

Leanna B. Aker & Arthur K. Ellis (USA)

A Meta-Analysis of Middle School Students' Science Engagement

Abstract: The extent to which middle school students are engaged in required science courses is an elusive but increasingly documented phenomenon. Anecdotal and empirical evidence alike raise concern with a perceived decline in science engagement reported by students as they transition into the middle school setting. Even what it means to be engaged is not thoroughly agreed on. Though an agreed-on operational definition of engagement is still nascent, an emerging consensus on a three-faceted model of student engagement exists in the research literature (Fredricks, Blumenfeld, & Paris, 2004). Thus, a synthesis of existing primary research of early adolescents' science engagement under this emerging conceptualization is warranted. The results of this meta-analysis indicate that instructional methods, class characteristics and competence predictors comprise the strongest relationship with self-reported science engagement in early adolescence. These predictors also show the strongest relationship with affective and cognitive engagement sub-types. Though affective and cognitive engagement were well-represented in primary studies, behavioral engagement was under-represented in student self-reports.

Keywords: meta-analysis, behavioral engagement, cognitive engagement, affective engagement, science, middle school, junior high school, early adolescence, self-determination theory, stage-environment fit theory

摘要 (Leanna B. Aker & Arthur K. Ellis: 关于初中生科学课参与度的元分析研究) : 初中生在必修科学课程中的参与度是一个难以捉摸, 但逐渐获得关注的现象。相关的轶事证据和实验性证据引起了各界对学生在进入初中阶段后所报告的科学课参与度下降问题的担忧。各界在如何定义参与度这一问题上也未能达成共识。虽然研究者对参与度的操作性定义的意见并不完全一致, 但是研究文献中已开始出现一些对学生参与度三面模型的共识 (Fredricks, Blumenfeld, & Paris, 2004)。因此, 在此新兴概念化背景下, 有必要对针对初中生的科学课参与度的各项研究进行综合分析。

本项元分析的结果表明, 教学方法、课堂特征和能力预测因素与初中学生自我报告的科学参与度有密切的关系。同时, 这些因素也显示出与情感和认知这一参与度子类型情感的密切关系。此外, 元分析的结果显示, 虽然目前对情感和认知参与的研究比较充分, 但对行为参与的研究相对缺乏。

关键词: 元分析, 行为参与, 认知参与, 情感参与, 科学, 初中, 青春期早期, 自我决定理论, 阶段-环境契合理论

摘要 (Leanna B. Aker & Arthur K. Ellis: 關於國中學生科學課參與度的元分析研究) : 國中生在必修科學課程中的參與度是一個難以捉摸, 但逐漸獲得關注的現象。相關的軼事證據和實驗性證據引起了各界對學生在進入國中階段後所報告的科學課參與度下降問題的擔憂。各界在如何定義參與度這一問題上也未能達成共識。雖然研究者對參與度的操作性定義的意見並不完全一致, 但是研究文獻中已開始出現一些對學生參與度三面模型的共識 (Fredricks, Blumenfeld, & Paris, 2004)。因此, 在此新興概念化背景下, 有必要對針對國中生的科學課參與度的各項研究進行綜合分析。

本項元分析的結果表明, 教學方法、課堂特徵和能力預測因素與國中學生自我報告的科學參與度有密切的關係。同時, 這些因素也顯示出與情感和認知這一參與度子類型情感的密切關係。此外, 元分析的結果顯示, 雖然目前對情感和認知參與的研究比較充分, 但對行為參與的研究相對缺乏。

關鍵詞: 元分析, 行為參與, 認知參與, 情感參與, 科學, 國中, 青春期早期, 自我決定理論, 階段-環境契合理論

Zusammenfassung (Leanna B. Aker & Arthur K. Ellis: Eine Meta-Analyse des wissenschaftlichen Engagements von Schülern der Mittelstufe): Das Ausmaß, in dem Schülerinnen und Schüler der Mittelstufe

an naturwissenschaftlichen Pflichtkursen teilnehmen, ist ein schwer zu erfassendes, aber zunehmend dokumentiertes Phänomen. Anekdotische und empirische Beweise werfen gleichermaßen Bedenken hinsichtlich eines wahrgenommenen Rückgangs des wissenschaftlichen Engagements der Schüler beim Übergang in die Mittelschule auf. Selbst was es bedeutet, engagiert zu sein, ist nicht gründlich vereinbart. Obwohl eine abgestimmte operative Definition von Engagement noch im Entstehen begriffen ist, gibt es in der Forschungsliteratur einen sich abzeichnenden Konsens über ein dreigliedriges Modell des studentischen Engagements (Fredricks, Blumenfeld, & Paris, 2004). Somit ist eine Synthese der bestehenden Primärforschung des wissenschaftlichen Engagements früher Jugendlicher im Rahmen dieser aufkommenden Konzeptualisierung gerechtfertigt. Die Ergebnisse dieser Meta-Analyse deuten darauf hin, dass Unterrichtsmethoden, Klassenmerkmale und Kompetenzprädikatoren die stärkste Beziehung zum selbstberichteten wissenschaftlichen Engagement in der frühen Adoleszenz darstellen. Diese Prädiktoren zeigen auch die stärkste Beziehung zu den Subtypen des affektiven und kognitiven Engagements. Obwohl affektives und kognitives Engagement in Primärstudien gut vertreten war, war das verhaltensbedingte Engagement in den Selbstberichten der Schüler unterrepräsentiert.

Schlüsselwörter: *Meta-Analyse, Verhaltens-Engagement, kognitives Engagement, affektives Engagement, Wissenschaft, Mittelschule, Gymnasium, frühe Adoleszenz, Selbstbestimmungstheorie, Stage-environment fit theory*

Аннотация (Леанна Б. Акер & Артур К. Эллис: Метаанализ научной вовлеченности учащихся средних классов): масштабы участия учащихся средних классов в естественно-научных курсах обязательных дисциплин сложно выразить в цифровом эквиваленте, однако данный феномен подвергается всё большему описанию. Как неverified, так и эмпирически подтвержденные случаи не проливают свет на причины регистрируемого уменьшения вовлеченности учащихся в научную жизнь при переходе на среднюю ступень обучения. Само понятие «вовлеченность» также не имеет общезакрепленного определения. Несмотря на то, что рабочее определение вовлеченности еще находится на стадии уточнения, в научно-исследовательской литературе отмечается единство взглядов в вопросе трехмерности модели студенческой вовлеченности (Фредрикс, Блюменфельд & Парис, 2004). Тем самым представляется оправданной попытка синтезировать актуальные данные первичных исследований степени вовлеченности подростков в научную жизнь в рамках зарождающейся концептуализации данного феномена. Результаты этого метаанализа указывают на то, что в раннем подростковом возрасте методы преподавания, характеристики класса и предикаторы компетенций имеют самое непосредственное отношение к рефлексии над собственной научной вовлеченностью. Эти предикаторы демонстрируют также самую тесную связь с подтипами аффективной и когнитивной вовлеченности. Аффективный и когнитивный типы вовлеченности в первичных исследованиях были представлены достаточно полно, в то время как поведенческая составляющая вовлеченности в процессе самозвальной оценки отмечалась у учащихся реже.

Ключевые слова: *метаанализ, поведенческая вовлеченность, когнитивная вовлеченность, аффективная вовлеченность, наука, средняя школа, гимназия, ранний подростковый возраст, теория самоопределения, модель взаимного соответствия индивидуума и среды*

Introduction

The problematic nature of student engagement with school science has been a concern of science researchers and practitioners for several decades as student interest in, and attitudes toward, science as a school subject appears to have waned (Jenkins & Pell, 2006; Lee & Anderson, 1993; Osborne, Simon, & Collins, 2003; Polkin & Hasni, 2014). This decline often coincides with the transition into middle school (Braund & Driver, 2005; Eccles et al., 1993; Eccles & Roeser, 2010; Mahatmya, Lohman, Matjasko, & Farb, 2012). However, researchers have demonstrated that declining engagement is not an inevitable outcome of the transition to middle school (Anderman & Maehr, 1994; Eccles et al., 1993; Vedder-Weiss & Fortus, 2011). There is good reason to think that early adolescence is a time of rich developmental potential to engage cognitively in abstract

reasoning, considering multiple perspectives, and weighing several strategies simultaneously (Mahatmya et al., 2012; Piaget, 1972).

Self-determination theory (SDT) and stage-environment fit (SEF) theory offer anchors to guide an evaluation of research about early adolescents' engagement with middle school science. SDT posits that students are most likely to be motivated when they feel a sense of competence, autonomy, and relatedness (Roeser & Eccles, 1998; Ryan & Deci, 2000). SEF theory suggests that an appropriate fit between the educational environment and students' developmental needs will lead to increased engagement (Eccles & Midgley, 1989, Eccles et al., 1993, p. 90). As early adolescents are unique in their increasing developmental need for autonomy and relatedness, these two theories provide a lens with which to evaluate engagement research at this age level (Steinberg & Silverberg, 1986; Soenens & Vansteenkiske, 2010).

As conceptual and operational clarity emerges about engagement, a meta-analysis of existing engagement studies in the area of school science is a logical next step toward increased coherence for this body of research. Studies exist in the research literature that purport to measure engagement but which use operationalizations that are incongruent with the emerging consensus about the construct. In 1991, a meta-analysis of engagement was conducted that focused almost exclusively on behavioral indicators of engagement with scant attention to affective or cognitive factors. While observable student behavior is indeed an indicator, it represents a limited subset of what is now considered a more complex description of engagement (Kumar, 1991). On the other hand, there are studies that are not identified as engagement-related, yet assess indicators of behavioral, affective, or cognitive engagement. A purposeful, updated synthesis of engagement and engagement-related research serves to solidify an operational definition of the construct.

The identification of practically significant predictors of engagement stands to benefit educational practitioners. Engagement is intuitively understood by educators and viewed as malleable and responsive to teacher practices (Finn & Zimmer, 2012; Singh, Granville, & Dika, 2002; Skinner & Pitzer, 2012). A synthesis of existing research can inform possible interventions that positively impact student engagement with specific science tasks. Identifying effective predictors of each type of engagement can inform targeted interventions to address certain specific engagement issues.

Engagement

The term "engagement" is ubiquitous in the educational field, appearing in teacher evaluation criteria, educator vernacular, and educational research. Part of the reason that the term is so pervasive is that it has such an intuitive meaning in education. This intuitive meaning is reflected in different definitions of engagement found in the research literature. Examples include the following: "the student's psychological investment in and effort directed toward learning, understanding, or mastering the knowledge, skills, or crafts that academic work is intended to promote" (Newmann, 1992, p. 12), "the attention...investment, and effort students expend in the work of school" (Marks, 2000, p. 155), and "constructive, enthusiastic, willing, emotionally positive, and cognitively focused participation with learning activities in school" (Skinner & Pitzer, 2012, p. 22). Thus, engagement refers to the nature and quality of a student's participation in school and its academic tasks.

Despite this intuitive meaning, or perhaps because of it, engagement has only recently begun to become operationalized as a construct. Some researchers criticize engagement as subsuming, duplicating, or overlapping existing educational constructs, such as motivation (Azevedo, 2015; Fredricks et al., 2004). Due to historical changes in both the construct itself and its grain size of interest—differentiating facilitators, indicators, and outcomes of engagement has also presented challenges. While differing engagement models exist in the research literature, each fundamentally attempts to describe and differentiate high and low quality engagement.

A seminal synthesis of engagement research proposes a model that has been increasingly adopted by educational researchers. Fredricks and her colleagues (2004) suggested that engagement is a meta-construct with three facets—behavioral, cognitive, and affective. Behaviorally engaged students show on-task actions such as attention and participation (Caraway & Tucker, 2003, p. 417). Affectively engaged students are interested, see value in the tasks they are given, and have positive emotions about what they are experiencing (Fredricks et al., 2004). Cognitively engaged students are self-regulated learners, use multiple strategies for learning, and show effort beyond what is required (Azevedo, 2015; Fredricks et al., 2004; Pintrich & DeGroot, 1990).

The three-faceted model of engagement has come to dominate the research literature—it has been validated psychometrically, used to examine and categorize psychometric instruments, taken up and cited by researchers in subsequent studies, and used to interpret existing research about engagement (Doğan, 2014; Fredricks et al., 2004; Fredricks, McColskey, Meli, Montrosse, Mordica, & Mooney, 2011; Sinatra, Heddy, & Lombardi, 2015; Veiga, Reeve, Wentzel, & Robu, 2014; Wang & Holcombe, 2010; Wang, Willett, & Eccles, 2011). Furthermore, each type of engagement can be disaggregated and understood as a distinct entity—one can, for example, imagine a situation in which a student is behaviorally but not cognitively engaged; and the cognitively engaged high achiever who works hard for good grades but professes no real interest in a subject represents a near folklore-like caricature.

Methodology

Literature Search. This meta-analysis includes a comprehensive literature review based on both published and grey literature. Included studies were published between 2006-2016, involved participants from 10 to 15 years old (grades 5-9), and written in or translatable to English. As causality was not desired, the search accommodated a variety of methodological designs, including experimental, quasi-experimental, repeated measures, correlational (e.g., correlational, regression), and ex post facto. Studies were excluded if they did not report effect sizes or the statistics necessary to calculate the effect size and its precision.

Characteristics of the engagement predictors and indicators further limited the number of included studies. Included studies examined science engagement predictors and outcomes that are malleable at the classroom or task level. For example, studies that primarily examined science content as predictors of engagement were excluded. The assessment of engagement indicators could be explicit or implicit, but could only be accomplished through student self-report. The decision about whether a study implicitly measured engagement was informed by guidelines from the research literature (Fredricks et al., 2004; Skinner & Pitzer, 2012).

Coding. A number of potential moderators of middle school science engagement were coded, including publication and peer-review status, grade level, school structure, school type, school setting, geographic location, socio-economic status, experimental design, instrument reliability and validity, and repeat authors. Additionally, engagement outcomes were coded as behavioral, affective, cognitive, or a combination thereof. Engagement predictors were coded by predictor type (instructional methods, technology, class characteristics, and social characteristics), as well as by self-determination theory component (autonomy, competence, and relatedness).

Statistical Analysis. Given that true effect sizes were expected to differ from study to study, a random effects meta-analysis was conducted, utilizing Comprehensive Meta-Analysis (CMA), Version 3 (Biostat, 2015) to conduct the meta-analysis, an online effect size calculator (Wilson, 2015) for effect size calculations not offered within the program, and Microsoft Excel to perform sub-calculations and examine descriptive statistics. Sub-analyses were conducted using random

effects meta-regression to determine the relationship between various predictors or moderators and engagement outcomes.

Hedges' g was selected as a common effect size metric for comparing studies. Effect sizes reflecting measures of association (the r -family of effect sizes) were converted to the d -family of effect sizes within CMA, and then Cohen's d values were then converted to Hedges' g within CMA. Guidelines for interpretation of effect sizes as strong ($g > 2.7$), moderate ($g > 1.15$), minimum ($g > .41$), and no practical effect ($g < .41$) were established by Ferguson (2009). While some studies produced single effect sizes, other studies reflected complex data structures. For independent groups, data were pooled together via a mini meta-analysis to yield a single effect size for each study. For non-independent subgroups, the pooling of data was conducted using a variance that corrected for the correlation among multiple outcomes. Values for high (.8), moderate (.5), and low correlation (.2), were assigned following guidelines proposed by Ferguson (2009). Identical engagement outcomes (e.g., affective and affective) were designated highly correlated, while different engagement outcomes (e.g., affective and cognitive) were designated moderately correlated.

Results

Seventy-nine studies met inclusion criteria. The majority of studies were published ($k = 58, 73.4\%$) and peer-reviewed ($k = 52, 67.6\%$). Sample sizes ranged from 20 to 10,437, with an overall sample size of 53,971 for the meta-analysis. Sixteen of the 79 studies yielded multiple engagement predictors. Predictors were coded both by type and by self-determination theory component. Instructional method ($n = 57, k = 40$) and class characteristics ($n = 60, k = 20$) were the most common predictor types (see Table 1). Autonomy SDT predictors were most common ($n = 94, k = 22$), followed by relatedness ($n = 35, k = 49$) and competence ($n = 29, k = 21$). The number of studies sums to more than 79 as some studies included more than one engagement predictor. A full list of studies can be found in Aker's dissertation (2016).

Table 1: *Descriptive Statistics for Predictor Classification*

Predictor classification	Point estimates		Studies	
	n	Percent	k	Percent
Type				
Instructional Method	57	36.1%	40	48.2%
Technology	15	9.5%	13	15.7%
Class Characteristics	60	37.9%	20	24.1%
Social Characteristics	26	16.5%	10	12%
Self-determination theory				
Autonomy	94	59.4%	22	23.9%
Competence	29	18.4%	21	22.8%
Relatedness	35	22.2%	49	53.3%

Twenty-three of the 79 studies yielded multiple engagement outcomes (see Table 2). The most common outcome provided by the studies was affective engagement ($n = 84, k = 56$), followed by cognitive engagement ($n = 49, k = 31$), combinations of two engagement outcomes ($n = 13, k = 9$),

behavioral engagement ($n = 10, k = 7$), and combinations of all three engagement outcomes ($n = 2, k = 2$). The number of studies summed to more than 79 ($k = 105$) because some studies provided data about more than one engagement type.

Table 2: *Descriptive Statistics for Engagement Outcomes*

Engagement type	Point estimates		Studies	
	n	Percent	k	Percent
Behavioral	10	6.3%	7	6.7%
Affective	84	53.2%	56	53.3%
Cognitive	49	31%	31	29.5%
Two outcomes combined	13	8.2%	9	8.7%
Three outcomes combined	2	1.3%	2	1.9%

One hundred fifty-eight effect sizes were calculated, representing each engagement predictor and outcome for the 79 included studies. The effect sizes ranged from $-.75$ to 2.51 , with the majority falling between $-.75$ and 1.8 . Positive effect sizes were most numerous ($n = 124$), though there were 33 negative effect sizes, and one effect size of zero.

Moderators of Engagement. A meta-regression was conducted for seven of the 12 coded moderators—five provided a minimum of ten point estimates for each moderator category, and two provided ten point estimates for most categories (see Table 3). Omnibus tests revealed statistically significant results for four of these seven moderators—geographic location, school setting, instrument reliability, and publication status.

Point estimates from studies sampling from countries outside the U.S. ($g = .42, 95\% \text{ CI } [.04, .49]$) showed the higher effect size while sampling U.S. schools showed the lower effect size ($g = .24, 95\% \text{ CI } [.16, .31]$). An examination of regression coefficients for the geographic location model showed that studies sampling schools outside the U.S. predicted increases in engagement point estimates ($\beta = .18, p = .0008$) when compared to studies sampled from schools within the United States ($\beta = .24, p < .00001$). However, 18 of the 44 studies from countries outside the United States originated from Turkey, where a K-8 school structure is common. The mean science engagement effect size for point estimates from middle schools was $g = .16, 95\% \text{ CI } [.06, .25]$, and from K-8 schools was $g = .42, 95\% \text{ CI } [.31, .52]$. These results suggest that the observed differences in science engagement due to geographic location might also be confounded by school structure.

Table 3: *Summary of Effect Sizes and Regression Models for Moderators*

Moderators	Point estimate categories			Regression model			
	n	Significant (n)	Minimum of 10 (n)	Q	df	p	Significant coeff. (n)
Publication status	2	2	2	15.70	1	.0007	2
Geographic location	2	2	2	11.28	1	.0007	2
School setting (w rural)	5	4	4	14.81	4	.0051	3
Instrument reliability	5	4	4	13.65	4	.008	0
School setting (no rural)	4	4	4	7.17	3	.067	2
Study methodology	4	3	4	6.41	3	.093	1

Peer review status	2	2	2	2.38	1	.123	1
Repeat authors	2	2	2	.03	1	.873	1
School structure	7	6	4	-	-	-	-
School type	4	3	3	-	-	-	-
Instrument validity	5	3	4	-	-	-	-
Socioeconomic status	5	3	3	-	-	-	-

School setting was reported for fewer than half of the point estimates ($n = 75$) within the study. Of those 75 point estimates, 18 reflected a mix of school settings (e.g., rural and suburban), and thus could be analyzed no further with respect to the effect of school setting on science engagement. Of the remaining 58 point estimates, those from urban schools reflected the highest effect size ($g = .40$, 95% CI [.25, .54]), and rural schools reflected the lowest effect size ($g = -.11$, 95% CI [-.42, .21]). Though the effect size for rural schools was not significant, the coefficient for rural schools was significant in the meta-regression ($\beta = -.50$, $p = .003$). This suggests that science engagement is expected to be lower in rural settings than in suburban or urban settings. However, an analysis of the lower mean science engagement effect size in rural schools was conducted with caution, as there were only five point estimates originating from schools in rural settings.

Instrument reliability was reported for all but six point estimates within the study. Point estimates from studies referencing an external instrument produced the highest mean effect size ($g = .60$, 95% CI [.39, .81]), followed closely by point estimates from studies referencing external instrument reliabilities ($g = .58$, 95% CI [.37, .78]). Though the effect sizes for both categories were statistically significant, the coefficients for each category within the regression model were not ($\beta = .33$, $p = .078$, and $\beta = .31$, $p = .099$, respectively). Point estimates from studies providing measures of internal reliability produced lower mean effect sizes, regardless of whether the internal measure was less or greater than .70 ($g = .26$, 95% CI [.12, .39], and $g = .30$, 95% CI [.22, .37], respectively). Neither coefficient was statistically significant in the regression model ($\beta = -.01$, $p = .965$, and $\beta = .03$, $p = .841$, respectively).

Point estimates from published studies showed the higher effect size ($g = .40$, 95% CI [.33, .46]), while those from unpublished studies showed the lower effect size ($g = .15$, 95% CI [.04, .25]). The regression coefficients for the publication status model showed that published studies predicted increases in engagement point estimates ($\beta = .25$, $p = .00007$) when compared to unpublished studies ($\beta = .15$, $p = .007$). Though these results suggest possible publication bias, no other analysis supported that conclusion (see Publication Bias).

Practically Significant Predictors. Fifty-one practically significant effect sizes ($g > .41$) represented 32.3% of the 158 point estimates and 46.8% ($n = 37$) of included studies (see Figure 1). Thirteen of 51 practically significant effect sizes reflected moderate effects ($g > 1.15$), and two had effect sizes approaching classification as strong—a science-technology-society curriculum approach ($g = 2.5$, 95% CI [2.079, 2.947]) and project-based learning ($g = 2.5$, 95% CI [1.954, 2.953]). The remaining 11 moderate effect size point estimates reflected a variety of predictors, including different instructional approaches (project-based learning, research, and scaffolding), self-determination theory components (autonomy and competence), and class characteristics (student-teacher relationship and perception of class goals).

Commonalities in Practically Significant Predictors. The distribution of engagement effect sizes for each predictor type was examined (see Table 4). Instructional method predictors had the highest frequency of practically significant effect sizes ($n = 24$; 46%), the highest frequency of

moderate effect sizes ($n = 7, 12.8\%$) and the lowest frequency of negative effect sizes ($n = 9, 15.8\%$). Though the other three categories of predictor types (technology, class characteristics, and social characteristics) yielded comparable frequencies of practically significant effects (26.7%, 28.3%, and 23%, respectively), technology had the highest frequency of negative effect sizes ($n = 5, 33.3\%$). Further, there were no practically significant technology point estimates that represented moderate effects ($g > 1.15$).

Table 4: *Distribution of Point Estimates by Predictor Classification*

Predictor Classification	Practically Significant Effect Sizes				Practically Insignificant Effect Sizes			
	Moderate ($2.7 > g > 1.15$)		Small ($1.15 > g > .41$)		Small ($.41 > g \geq 0$)		Negative ($g < 0$)	
	<i>n</i>	Percent	<i>n</i>	Percent	<i>n</i>	Percent	<i>n</i>	Percent
Type								
Instructional method	7	12.8%	17	29.8%	24	42.1%	9	15.8%
Technology	0	0%	4	26.7%	6	40%	5	33.3%
Class characteristics	5	8.3%	12	20%	33	55%	10	16.7%
Social characteristics	1	3.8%	5	19.2%	15	57.7%	5	19.2%
Self-determination theory								
Autonomy	5	5.3%	17	18%	51	54%	21	22.2%
Competence	4	13.8%	13	44.8%	11	37.9%	1	3.4%
Relatedness	4	11.4%	7	20%	16	45.7%	8	22.9%

Mean effect sizes were calculated for each category of predictor. Instructional method predictors showed the highest effect size ($g = .42, 95\% \text{ CI } [.34, .51]$), followed by class characteristics ($g = .34, 95\% \text{ CI } [.25, .42]$), and social characteristics ($g = .25, 95\% \text{ CI } [.12, .38]$). For technology predictors ($g = .10, 95\% \text{ CI } [-.06, .27]$), it was possible that the effect size was zero ($Z = 1.23, p = .22$). Only the mean effect size for instructional methods predictors achieved a minimum practical effect size of $g > .41$. See Table 5 for effect sizes and null tests of each predictor.

Table 5: *Effect Sizes and Null Tests for Predictor Classification: Type*

Predictor type	<i>n</i>	<i>g</i>	95% CI	<i>Z</i>	<i>p</i>
Instructional methods	57	.42	[.34, .51]	9.82	.0000
Technology	15	.10	[-.06, .27]	1.23	.2201
Class characteristics	60	.34	[.25, .42]	7.87	.0000
Social characteristics	26	.25	[.12, .38]	3.72	.0002

A test of the predictor type regression model reveals that it was likely effect size differed by predictor type ($Q = 13.56, p = .004$). The predictor type model explained 5% of the total between-studies variance in effect sizes ($R^2 = .05$). An examination of the regression coefficients for the model suggested that technology, class, and social predictors predicted decreased engagement point estimates when compared to instructional methods. However, only the coefficients for technology ($\beta = -.32, p = .0006$) and social characteristics ($\beta = -.18, p = .027$) were statistically significant (see Table 6). Though the null test of technology ($Z = 1.23, p = .2201$) indicated that the mean effect size point estimate for technology predictors on could be zero, the regression model suggested the impact of technology predictors on the model was significant.

Competence was the self-determination theory predictor with the highest frequency of practically significant effect sizes ($n = 17, 58.6\%$), the highest frequency of moderate effect sizes ($n = 4,$

13.8%), and lowest frequency of negative effect sizes ($n = 1$, 3.4%). Autonomy and relatedness yielded similar frequencies of practically significant point estimates ($n = 22$, 23.3% and $n = 11$, 31.4%, respectively) and negative point estimates ($n = 21$, 22.2% and $n = 8$, 22.9%, respectively).

Table 6: *Meta-regression Model for Predictor Classification: Type*

Predictor type	Variance		Test of model			Regression		
	T^2	R^2	Q	df	p	Coeff.	Z	p
Instructional method (intercept)	.0898	0				.42	9.82	.000
Technology	.09	0	7.92	1	.005	-.32	-3.40	.0006
Class characteristics	.0888	.01	8.42	2	.015	-.09	-1.44	0.149
Social characteristics	.0857	.05	13.56	3	.004	-.18	-2.22	0.026

Mean effect sizes were calculated for each category of SDT predictor. Competence showed the highest effect size ($g = .56$, 95% CI [.44, .69]), and autonomy showing the lowest effect size ($g = .26$, 95% CI [.19, .33]). All of the SDT predictors were statistically significant. See Table 7 for effect sizes and null tests of each predictor.

Table 7: *Effect Sizes and Null Tests for Predictor Classification: SDT*

SDT predictor type	n	G	95% CI	Z	p
Autonomy	94	.26	[.19, .33]	7.31	.0000
Competence	29	.56	[.44, .69]	8.90	.0000
Relatedness	35	.34	[.22, .46]	5.74	.0000

Though it was likely that the effect size differed by SDT predictor type ($Q = 17.80$, $p = .0001$), the model explained a negligible amount of the between-studies variance in effect sizes ($R^2 < .001$). An examination of the incremental changes to the model suggested that a model with just autonomy and competence explained 6% of the variance in effect sizes ($R^2 = .06$).

An examination of the regression coefficients for the model suggested that each SDT component predicted increased engagement (see Table 8). Furthermore, the coefficient for competence was statistically significant ($\beta = .31$, $p = .00002$) when compared to the intercept for autonomy. Though relatedness predicted increased engagement ($\beta = .08$), it was possible that the effect of relatedness predictors on engagement could be zero ($Z = 1.18$, $p = .236$).

Table 8: *Meta-regression Model for Predictor Classification: SDT*

SDT predictor type	Variance		Test of model			Regression		
	T^2	R^2	Q	df	p	Coeff.	Z	p
Autonomy (intercept)	.0898	0				.26	7.31	.000
Competence	.0840	0.06	18.13	1	.0000	.31	4.22	.000
Relatedness	.0950	0.00	17.80	2	.0001	.08	1.18	.236

Predictors of Engagement Types

Of 84 affective engagement point estimates, 28 were practically significant ($g > .41$). Of the predictor types, class characteristics and instructional methods showed the highest affective engagement effect sizes ($g = .42$, 95% CI [.30, .53], and $g = .38$, 95% CI [.28, .48], respectively). Similar to the holistic engagement results, technology showed the lowest effect size ($g = .09$, 95% CI [-.08, .25], and was not statistically significant. The regression model for predictor type on affective engagement explained 13.2% of the between-studies variance in affective engagement effect sizes

($R^2 = .132$), though only the instructional methods coefficient was statistically significant, as the effects of class characteristics and instructional methods were so similar.

Of the self-determination theory predictors, competence yielded the highest affective engagement mean point estimate ($g = .53$, 95% CI [.34, .71]), followed by relatedness ($g = .35$, 95% CI [.21, .50]), and autonomy ($g = .27$, 95% CI [.17, .37]). However, it was unlikely that affective engagement differed by SDT predictor type ($Q = 4.49$, $p = .06$), and the regression model explained a negligible amount of the variance in affective engagement ($R^2 < .001$).

The results for cognitive engagement paralleled those of engagement overall, with instructional methods and competence showing the highest mean cognitive point estimates ($g = .49$, 95% CI [.33, .66]) and ($g = .61$, 95% CI [.41, .81], respectively). Both predictor type and the SDT predictor models explained negligible variance between studies ($R^2 < .001$). With only 10 behavioral engagement point estimates, it was inadvisable to analyze mean effect sizes by category or through meta-regression.

Publication Bias. A comparison of unpublished ($g = .15$, 95% CI [.04, .25] $n = 39$) and published studies ($g = .40$, 95% CI [.33, .46], $n = 119$) warranted an examination of potential public bias. However, other suggested analyses did not find evidence for publication bias in this study. Though the regression model for publication status was statistically significant, it explained a negligible portion of the effect size variance ($R^2 < .0001$). The funnel plot revealed studies missing to the *right*, rather than the left of the mean. The adjusted mean effect size produced through a trim and fill procedure ($g = .42$, 95% CI [.35, .48]) was *larger* than the original ($g = .37$, 95% CI [.30, .42]). Last, Orwin's classic fail-safe N indicated that 9197 studies would be required to bring the mean Hedges' g to a value that would no longer be statistically significant.

Discussion

Of the predictor types, instructional methods were the best predictors of engagement. Though technology, class characteristics, and social characteristics all generated positive mean effect sizes, they also predicted decreases in science engagement in the regression model, with respect to instructional methods. Technology predicted the greatest decreases in engagement and had the highest representation of negative point estimates of all of the predictors ($n = 5$, 33%). Class characteristics and social characteristics predicted smaller decreases ($\beta = -.09$, $p = .149$, and $\beta = -.18$, $p = .026$, respectively), and the predicted decrease for class characteristics was not statistically significant.

Though causality was not established by this study, these results suggest that interventions focusing on technology, class characteristics, and social characteristics could be less effective at increasing science engagement than interventions focusing on instructional methods. The fact that technology predictors showed the lowest mean effect size and predicted the greatest decrease in engagement with respect to instructional methods runs counter to rationales given for technology integration in science classrooms—authenticity with the scientific discipline, equity, novelty, and autonomy support (Guillén-Nieto & Aleson-Carbonell, 2012; Zucker, Tinker, Staudt, Mansfield, & Metcalf, 2008). A common rationale given for the incorporation of technology games into the curriculum is that students receive more immediate feedback on their progress in a gaming situation (Garris, Ahlers, & Driskell, 2002). One explanation for the disconnect between rationales for technology integration and the relationship of technology with engagement in this study is that technology is one of many conduits through which authenticity, equity, novelty, autonomy, and feedback can be enhanced. The mere integration of technology does not ensure that any of the aforementioned desired qualities are implemented, or implemented effectively.

The predicted decrease in engagement from social characteristics when compared to instructional methods is also contradictory to educational research. Examples of social characteristics within

this study included perceptions of teacher characteristics—approachability, social support, and strictness—as well as more holistic social characteristics, such as perceptions of belonging, cooperative learning, and respect for differences. Research supports the efficacy of social interventions such as cooperative learning (Slavin, Hurley, & Chamberlain, 2003). Further, extensive research on the middle school transition suggests that students report their teachers to be more controlling and less nurturing, and also that social comparison and competition increases (Eccles & Midgley, 1989; Eccles et al., 1993; Lepper, Corpus, & Iyengar, 2005; Roeser & Eccles, 1998). Thus, perceptions of social characteristics should predict students' engagement.

There are a number of possible explanations for the incongruity between the observed relationship of social characteristics with engagement in this study and other educational research findings. One is that the vast majority of social characteristics point estimates ($n = 22$) reflected correlations between perceptions of those characteristics and engagement; only four of the point estimates in this category involved an intervention. Thus, it is possible that a student could report being engaged, while also reporting that his or her teacher was not approachable—in a correlational study there is no reason for one to explain the other. While the social characteristics category reflected 26 point estimates, they originated from only ten studies. In fact, one study produced 10 of the 26 point estimates. Additionally, six of the 26 point estimates reflected predictors that would be expected to have a negative relationship with engagement: perceptions of the teacher as admonishing, strict, or dissatisfied. When considering these different explanations in concert, a more likely explanation for the incongruity between observed and expected relationships between social characteristics and students' science engagement is that there were not enough point estimates to draw a definitive conclusion.

The class characteristics category, which predicted a statistically nonsignificant decrease in engagement with respect to instructional methods, was comprised of a variety of predictors, such as relevance, critical voice, autonomy support, and democratic versus traditional environments. The duration of more abstract interventions such as autonomy support could impact their efficacy, with students experiencing some discord with the intervention at early stages, and becoming more comfortable and benefitting from such interventions over time. Alternately, the novelty of such interventions could cause positive initial effects, with decreases over time as the intervention becomes more routine. In studies with multiple measures of engagement over time, the investigator selected the most proximal measure of engagement to the intervention. Thus, it is possible that longer-duration measures of the relationship between class characteristics and science engagement could show higher or lower point estimates than the more proximal measures within this study.

To further complicate the analysis of predictor type classification, many instructional methods can incorporate aspects of technology, class characteristics, or social characteristics. For example, project-based learning (instructional method) can include cooperative learning (social characteristic), and/or relevance (class characteristic) components. Thus, while one can conclude that a broad focus on technology, class characteristics, and social characteristics predicts decreases in science engagement, one cannot conclude that instructional methods incorporating these other components would be less effective than instructional methods that do not. Because the instructional methods category is a broad one—encompassing varied predictors such as project-based learning, graphic organizers, and whole brain teaching—further analysis is needed to fully answer the research question about commonalities in practically significant science engagement predictors.

Despite research on the middle school transition that shows students report negative perceptions of their teachers as more controlling, and their classrooms as more heavily focused on social comparison (Eccles & Midgley, 1989; Eccles et al., 1993; Lepper et al., 2005; Roeser & Eccles, 1998), competence was the best SDT predictor of increased science engagement over autonomy and relatedness. This finding is not entirely unexpected, as another defining characteristic of the

middle school transition is an increased focus on academic content standards (Ryan & Patrick, 2001). Cognitive mismatches between science classroom tasks and the changing early adolescent brain were not a neglected component of students' self-reports of their middle school classrooms or their science classes (Anderman & Mueller, 2010; Mahatmya et al., 2012; Ryan & Patrick, 2001, Uekawa, Borman, & Lee, 2007).

The finding that competence yielded the greatest effect size of the SDT predictors could suggest that science engagement is fundamentally different than engagement in other content areas. This premise is supported by Deci and Ryan's (2002) assertion that the relative importance of one given self-determination theory need to another can change depending on classroom characteristics. Science engagement may benefit more from explicit attention to competence as the content becomes more complex during middle school than engagement benefits from attention to autonomy or relatedness concerns. In other words, a perceived competence deficit could be a bigger problem than a perceived autonomy or relatedness problem. Though autonomy and relatedness may be the most prevalent unmet needs of early adolescents in science classrooms, competence predictors could be most effective at meeting those autonomy and relatedness needs. The relationship among autonomy, competence, and relationship is iterative; students' emotions related to perceived competence with a task can serve to increase or decrease their sense of autonomy and relatedness. Competence predictors could be more effective at increasing engagement in the early stages of engagement interventions.

Though instructional methods and competence produced the highest mean effect sizes, both predictor type and SDT predictor type regression models left a large amount of engagement variance unexplained. This finding parallels research that suggests only a small portion of engagement variance was explained by teacher and class-level variables, with the majority of variance occurring between and within individuals (Uekawa et al., 2007). Though this study examined classroom and task level science engagement predictors, it did not capture between individual and within individual variance.

Conclusion

Though much of the literature concerning early adolescents' perceptions about the middle school transition suggests that autonomy and relatedness are the most prevalent unmet needs, the results of this study suggest that academic predictors, such as instructional methods and competence, were more effective predictors of science engagement. Though these results are somewhat unintuitive, they do not fundamentally contradict interpretations through the lens of SEF theory or SDT. A lack of engagement indicates a mismatch between a learner's needs and the classroom environment.

Comprehensive instruments, or collections of instruments representing all three facets of engagement, should be utilized to examine trajectories of engagement for individual students. This recommendation is supported by the finding that within or between person variables explained more engagement variance than classroom or teacher-level variables (Lau & Roeser, 2008; Uekawa et al., 2007). The Experience Sampling Method (ESM) is a promising technique to examine these changes in student engagement. When self-reports of engagement through ESM are matched to the characteristics of tasks and activities occurring at the time of the self-reports, researchers can analyze nuanced changes in engagement for individuals. The Uekawa et al. (2007) study provides an exemplar of how students' self-reports of engagement, gathered through ESM, can be matched with temporally-immediate reports of class activities to produce a complete picture of students' changing engagement and possible antecedents of those changes.

Another benefit to assessing engagement longitudinally through ESM is the identification of possible engagement trajectories. Some research suggests that affective engagement is a precursor or regulator of other types of engagement (Ajzen, 1991; Ajzen & Fishbein, 1977, p. 888; Eccles & Wang, 2012; Pekrun & Linnenbrink-Garcia, 2013; Schank, 1979). Other researchers suggest that

cognitive and affective engagement predict behavioral changes (Reschly & Christenson, 2006). The use of ESM could afford the kind of detailed observation necessary to elucidate temporal changes and trajectories of engagement changes. Such information could inform decisions of which engagement types are appropriate targets in initial engagement interventions, compared with interventions that would better be targeted later in the sequence.

Another recommendation is to purposefully sample disengaged students in order to determine what practices *change* engagement for those students. In other words, though the results from this study may indicate that certain predictors have a more positive relationship with engagement than others, the study cannot inform conclusions about which predictors show the largest changes in engagement, nor can the study inform conclusions about which predictors show the largest changes in engagement for specific groups. As an implicit purpose of this study was to identify practices that engage or re-engage students with science coursework, an analysis of predictors that improve engagement for disengaged students is critical to inform best engagement practices in science classrooms.

The results from this meta-analysis suggest the inclusion of certain predictors in future studies. Categories that predicted the largest mean engagement effects included instructional methods, class characteristics, and competence. The finding that instructional methods best predict science engagement bears further examination. Do some instructional methods work better for disengaged students? Does the order in which instructional method interventions are implemented matter? What types of instructional methods work best? Similar questions emerge for class characteristics and competence predictors. Further analyses of effective engagement predictors will also be enhanced by the aforementioned use of longitudinal methods and purposeful sampling.

Though effective predictors of early adolescents' science engagement were identified in this study, it would be premature to eliminate less effective predictor categories from consideration in future science engagement studies. For example, though technology predicted a statistically significant decrease in engagement, the mean effect of technology on each engagement type was positive, and there were limited numbers of technology point estimates. Thus, the results of this study might inform hypotheses about expected results in future studies, but would not be cause for exclusion of particular predictors. Simple models with only predictor type or predictor SDT type did not predict a great deal of engagement variance, and there were also four statistically significant moderators of engagement—publication status, instrument reliability, school setting, and geographic location. These variables deserve further elucidation before definitive conclusions about predictors worthy of inclusion in future studies can be made.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), pp. 179-211. URL: [http://dx.doi.org/10.1016/0749-5978\(91\)90020-T](http://dx.doi.org/10.1016/0749-5978(91)90020-T)
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84(5). URL: <http://dx.doi.org/10.1037/0033-2909.84.5.888>
- Aker, L., B. (2016) A Meta-Analysis of Middle School Science Engagement (Doctoral dissertation). Retrieved from Dissertation Abstracts Database (Order No. 10157330).
- Anderman, E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research*, 64(2), pp. 287-309. URL: <http://dx.doi.org/doi:10.3102/00346543064002287>
- Anderman, E. M., & Mueller, C. (2010). Middle school transitions and adolescent development. In J. L. Meece & J. S. Eccles (Eds.), *Handbook of research on schools, schooling, and human development* (pp. 198-215). New York, NY: Routledge.

- Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist*, 50(1), pp. 84-94. URL: <http://dx.doi.org/10.1080>
- Biostat. (2015). Comprehensive Meta-Analysis, Version 3. [Software] Englewood, NJ.
- Braund, M., & Driver, M. (2005). Pupils' perceptions of practical science in primary and secondary school: Implications for improving progression and continuity of learning. *Educational Research*, 47(1), pp. 77-91. URL: <http://dx.doi.org/10.1080/0013188042000337578>
- Caraway, K., & Tucker, C. M. (2003). Self-efficacy, goal orientation, and fear of failure as predictors of school engagement in high school students. *Psychology in the Schools*, 40(4). URL: <http://dx.doi.org/10.1002/pits.10092>
- Deci, E. L., & Ryan, R. M. (2002). *Handbook of self-determination research*. Rochester, NY: University Rochester Press.
- Doğan, U. (2014). Validity and reliability of student engagement scale. *Journal of Faculty of Education*, 3(2), pp. 390-403. URL: <http://dx.doi.org/10.14686/BUFEFAD.201428190>
- Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), pp. 455-463. URL: <http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x>
- Eccles, J. S., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. *Research on Motivation in Education*, 3, pp. 139-186.
- Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & MacIver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, 48(2). URL: <http://dx.doi.org/10.1037/0003-066X.48.2.90>
- Eccles, J. S., & Roeser, R. W. (2010). An ecological view of schools and development. In J. L. Meece, & J. S. Eccles (Eds.), *Handbook of research on schools, schooling, and human development* (pp. 6-21). New York, NY: Routledge.
- Eccles, J., & Wang, M. T. (2012). Part I commentary: So what is student engagement anyway? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 133-145). New York, NY: Springer. URL: http://dx.doi.org/10.1007/978-1-4614-2018-7_6
- Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97-131). New York, NY: Springer. URL: <http://dx.doi.org/10.1007/978>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), pp. 59-109. URL: <http://dx.doi.org/10.3102/00346543074001059>
- Fredricks, J., McColskey, W., Meli, J., Montrosse, B., Mordica, J., & Mooney, K. (2011). *Issues & Answers Report: Measuring student engagement in upper elementary through high school: A description of 21 instruments* (REL Report No. 098).
- Garris, R., Ahlers, R., Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4), pp. 441-467. URL: <http://dx.doi.org/10.1177/1046878102238607>
- Guillén-Nieto, V., & Aleson-Carbonell, M. (2012). Serious games and learning effectiveness: The case of It's a Deal! *Computers & Education*, 58, 1, pp. 435-448. URL: <http://dx.doi.org/10.1016/j.compedu.2011.07.015>
- Jenkins, E. W., & Pell, R. G. (2006). *The Relevance of Science Education Project (ROSE) in England: A summary of findings*. Leeds, UK: Centre for Studies in Science and Mathematics Education.

- Kumar, D. D. (1991). A meta-analysis of the relationship between science instruction and student engagement. *Educational Review*, 43(1), pp. 49-61. URL: <http://dx.doi.org/10.1080/0013191910430105>
- Lau, S., & Roeser, R. W. (2008). Cognitive abilities and motivational processes in science achievement and engagement: A person-centered analysis. *Learning and Individual Differences*, 18(4), pp. 497-504.
- Lee, O., & Anderson, C. W. (1993). Task engagement and conceptual change in middle school science classrooms. *American Educational Research Journal*, 30(3), pp. 585-610. URL: <http://dx.doi.org/10.3102/00028312030003585>
- Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: age differences and academic correlates. *Journal of Educational Psychology*, 97(2), pp. 184-196. URL: <http://dx.doi.org/10.1037/0022-0663.97.2.184>
- Mahatmya, D., Lohman, B. J., Matjasko, J. L., & Farb, A. F. (2012). Engagement across developmental periods. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 45-63). New York, NY: Springer. URL: http://dx.doi.org/10.1007/978-1-4614-2018-7_3
- Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, 37(1), pp. 153-184. URL: <http://dx.doi.org/10.3102/00028312037001153>
- Newmann, F. M. (1992). *Student engagement and achievement in American secondary schools*. New York, NY: Teachers College Press.
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), pp. 1049-1079. URL: <http://dx.doi.org/10.1080/0950069032000032199>
- Pekrun, R., & Linnenbrink-Garcia, L. (2013). Academic emotions and student engagement. In S. L. Christensen, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on school engagement*, pp. 259-282. New York, NY: Springer. URL: http://dx.doi.org/10.1007/978-1-4614-2018-7_12
- Piaget, J. (1972). *The psychology of intelligence*. Totowa, NJ: Littlefield.
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), pp. 33-40. URL: <http://dx.doi.org/10.1037/0022-0663.82.1.33>
- Potkin, P., and Hasni, A. (2014), Analysis of the decline in interest towards school science and technology from grades 5 through 11. *Journal of Science Education and Technology*. 23(6), pp. 784-802.
- Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement*, pp. 3-19) New York, NY: Springer.
- Roeser, R. W., & Eccles, J. S. (1998). Adolescents' perceptions of middle school: Relation to longitudinal changes in academic and psychological adjustment. *Journal of Research on Adolescence*, 8(1), pp. 123-158. URL: http://dx.doi.org/10.1207/s15327795jra0801_6
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68. URL: <http://dx.doi.org/10.1037/0003-066X.55.1.68>
- Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, 38(2), pp. 437-460. URL: <http://dx.doi.org/10.3102/00028312038002437>
- Schank, R. C. (1979). Interestingness: Controlling inferences. *Artificial Intelligence*, 12(3), pp. 273-297. URL: [http://dx.doi.org/10.1016/0004-3702\(79\)90009-2](http://dx.doi.org/10.1016/0004-3702(79)90009-2)

- Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, 50(1), pp. 1-13. URL: <http://dx.doi.org/10.1080/00461520.2014.1002924>
- Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research*, 95(6), pp. 323-332. URL: <http://dx.doi.org/10.1080/00220670209596607>
- Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 21-44). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-2018-7_2
- Slavin, R. E., Hurley, E. A., & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. *Handbook of psychology*, 3(9), pp. 177-198. URL: <http://dx.doi.org/10.1002/0471264385.wei0709>
- Steinberg, L., & Silverberg, S. (1986). The vicissitudes of autonomy in early adolescence. *Child Development*, 57(4), pp. 841-851.
- Uekawa, K., Borman, K., & Lee, R. (2007). Student engagement in US urban high school mathematics and science classrooms: Findings on social organization, race, and ethnicity. *The Urban Review*, 39(1), pp. 1-43. URL: <http://dx.doi.org/10.1007/s11256-006-0039-1>
- Vedder-Weiss, D., & Fortus, D. (2011). Adolescents' declining motivation to learn science: Inevitable or not? *Journal of Research in Science Teaching*, 48(2), pp. 199-216. URL: <http://dx.doi.org/10.1002/tea.20398>
- Veiga, F., Reeve, J., Wentzel, K., & Robu, V. (2014). Assessing students' engagement: A review of instruments with psychometric qualities. In F. Veiga (Coord.), *Students' engagement in school: International perspectives of psychology and education* (pp. 38-57). Lisboa, Portugal: Instituto de Educação da Universidade de Lisboa.
- Wang, M. T., & Holcombe, R. (2010). Adolescents' perceptions of school environment, engagement, and academic achievement in middle school. *American Educational Research Journal*, 47(3), pp. 633-662. URL: <http://dx.doi.org/10.3102/0002831209361209>
- Wang, M. T., Willett, J. B., & Eccles, J. S. (2011). The assessment of school engagement: Examining dimensionality and measurement invariance by gender and race/ethnicity. *Journal of School Psychology*, 49(4), pp. 465-480. URL: <http://dx.doi.org/10.1016/j.jsp.2011.04.001>
- Wilson, D. B. (2015). Practical meta-analysis effect size calculator [Web software]. Fairfax, VA: George Mason University.
- Zucker, A., Tinker, R., Staudt, C., Mansfield, A., & Metcalf, S. (2008). Learning science in grades 3-8 using probeware and computers: Findings from the TEEMSS II project. *Journal of Science Education and Technology*, 17, pp. 42-48. URL: <http://dx.doi.org/10.1007/s10956-007-9086-y>

About the Authors

Dr. Leanna B. Aker: City University of Seattle (USA); e-mail: laker@cityu.edu

Dr. Arthur K. Ellis: Prof., Director, Center for Global Curriculum Studies at Seattle Pacific University (USA); e-mail: aellis@spu.edu

