choice as a learning construct.Study 1 is correlational: itinvestigates whether the choices to seek negative feedback and to revise correlate with internallearning outcomes for college students. However, being correlational, it raises the concern that students who chose negative feedback could have learned anyway, being successful learners in general. Thus, Study 2 is correlational and experimental:it investigates 1) whether Study 1 findings replicateeven when the learning measures change and 2) whether the concern raised in Study 1 is legitimate.It compares college students who did not play (Non-players or Control condition) and who played(Players or Treatment condition) Posterlet to gauge whether playing the game helps students learn design principles. Finally, Study 3 is correlational:1) it investigates if our findings generalize to other ages by sampling middle-school students and 2) it examines external validity through correlations between choices and external learning outcomes (in-schoolstandardized achievement).

5.1 Study 1

5.1.1 Design

Participants are N=109 community college students (63 females, 45 males, and 1 not reported) from California, aged 15-52, M_{age}=22.13 (SD=5.40). They played Posterlet individually (M=13 min),designing three posters,and then took an individual online posttest (M=3 min). We measured:

Learning Choices.*Negative Feedback* counts the number of times (out of 9)students choose the "I don't like" feedback, while*Revision* counts the number of revisions (out of 3) across posters.

Internal Learning Outcomes.*Poster Quality* measures in-game poster performance, described in Section 4. *Posttest* measures learning of the graphical design principles, adding the normalized scores of the four posttest questions in the**Critique** and **Principle Selection**categories, described in Section 4.

5.1.2 Results

Do learning choices correlate with internal learning outcomes?Table 1 shows the correlations among the learning choices and internallearning outcomes (poster performance measured by the game and learning of the graphical design principles measured by the posttest). Performance (*Poster Quality*) improved across levels: Level 1=9.76, Level 2=11.51, Level 3=12.10; $F(2,107)=13.46$, $p<.001$, so it can also be considered as a learning measure.Both choices(*Negative Feedback*, *Revision*) correlate with both internal learning outcomes (*Poster Quality*,*Posttest*), and strongly with each other.Also, *Poster Quality*correlated strongly with the posttest on graphical design principles, providing convergent validity of our learning measures.

Table 1.Correlations between negative feedback, revision, and internallearning outcomes ($^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$)

Measures	Negative Feedback	RevisionN=109	Poster QualityN=109	Posttest N=103
Negative Feedback	--	.52***	.21*	.20*
Revision		--	.37***	.33**
Poster Quality			--	.44***

*Negative Feedback*correlates strongly with*Revision*.To investigate whether they are unique learning predictors, we conducted stepwise linear regressionanalyses using*Poster Quality* and *Posttest*as the separate, dependent variables. For *Poster Quality*, only*Revision*enters as asignificant predictor(β=.37, F(1,107)=17.23, *p*<.001), accounting for 13.9% of the variance in performance (Adjusted R^2=.13).For*Posttest,Revision* is the significant predictor(β=.33, F(1,101)=12.57, *p*=.001), accounting for 11.1% of the variance in learning(Adjusted R^2=.10).We also explored the relation between learning choices and internal learning outcomes through a series of partial correlations, which revealed that *Revision* explained bothlearning outcomesmore than *Negative Feedback* did. Thus,*Negative Feedback*and *Revision*predictall internallearning outcomes, and *Revision* accounts for a little more of the variance in learning outcomes.

5.1.3 Discussion

Students seekingnegative feedback performed better on bothinternal learning outcomes and revised more. Students who chose to revise also performed better on bothinternal learning outcomes. Here, revisingseemsmore important for learning thanseekingnegative feedback, being a strong predictor of bothinternal learning outcomes.Finally, students who performed better on the posters also performed better on the posttest.

5.2 Study 2

5.2.1 Design

Participants were N=31 students (22 females, 9 males), aged 18-24, M_{age}=20.06 (SD=1.70), from the same college as Study 1. They were randomly assigned to one of two conditions, control and treatment. In the control condition, N=15 students (11 females, 4 males) took an online individual posttest (M=14 min). In the treatment condition, N=16 students (11 females, 5 males) played the Posterlet game individually (M=9 min, players were limited to 5 minutes per poster)designing two posters, before taking the same individual online posttest (M=13 min). Here, the posttest consisted of the two alternative learning measures,*Poster Ranking* and *Justify Ranking*,dimensions of the**Recognition**categorydescribed in Section 4.

5.2.2 Results

2a. Do learning choices correlate with internal learning outcomes?Table 2 shows the correlations among the learning choices and internal learning outcomes for Posterlet players. Performance (*Poster Quality*) improved across game levels: Level 1=10.00, Level 2=13.12; Wilks' Lambda=.66, F(1,15)=7.87, *p*=.01. Thus, performance can also be considered as a learningmeasure.*Negative Feedback*strongly correlates with *Revision*and with internal learning outcomes (*Poster Quality*,*Poster Ranking*, and*Justify Ranking*).Also, *Poster Quality* correlates with both posttest outcomes (*Poster Ranking* and*Justify Ranking*), providing convergent validity for thelearning measures.

Table 2.Correlations between choices and internal learning outcomes for Posterlet players ($^{**}p<.01$, $^{*}p<.05$)

Measures	Negative Feedback	Revision N=16	Poster Quality N=16	Poster Ranking N=16	Justify Ranking N=16
Negative Feedback	--	.64**	.59*	.56*	.57*
Revision		--	.37	.36	.31
Poster Quality			--	.64**	.64**

As before,*Negative Feedback* and *Revision* are strongly correlated. To investigate their uniqueness as predictors of learning, we conducted stepwise linear regressions using *Poster Quality* and *Poster Ranking* as separate, dependent variables.For both measures, only *Negative Feedback*enters as asignificant predictor:

β=.59,F(1,14)=7.61, *p*=.01, accounting for 35.2% of the variance inperformance (Adjusted R^2=.31) and, similarly,β=.56, F(1,14)=6.43, *p*=.02, accounting for 31.5% of the variance in the *Poster Ranking* (Adjusted R^2=.27).We also explored the relation between learning choices and internal learning outcomesthrough a series of partial correlations, which revealed that *Negative Feedback* explained both*Poster Quality*and *Poster Ranking* more than *Revision* did.We replicated these results for the alternative posttest measure, *Justify Ranking*.Thus,*Negative Feedback*predictsall internallearning outcomes.*Revision*, which strongly correlates with *Negative Feedback*, does not reach significance with the small sample size.

2b. Do treatment students outperform control students on the internal learning outcomes?We wanted to assess whether playing Posterlet improved students' perception of graphical design principles.A *t*-test comparedposttest learning between students in the treatment and control condition. On *Poster Ranking*, control students (M=1.20, SD=.68) were outperformed [*t*(29)=-3.74, *p*<.01] by treatment students (M=2.38, SD=1.02).On *Justify Ranking*, control students (M=1.27, SD=.80)were outperformed [*t*(29)=-3.52, *p*<.01]by treatment students (M=2.44, SD=1.03).On each individual posttest question, treatment students outperformed control students (left side of Figure 2). Figure 2 (left) shows that, on all questions (except for*Text not on Edge*), treatment students performed above chance (i.e., greater than -0.33, because only one out of three options is correct, the scores being 1 for correct and -1 for incorrect responses).

5.2.3 Discussion

We found the same correlation pattern between seeking negative feedback and internal learning outcomesas inStudy 1, although the posttest was replaced. Students who choose negative feedback also tend to revise. Although learning choices correlate strongly with each other, only seeking negative feedbackpredicts all internal learning outcomes.As before, performance correlates with posttest learning, providing convergent validityfor the learning measures.We show that students do learn from Posterlet: they did not already know design principles, because treatment students outperformedcontrol students on both dimensions of **Recognition**: they learned to better judge posters andshowed increased perception of the graphical design principles. We not only helped students learn, butquite possiblyimproved their perception of the world.

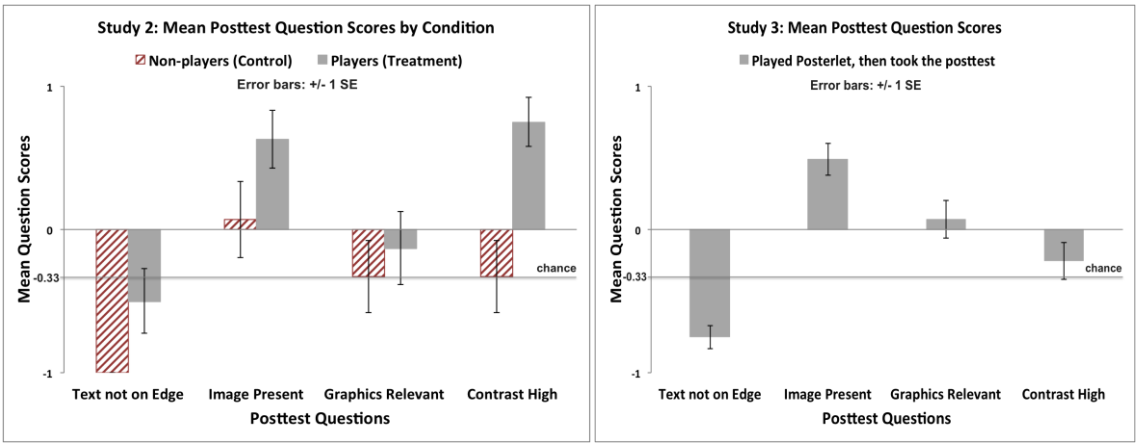

Figure 2. Study 2 posttest scores by condition (left) and Study 3 posttest scores (right)

5.3 Study 3

5.3.1 Design

Participants were N=976th-grade middle-school students from California, aged 11-12 years (46females, 51males).Students played Posterlet individually, in which they designed three posters, before taking an individual online posttest. Only N=80students provided consent and completed Posterlet (M=14 min)and, of these, only N=56 students completed the posttest (M=7 min), due to time constraints. We used the same choice and performance measures. Here, *Posttest* combines the measures of the **Principle Selection** category from Study 1 and *Poster Ranking* from Study 2, described in Section 4, summing their normalized Z-scores.

5.3.2 Results

3a. Do learning choices correlate with internal learning outcomes? *Negative Feedback* correlates strongly with *Revision* (Table 3). Performance (*Poster Quality*) improved significantly from the first poster: Level 1=9.76, Level 2=11.40, Level 3=11.35; Wilks' Lambda=.87, F(2,78)=6.01, p=.004. Thus, performance can also be considered as a learning measure. Only *Revision* correlates with the internal learning outcomes.

Table 3.Correlations between negative feedback, revision, and internal learning outcomes ($^{***}p$<.001, $^{**}p$<.01, $^{*}p$<.05)

Measures	Negative Feedback	RevisionN=80	Poster QualityN=80	Posttest N=56
Negative Feedback	--	.54***	.13	.22
Revision		--	.23*	35**
Poster Quality			--	.02

To investigate the uniqueness of choices as predictors of learning outcomes, we conducted stepwise linear regression analyses using *Poster Quality* and *Posttest* as separate, dependent variables. Only *Revision* enters as a predictor in the regression:β=.23, F(1,78)=4.19, p=.04, accounting for 5.1% of the variance in *Poster Quality* (Adjusted R^2=.04) and, similarly, β=.35, F(1,54)=7.65, p=.008, accounting for 12.4% of the variance in *Posttest* (Adjusted R^2=.11).As in Study 1, *Revision* predicts all internal learning outcomes. Figure 2 (right side) shows that, on all questions (except for *Text not on Edge*), students performed above chance(-0.33).

3b. Do learning choices correlate with external learning outcomes? Table 4 includes the correlations between internal and external learning outcomes. We received achievement scores for many of the students to answer this question (samples vary depending on the achievement scores available). Learning choices show stable correlations with *ELA* scores. Additionally, seeking negative feedback correlates with *Science*, while revising correlates with *Math*. Thus, learning choices in the game(internal)correlate with learning outcomes in the school (external). Also, learning on the post test correlates with *ELA* and *Science* scores. Thus, learning choices measured by the game and learning of the graphical design principles measured by a separate post test correlate with external learning outcomes measured by standardized assessments.

Table 4.Correlations between negative feedback, revision, posttest, and outside assessments ($^{**}p$<.01, $^{*}p$<.05)

Measures	ELA N=71	Science N=69	Math N=42
Negative Feedback	.39**	.26*	.30
Revision	.30*	.18	.35*
Posttest	.42**(N=48)	.41**(N=46)	.30(N=27)

5.3.3 Discussion

Students performed above chance on all posttest questions but the first, thus they learned design principles playing Posterlet. The first posttest question was challenging for most students, because it targeted a subtle design rule that not all students encountered as feedback. Learning choices strongly correlate with each other, but only revision predicts internal learning outcomes. More tuning of the posttest is necessary to ensure that it measures learning appropriately. Both learning choices in Posterlet predict in-school learning(*ELA*).Additionally, seeking negative feedback predicts *Science* and revising predicts *Math*. Notably, posttest learning (*Posttest*) predicts in-school learning(*ELA* and *Science).*

6. CONCLUDING DISCUSSION

We hypothesized that the choices to seek negative feedback and to revise would predict learning outcomes. Across the three studies, we found that the learning choices in the environment correlated with each other and with both learning outcomes, correlations between choices and learning outcomes being in the same range (.2 to .4) for bigger samples. A limitation of the work so far is that, while we find that choices correlate with learning outcomes, these relations bounce from Study 1 to Study 3, different choices being better for different outcomes. In Study1(college students) and Study 3 (middle-school students), revision predicts internal learning outcomes (poster performance and posttest learning), while in Study 2 (college students)negative feedback predicts internal learning outcomes. For college students (Studies 1 and 2), seeking negative feedback correlated with the internal learning outcomes (poster performance and posttest learning), which

also correlated with each other, supporting the convergent validity for the learning measures. This result did not generalize to middle-school students, likely due to developmental differences (Peters et al., 2014) or to the different posttest used in Study 3. Instead, for middle-school children (Study 3), the choice to revise correlated with children's internal learning outcomes (poster performance and posttest learning), as well as with their external learning outcomes (standardized achievement scores). For children, in-game choices and posttest learning correlated with external outcomes, indicating that choices can predict out-of-game learning.

We showed through both internal and external validation that choices such as these are beneficial in predicting learning, although more research is needed to clarify the contribution of each learning choice to learning outcomes. Students who played Posterlet learned to judge posters and to perceive design principles better than control students. Thus, we not only helped students learn, but quite possibly improved their perception of the world in the context of graphical design principles and poster design. Given these results, we showed that choices are important, but that the reliability of the assessment is not yet established. Our assessment was focused more on validating the choices rather than the learning outcomes. In summary, we developed a choice-based assessment game, Posterlet to track behaviors that we hypothesized are important to foster independent learning. We provided a first-of-kind demonstration that choices to seek negative feedback and to revise predict better learning in our assessment and in school. Our work may help design game environments focused on choices to enhance learning, by embedding learning choices into instruction models in both formal and informal environments.

ACKNOWLEDGEMENT

This work was supported by the NSF under Grant EHR-1228831. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the granting agencies. We would like to thank Neil Levine for his artwork and list of graphical design principles, as well as Jacob Haigh and Howard Palmer for their work on the Choicelet platform. We would also like to thank the students who participated in these studies, as well as their teachers.

REFERENCES

Elkind, D., 1968.Giant in the nursery – Jean Piaget.*The New York Times Magazine*, May 26, 1968, pp.SM25-SM80.Retrieved from http://search.proquest.com/docview/118265693.

Hattie, J. andTimperley, H., 2007. The power of feedback.*Review of Educational Research*, Vol. 77, No. 1, pp. 81–112.

Klehe, U.C. and Anderson, N., 2007. Working hard and working smart: motivation and ability during typical and maximum performance. *Journal of Applied Psychology*, Vol. 92, No. 4,pp. 978.

Kluger, A.N.and DeNisi, A., 1998. Feedback interventions: toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, Vol. 7, No. 3, pp. 67–72.

Koedinger, K.R. et al., 1997. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, Vol. 8, pp. 30-43.

Peters, S. et al.,2014. The neural coding of feedback learning across child and adolescent development, *Journal of Cognitive Neuroscience 26*(8), pp. 1705-1720.

Roll, I. et al., 2011. Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, Vol. 21, No. 2, pp. 267-280.

Schwartz, D.L. and Arena, D., 2009. Choice-based assessments for the digital age. *MacArthur 21st Century Learning and Assessment Project*.

Schwartz, D.L. and Arena, D., 2013. *Measuring what matters most: Choice-based assessments for the digital age*. MIT Press.

Schwartz, D.L. and Bransford, J.D., 1998. A time for telling.*Cognition and Instruction*, Vol. 16, pp. 475–522.

Schwartz, D.L.et al., 2009. Constructivism in an age of non-constructivist assessments. *Constructivist Instruction*,34-61.

Standing, L., 1973. Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, Vol. 25, pp. 207-222.

Vygotsky, L.S., 1934. *The collected works of LS Vygotsky: Problems of the theory and history of psychology*, Vol. 3.