Individual Differences as Predictors of Learning and Engagement

Sidney D'Mello (sdmello@memphis.edu)
Claire Williams (mcwllams@memphis.edu)
Patrick Hays (dphays@memphis.edu)
Andrew Olney (aolney@memphis.edu)

Institute for Intelligent Systems, University of Memphis
Memphis, TN 38152 USA

Abstract

We investigated the possibility of predicting students' engagement and learning gains during a tutoring session from trait measures of motivation, engagement, burnout, cognitive ability, prior knowledge, and task related measures. Participants completed a multiple choice pretest, a learning session, a posttest, and a battery of individual differences tests and questionnaires. Multiple regression and exploratory factor analyses indicated that the individual differences measures yielded medium sized effects at predicting learning gains as well as engagement levels that were self-reported during the tutorial session. In general, self-reported interest in the task and confidence in learning from a computer tutor coupled with working memory capacity and attentional abilities were the major predictors of both engagement and learning.

Keywords: learning, engagement, individual differences, cognitive abilities, motivation, burnout, ITS

Introduction

There is no one-size-fits-all approach when it comes to promoting student engagement and learning. Engagement and learning are affected by a number of factors such as, the learning environment (classroom, human tutor, high stakes learning), the task (acquiring shallow facts versus obtaining a deeper conceptual understanding), and characteristics of the learners themselves (e.g., visual versus verbal learners, performance versus mastery-oriented learners) (Ackerman, Sternberg, & Glaser, 1989; Jonassen & Grabowski, 1993; Schmeck & Geisler-Brenstein, 1989). Therefore, understanding how a particular student will be engaged in and benefit from a learning session requires an analysis of how the learning environment, the task, and the characteristics of the learner (i.e. individual differences) interact and influence learning outcomes.

For a given learning activity (e.g., learning conceptual physics from a human tutor), the context and the task are fixed, however the individuals involved in the activity vary. Hence, it is important to discriminate learners that actively engage and benefit from a learning session from others who passively attend the session and do not demonstrate dramatic improvements in their knowledge levels. Consequently, individual differences research has been a long standing and valuable tradition in the fields of psychology and education (Ackerman et al., 1989; Jonassen & Grabowski, 1993). Although research efforts along this front have yielded some important insights, there is little data on how individual differences influence engagement and learning within the context of intelligent learning environments such as Intelligent Tutoring Systems (ITSs). Understanding how individual differences impact learning sessions with ITSs is important, because ITSs are emerging as effective alternatives to deliver individualized instruction to large numbers of students (Corbett, Anderson, Graesser, Koedinger, & VanLehn, 1999; Graesser, Person, & Magliano, 1995; Koedinger & Corbett, 2006).

It is generally acknowledged that all students do not benefit equally from learning sessions with ITSs (VanLehn et al., 2007). Some learners show dramatic improvements in learning gains from pre to post tests, while tutoring has a negligible impact on others. Some learners actively attend the session by carefully listening to the tutor, taking initiative by asking questions, and providing verbose responses to the tutor's questions (Graesser et al., 1995). However, other non-critical learners, socially attend the session, and are comfortable being passive information receivers rather than active problem solvers. Who are these learners? Can they be discriminated from standard individual difference measures? What are the individual differences that are predictive of engagement and learning gains? These are the questions that motivated the present study.

The present study investigated whether trait measures of individual differences in (a) motivation, engagement, and burnout, (b) cognitive abilities, and (c) task related measures, could predict state measures consisting of engagement levels and learning gains in a one-on-one tutoring session with an ITS. Our focus on trait measures of motivation, engagement, and burnout is motivated by numerous studies that have related these measures to engagement and learning (Bartels & Magun-Jackson, 2009; Pekrun, Elliot, & Maier, 2006). For example, learners with mastery-approach motivation orientations are expected to be absorbed in the learning process (i.e., more engaged) and process the material deeply, presumably resulting in higher learning gains (Elliot & McGregor, 2001). In contrast, learners with performance-approach characteristics process the material at relatively shallow levels and do not demonstrate impressive learning gains. Similarly, some research has linked trait measures of engagement and burnout to performance outcomes (Schaufeli, Martinez, Pinto, Salanova, & Bakker, 2002).

Individual differences in cognitive abilities have previously been related to a variety of outcomes, hence, we expect them to be predictive of both engagement and learning with ITSs. For example, working memory capacity has been linked to performance on tests of fluid intelligence (Yuan, Steedle, Shavelson, Alonzo, & Oppezzo, 2006). Sustained attention has been related to academic achievement in school contexts (Steinmayr, Ziegler, & Träuble, 2010). In general, existing research has empirically demonstrated interactions between affect, working memory capacity, attention, intelligence, and performance outcomes (Linnenbrink, Ryan, & Pintrich, 1999; Steinmayr et al., 2010; Vergus & Boeck, 2002; Yuan et al., 2006). Hence, the present study focused on working memory capacity, selective and sustained attention, and general intelligence as predictors of engagement and learning gains.

In addition to the motivation, engagement, burnout, and cognitive variables, there is reason to suspect that individual differences pertaining to the learning task itself might be predictive of both engagement and learning gains. For example, task interest is likely to trigger curiosity and promote engagement (Berlyne, 1978), while prior knowledge is expected to be predictor of learning gains (VanLehn et al., 2007). More interestingly, there is some recent evidence that suggests that students' confidence of learning from a computer can be a better predictor of learning gains that other variables (e.g., initial motivation, prior knowledge) (Jackson, Graesser, & McNamara, 2009).

The present study investigated whether engagement and learning gains from a tutoring session in biology could be inferred from the aforementioned individual differences measures. More specifically, our analyses focused on (a) comparing the predictive power of three banks of predictors (motivation/engagement/burnout *versus* cognitive *versus* task), (b) assessing the predictive power of combined models that simultaneously include predictors from all three banks, (c) deriving principal components from the individual difference measures, and (d) correlating the derived components with engagement and learning gains.

Methods

Participants

Participants were 90 college students (non biology majors) who participated for course credit.

Description of Learning Environment

The study used a dialogue-based ITS that tutored students on eight topics in biology (e.g., cellular respiration, mitosis, ecological succession) via natural language dialogues. The ITS was designed to mirror the pedagogical and motivational strategies of lectures delivered by expert human tutors (D'Mello et al., in review).

Participants were randomly assigned to one of three versions of the ITS. In the *dialogue* version, the tutor primarily transmitted information (68% of the time) but occasionally provided cues for acknowledgements (e.g., "Right?", "ok?"), asked comprehension gauging questions (e.g., "Do you understand?"), and prompted the student for answers (e.g., "X is a type of what?"). Alternatively, in the *monologue* version, the tutor did all the talking and the student was a passive recipient. The third version consisted of *vicarious dialogues*, where the discourse patterns were structurally similar to the dialogue condition, but with one important exception. Here, it was a virtual student, instead of the learner, that answered the tutor's comprehension gauging questions and prompts. The virtual student always provided the correct answer (via simulated keystrokes) and the human learner simply watched the interaction.

The lectures were delivered via a simple conversational interface that consisted of an animated conversational agent that delivered the content of the lectures by means of synthesized speech, a media panel that displayed images relevant to the lectures, and an input box for students to type their responses for the dialogue condition. In the vicarious dialogue condition, the virtual student's responses were provided in the input box with simulated keystrokes. The simulated keystrokes were carefully calibrated in order to mirror the temporal dynamics of actual typing (i.e., onset delay, variable interstroke delay, and delay before hitting enter key to submit response).

Dependent Measures

Engagement Measures. Participants engagement levels were tracked at multiple points in the tutorial session with the affect grid (Russell, Weiss, & Mendelsohn, 1989) and through post-lecture questionnaires. The affect grid is a validated single item affect measurement instrument consisting of a 9 × 9 (valence × arousal) grid. Valence and arousal are the primary dimensions that underlie affective experiences. The arousal dimension ranges from sleepiness to high-arousal, while the valence dimension ranges from unpleasant feelings to pleasant feelings. Participants indicate their affective state by marking an X at the appropriate location on the grid.

The post-lecture questionnaire asked participants to self-report their engagement levels after each lecture. There were three questions which asked the participant to rate their engagement at the beginning, middle, and end of each lecture. Participants indicated their ratings on a six-point scale ranging from *very bored* to *very engaged.*

Knowledge Tests. The knowledge tests (used to measure prior knowledge and learning gains) were 24-item multiple-choice tests with three questions for each lecture. *Prompt* questions tested participants on content for which the tutor explicitly prompted the student in the dialogue and vicarious conditions. Although there were no explicit prompts in the monologue condition, we verified that the content of the prompts was explicitly covered in the monologue. *Assertion* questions tested participants on content that the tutor explicitly asserted to the student via direct instruction. Finally, there were *deep reasoning* questions that required causal reasoning, inference, etc. rather than recall of shallow facts. Participants completed alternate test versions for pretest and posttest that were counterbalanced across participants.

Individual Difference Measures

Motivation, Engagement, and Burnout. These measures consisted of: the Achievement Goals Questionnaire (AGQ) for motivation, the Utrecht Work Engagement Scale for Students (UWES-S) for trait engagement, and the Maslach Burnout Inventory Student Survey (MBI-SS) for burnout (Elliot & McGregor, 2001; Schaufeli et al., 2002).

The AGQ, a validated 12 item questionnaire, was used to classify participants' motivation levels as *performance-approach, performance-avoidance, mastery-approach,* and *mastery-avoidance* (Elliot & McGregor, 2001).

The UWES-S is a validated 14-item self-report measure of three dimensions of student engagement: *vigor*, *dedication*, and *absorption* (Schaufeli et al., 2002).

The MBI-SS is a validated 15-item self-report measure of three dimensions of student burnout: *exhaustion*, *cynicism*, and *professional efficacy* (Schaufeli et al., 2002).

Task Related Individual Differences. These measures consisted of pretest scores as a measure of *prior knowledge* in biology (see above) and a locally created Perceptions of Learning Biology Questionnaire (PLB). The PLB consisted of three questions that were designed to gauge participants' *interest* in learning biology, their perceived *usefulness* of learning biology, and their *confidence* that they could learn biology from a computer tutor.

Cognitive Measures. The cognitive measures consisted of: self-reported ACT or SAT scores as a measure of aptitude (these are standardized tests required for admission to universities in the US; SAT scores were converted to ACT scores in the present study), the validated Reading Span test (RSpan) to measure working memory capacity (Daneman & Carpenter, 1980), and the validated Ruff 2 and 7 Selective Attention test (Ruff 2 and 7) which measures selective and sustained attention (Ruff, Neimann, Allen, Farrow, & Wylie, 1992).

In each trial of RSpan, participants are presented with a logical or nonsensical sentence and an arbitrary letter that appears at the end of the sentence. They have to read the sentence out loud, determine if it was logical or nonsensical, and try to remember the unrelated letter. At recall, the participant typed the letters from the current set of trials in the correct order. The set sizes ranged from 2 to 5 letter strings (there were 3 trials of 2 character strings, 3 trials of 3 character strings, 4 trials of 4 character strings, and 2 trials of 5 character strings).

The measures from the RSpan include the *absolute span*, which is the highest set size (i.e., 2, 3, 4, or 5) that the participant recalled correctly, the *weighted span* (i.e., a score computed by weighting set size and items recalled), and the *total recalled* (i.e., the total number of items that the participant recalled correctly).

The Ruff 2 and 7 is a measure of selective and sustained attention (Ruff et al., 1992). It is a five-minute timed task with 20 trials (each trial is 15 seconds). For each trial, 30 targets (2's and 7's) were embedded in either a string of alphabetical capital letters (known as the automatic detection trials), or among strings of digits (known as the controlled search trials). Participants are required to spot the 2's and 7's from the distracters and click on them.

Selective attention was measured by the *automatic detection speed* and *accuracy* (the 10 letter trials) and by the *controlled search speed* and *accuracy* (the 10 digit trials). Sustained attention is measured by the *total speed* and *total accuracy* in the 20 trials.

Procedure

Participants were tested individually over a two hour session. They first completed an informed consent followed by the pretest and the Perceptions of Learning Biology questionnaire. Next, they read instructions on how to use the affect grid. On the basis of random assignment, participants then completed a tutorial session with either the monologue, dialogue, or vicarious version of the tutor. There were 30 participants in each condition. The tutoring session consisted of eight lectures that were randomly ordered for each participant. Random ordering was permissible because there was no major content overlap across lectures. Participants completed the affect grid and the post-lecture questionnaire after each lecture. They completed the posttest after the completion of all eight lectures. Finally, they completed the battery of individual difference measures after which they were fully debriefed.

Results and Discussion

We analyzed the data with multiple regression (MLR) and exploratory factor analysis techniques. The goal of the MLR analyses was to assess the predictive power of the three banks of predictors by comparing each bank separately, as well as building combined models that collectively considered all three banks. The factor analysis was used to extract principal components from the individual difference measures and to correlate the extracted components to the dependent measures (engagement and learning gains).

It is important to highlight some important points before describing the results. First, there were seven dependent variables: four learning gains measures and three engagement measures. The four learning gains measures were the corrected learning gains [(post – pre)/(1-pre)] for the prompt, assertion, and deep-reasoning questions, and an overall learning gains score (gains computed on all the items without segregating them into the different categories).

The three measures for engagement consisted of valence and arousal scores from the Affect Grid and a *composite engagement* score, which was the average engagement from the post lecture questionnaire (i.e., mean for each lecture of beginning engagement, middle engagement, and end engagement). Since the Affect Grid and post lecture questionnaires were administered eight times, once after each lecture, an aggregate value for valence, arousal, and composite engagement was computed for each participant by averaging the scores across lectures.

It is important to emphasize that the goal of the present paper is to identify the individual difference measures that predict learning and engagement and not to assess the impact of the tutor version (i.e., dialogue, monologue, vicarious). Previous analyses have compared our dependent measures as a function of tutor type (D'Mello et al., in review). Hence, the present analyses collectively analyzed all participants without considering tutor version.

Comparing Individual Predictor Banks

The goal of this analysis was to compare the predictive power of the different banks of predictors. This was accomplished by constructing 21 multiple regression models for the seven dependent variables and the three predictor banks. There were ten motivation and engagement predictors, four task related measures, and ten cognitive predictors.

Prior to constructing the regression models, we performed a correlational analysis to identify the most diagnostic set of predictors. In particular, any predictor that marginally-significantly correlated ($p < .10$) with at least one of the seven dependent measures was preserved for the subsequent analyses. This reduced the predictor set to four motivation and engagement predictors (performance-approach, performance-avoidance, vigor, and exhaustion), three task related predictors (prior knowledge, confidence, and interest), and seven cognitive predictors (ACT; absolute span, weighted span, total recalled from the RSpan test; automatic detection speed, controlled search speed, and total speed from the Ruff 2 and 7). Multicollinearity problems among these predictor sets were diagnosed and corrected with tolerance analyses prior to constructing the regression models.

Space constraints preclude an extensive discussion of the regression models constructed by examining each predictor set independently. Hence, the current discussion is limited to comparison of the predictive power of the three feature sets (coefficients will be examined in the subsequent analysis). R^2 adj. values as a measure of goodness of fit for regression models are presented in Table 1.

It appears that on average the cognitive predictors explained 10.2% of the variance for the learning gains measures, which is consistent with a small to medium sized effect (Cohen, 1992). Variance explained by the cognitive set was also quantitatively greater than the variance explained by the motivation/engagement/burnout and task related predictors, which were on par with each other (mean R^2 adj. = .044 and .053, respectively). In contrast, the three predictor sets were equally effective in predicting the engagement measures.

Multiple Predictor Sets

The next set of regression models were constructed from the predictors that were significant in the previous set of analyses. Here, predictors from all three feature sets were simultaneously considered and the significant predictors were identified via stepwise regression.

Table 1. R^2 *adj*. for regression models

Dependent Measure	**Individual Banks**			
	M,E,B	*Task*	*Cog*	**Combined**
Learning				
Prompt	0[c]	0[c]	.085	.113
Assertion	.111	.027[b]	.039	.122
Deep	0[c]	.053	.129	.194
Overall	0[c]	.062	.156	.149
Mean	*.028*	*.036*	*.102*	*.145*
Engagement				
Valence	.047	.030	.067	.082
Arousal	.066	.111[b]	.061	.197
Composite	.081	.086	.136	.169
Mean	*.065*	*.076*	*.088*	*.149*

Notes. All models significant at $p < .05$ unless noted otherwise. [b] significant at $p < .10$, [c] not significant ($p > .10$). *M,E,B* = motivation, engagement, burnout. *Cog* = Cognitive.

Learning Gains. There were statistically significant models for learning gains on prompt questions, assertion questions, deep reasoning questions, as well as for total learning gains (see Table 1). On average, the combined feature sets explained .145 of the variance, which approaches a medium sized effect (Cohen, 1992) and represents a 43% improvement in the variance explained by considering the best feature set independently (i.e., cognitive features).

Turning our focus to the significant predictors of the regression models (see Table 2), it appears that students with higher working memory abilities performed well on prompt questions. Surprisingly, self-reported exhaustion scores positively predicted performance on assertion questions; this finding warrants further analysis.

Deep reasoning questions, however, were predicted by a combination of self-reported interest in learning biology as well as a high ability to sustain attention. Total learning gains, however, were predicted by a combination of working memory capacity and sustained attention, indicating that the cognitive variables are the most relevant.

Table 2. Direction (+, -) of significant predictors

Predictor	**Learning Gains**				**Engagement**		
	P	*R*	*D*	*O*	*A*	*V*	*C*
Perf-Approach						+	
Exhaustion		+					
Interest			+		+	+	+
Absolute Span	+				+		
Weighted Span						+	
Total Recalled				+			+
Total Speed			+	+			
Contrl. Srch. Speed							+[a]

Notes. [a] $p = .056$; $p < .05$ for other predictors; P, R, D = gains for prompt, assertions, and deep questions, respectively. O = overall learning gains. A, V, C = arousal, valence, and composite engagement, respectively.

Engagement. Statistically significant models were obtained for arousal, valence, and the composite engagement score. These models explained an average of 14.9% of the variance, which is consistent with a 70% improvement over the best individual model (cognitive features; see Table 1).

An examination of the significant coefficients of the regression models for engagement indicated that task interest and working memory capacity were the most diagnostic predictors (see Table 2). In particular, arousal was predicted by task interest and absolute span. Valence was predicted by task interest, weighted span, and with a performance-approach motivational orientation. Finally, composite engagement was predicted by task interest, total items recalled during the RSpan test, and controlled search speed (an important characteristic of selective attention). Simply put, being interested in the learning session and having the requisite cognitive ability (working memory span and attention) to handle the difficulties and demands of the session were the major predictors of engagement.

Factor Analysis

We analyzed the individual differences with an exploratory factor analysis (principal components analysis with varimax rotation and Kaiser normalization). The analysis was conducted on 18 out of the 24 predictors because the inclusion of some of the predictors from the RSpan and Ruff 2 and 7 tests posed problems with respect to the factorability of the data. Specifically, only the absolute span measure from the RSpan test and the total speed and total accuracy scores from the Ruff 2 and 7 test were included.

Several indicators of factorability on the model with 18 predictors indicated that the data were in fact factorable. In particular, (a) the Kaiser-Meyer-Olkin measure of sampling adequacy was .72, which is above the recommended value of .6, (b) Bartlett's test of sphericity was significant (χ^2 (153) = 287.16, $p < .05$), (c) the diagonals of the anti-image correlation matrix were all above .5, which supports the inclusion of each item in the factor analysis, and (d) the commonalities were above .3, which indicates that each item shared a degree of common variance with the other items.

The analysis yielded six components with eigen values greater than 1 that collectively accounted for 63.4% of the variance (see Table 3). It appears that Component 1, which consists of a combination of predictors from the UWESS-S, MBI-SS, and AGQ represents highly engaged, low burnout, and mastery-approach oriented learners. This component accounted for 18.9% of the variance. In contrast, Component 2 (10.3% variance) represents learners with mastery and performance-approach tendencies. Component 3 (9.5% variance) represents learners that have some prior knowledge in biology and they find it interesting and useful, while Component 4 (9.4% variance) is consistent with learners that are intelligent and have high attention abilities. Component 5 (8% variance) represents learners have a large working memory and are confident that they can learn biology from a computer tutor. Finally, Component 6 (7.2% variance) consists of learners that are absorbed, but have a performance-avoidance motivational orientation.

Our analyses proceeded by correlating the individual difference measures with the six extracted components (see Table 4). As evident from the table, components 4 and 5 are the major predictors. In particular, component 5 correlates with six out of the seven dependent measures, thereby indicating that confidence in learning biology from a computer tutor coupled with large working memory capacity and attentional ability is the individual difference component that predicts engagement and learning.

Table 3. Factor loadings

	Components					
Item	1	2	3	4	5	6
Dedication	.83					
Cynicism	-.80					
Pro Efficacy	.76					
Exhaustion	-.68					
Vigor	.61					.40
Mast Approach	.61	.50				
Mast Avoid		.73				
Perf Approach		.68				
Interest			.75			
Useful			.73			
Prior Knowledge			.60	.42		
ACT				.84		
Total Accuracy				.67		
Total Speed		.39		.40	.36	-.33
Absolute Span					.73	
Confidence			.35		.73	
Perf Avoid		.48				.71
Absorption	.36					.62

Note. Items sorted by size and values < .3 are suppressed

Table 4. Correlations between dv's and components

Dependent Measure	**Components** 1	2	3	4	5	6
Learning						
Prompt	-.111	-.018	-.008	.183 [b]	.200 [b]	-.036
Assertion	.016	-.041	.128	.030	.131	-.133
Deep	.133	-.017	.160	.302 [a]	.264 [a]	-.055
Total	.035	-.049	.162	.316 [a]	.288 [a]	-.059
Engagement						
Valence	.052	.209 [b]	.259 [a]	.028	.202 [b]	.108
Arousal	.047	-.041	.113	.101	.252 [a]	.062
Mean E.	.075	.136	.242 [a]	.214 [a]	.291 [a]	.101

Notes. [a] significant at $p < .05$, [b] significant at $p < .10$

General Discussion

The present study investigated the possibility of predicting students' engagement and learning gains during a tutoring session with an ITS on the basis of individual differences in motivation, engagement, burnout, cognitive abilities, and task related measures. The results supported the conclusion that the cognitive factors reigned supreme when it comes to predicting learning outcomes; however, all three predictor banks were equivalent for predicting engagement. When models were combined, the individual difference measures explained 15% of the variance in engagement and learning gains, which is consistent with a medium effect (Cohen, 1992). In general, interest in the task, confidence in learning from a computer tutor, large working memory capacity, and heightened attentional abilities were the major predictors of both engagement and learning.

Our findings have important implications for the design of ITSs that aspire to be dynamically adaptive to individual learners. These ITSs construct sophisticated student models and utilize them to tailor the instruction to each students zone of proximal development (Koedinger & Corbett, 2006). The models are usually constructed on the basis of how students' knowledge in a particular domain meshes with the material that the tutor is expected to cover. In our view, a brief pretesting session on some of the individual difference measures coupled with the existing student modeling approaches will yield more accurate models that can guide individualized instruction. How these models are utilized to heighten engagement and enhance learning gains awaits further research and technological development.

Acknowledgments

This research was supported by the Institute of Education Sciences (R305A080594). The opinions expressed are those of the authors and do not represent views of IES.

References

Ackerman, P. L., Sternberg, R. J., & Glaser, R. (Eds.). (1989). *Learning and individual differences: Advances in theory and research*. New York: Freeman.

Bartels, J. M., & Magun-Jackson, S. (2009). Approach-avoidance motivation and metacognitive self-regulation: The role of need for achievement and fear of failure. *Learning and Individual Differences, 19*(4), 459-463.

Berlyne, D. (1978). Curiosity in learning. *Motivation and Emotion, 2*, 97-175.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Corbett, A., Anderson, J., Graesser, A., Koedinger, K., & VanLehn, K. (1999). Third generation computer tutors: Learn from or ignore human tutors? In *Proceedings of CHI Conference on Human Factors in Computing Systems* (pp. 85 - 86). New York: ACM.

D'Mello, S., Hays, P., Williams, C., Cade, W., Brown, J., & Olney, A. (in review). *Collaborative Lecturing by Human and Computer Tutors*

Daneman, M., & Carpenter, P. A. (1980). Individual difference in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(450-466).

Elliot, A., & McGregor, H. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*(3), 501-519.

Graesser, A., Person, N., & Magliano, J. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*(6), 495-522.

Jackson, G. T., Graesser, A. C., & McNamara, D. (2009). What Students Expect May Have More Impact Than What They Know or Feel. In V. Dimitrova, R. Mizoguchi, B. DuBoulay & A. Graesser (Eds.), *Artificial Intelligence in Education - Building Learnning Systems That Care: from Knowledge Representation to Affective Modelling* (Vol. 200, pp. 73-80).

Jonassen, D. H., & Grabowski, B. L. (Eds.). (1993). *Handbook of Individual Difference, Learning, and Instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Koedinger, K., & Corbett, A. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61-78). New York, NY: Cambridge University Press.

Linnenbrink, E. A., Ryan, A. M., & Pintrich, P. R. (1999). The role of goals and affect in working memory functioning. *Learning and Individual Differences, 11*(2), 213-230.

Pekrun, R., Elliot, A., & Maier, M. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology, 98*(3), 583-597.

Ruff, R. M., Neimann, H., Allen, C. C., Farrow, C. E., & Wylie, T. (1992). The Ruff 2 and 7 Selective Attention Test: A neuropsychological application. *Perceptual and Motor Skills, 75*(1311-1319).

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid - a Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology, 57*(3), 493-502.

Schaufeli, W. B., Martinez, I. M., Pinto, A. M., Salanova, M., & Bakker, A. B. (2002). Burnout and engagement in university students - A cross-national study. *Journal of Cross-Cultural Psychology, 33*(5), 464-481.

Schmeck, R. R., & Geisler-Brenstein, E. (1989). Individual differences that affect the way that students approach learning. *Learning and Individual Differences, 1*(1), 85-124.

Steinmayr, R., Ziegler, M., & Träuble, B. (2010). Do intelligence and sustained attention interact in predicting academic achievement? *Learning and Individual Differences, 20*, 14-18.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*(1), 3-62.

Vergus, T., & Boeck, P. D. (2002). On the correlation between working memory capacity and performance on intelligence tests. *Learning and Individual Differences, 13*, 37-55.

Yuan, K., Steedle, J., Shavelson, R., Alonzo, A., & Oppezzo, M. (2006). Working memory, fluid intelligence, and science learning. *Educational Research Review, 1*, 83-98.