www.activationlab.org
info@activationlab.org

Measures Technical Brief

Engagement in Science Learning Activities version 3.2 Aug 2016

Overview

Description of the Construct

We conceptualize **engagement** as one's focus, participation, and persistence on a task (Carini, et al., 2006; Finn, Pannozzo, & Voelkl, 1995; Fredricks, et al, 2004; Fredricks, et al, 2011). Following the research literature, within this conceptualization, we envision three dimensions of engagement: (1) **behavioral** engagement focuses on whether learner behaviors are related to completing the task or are off task; (2) **cognitive** engagement focuses on whether thought processes and learner attention are directed towards meaningful processing of information involved in completing the task; and (3) **affective** engagement focuses on whether the emotions that occur as part of completing a task are positive and high arousal rather than negative and low arousal. Research suggests that a combination of these three facets of engagement support students to learn (Dorph, et. al., 2013; Fredricks, et al, 2004).

Intended Uses of the Instrument

The Engagement survey was written for use with 10-14 year-old respondents immediately after a science activity, whether in a class or in an informal learning context. Such contextualized and immediate use minimizes memory biases or inferences based on beliefs the learner has about themselves or the learning context. Accordingly, it should be used after a focused science activity rather than as a measure of general engagement over a series of activities. No particular assumptions are made about task structure (e.g., brief or extended, alone or collaborative, adult guided or student guided) other than there is a particular task that should have been completed. The *engagement survey* is used to measure an individual's behavioral, cognitive, and affective engagement. Our analysis of the internal structure of the instrument indicates that valid inferences can be made regarding the overall engagement (i.e., a combination of affective, behavioral, and cognitive engagement) during an activity using responses across all items. Equally valid inferences can be made for two sub-factors of the scale, specifically an affective score or a behavioral/cognitive score. Responses to the cognitive and behavioral co-occur so tightly that separating those scores is not typically meaningful. Due to the self-report nature, the survey is not intended for high-stakes decisions about students (e.g., pass/fail determination, selection of program participants) or programs. Instead, the instrument is intended for formative feedback and/or for research purposes.

Evidence of Reliability and Internal Structure

Analyses were based on a total sample of 2,600 6th and 8th grade students responding after completing a diverse range of in-class science activities. Both the raw (Cronbach's) and polychoric alpha coefficients were found to be good (.80 and .85, respectively). Further description of analytical procedures and results are available in the psychometric properties section of this report.

How to Score

The *engagement survey* has the highest model fit when scores are calculated using a bi-factor model within confirmatory factor analysis. This provides three scores, an overall engagement score, a score for affective engagement and a score for behavioral/cognitive engagement. Pragmatically, scores can be produced from simple averages of all items (all of which are based on a 4-point Likert scale, with reverse coding for four of the items) to give an overall engagement score, or for the sub-parts the sum of items for affective engagement or behavioral/cognitive engagement. In fact, simple averages appear to have stronger predictive validity than do factor scores.

Analytical Options

Once scores are generated for the scale, researches and evaluators may be interested in using these scores in various analyses. The average scores can be treated as continuous dependent or independent variables for *t*-tests, ANOVA, and regression-type analyses.

The Instrument
Engagement in Science

Item ID Number	Prompt	Sub-factor	Response Options and Coding
E01*	During this activity: I felt bored.	Affect	1=YES! 2=yes 3=no 4=NO!
E02	During this activity: I felt happy.	Affect	4=YES! 3=yes 2=no 1=NO!
E03	During this activity: I felt excited.	Affect	4=YES! 3=yes 2=no 1=NO!
E04*	During this activity: I was daydreaming a lot.	Cognitive	1=YES! 2=yes 3=no 4=NO!
E05	During this activity: I was focused on the things we were learning most of the time.	Cognitive	4=YES! 3=yes 2=no 1=NO!
E06	During this activity: Time went by quickly.	Behavior	4=YES! 3=yes 2=no 1=NO!
E07*	During this activity: I was busy doing other tasks.	Behavior	1=YES! 2=yes 3=no 4=NO!
E08*	During this activity: I talked to others about stuff not related to what we were learning.	Behavior	1=YES! 2=yes 3=no 4=NO!

*Item is reverse-coded.

Psychometric Properties
Classical test theory statistics (reliability and exploratory factor analysis) were utilized to determine the Fascination scale's unidimensionality. Analyses were based on a total sample of 2,611 youth from 6th and 8th grade science classrooms. The sample was reduced to 2,600 after excluding cases where more than 50% of the items (5) were invalid (i.e., omissions or inappropriate multiple selection).

Table 1: Engagement Results

	Alpha if Deleted		Bi-factor Model Loadings		
Items	Raw	Polychoric	Factor 1	Factor 2	Factor 3
E01	0.76	0.81	0.270	.	0.800
E02	0.77	0.82	0.637	.	0.636
E03	0.77	0.82	0.626	.	0.621
E04	0.78	0.82		0.382	0.628
E05	0.78	0.83	.	0.253	0.600
E06	0.80	0.84	0.183	.	0.496
E07	0.79	0.84	.	0.559	0.449
E08	0.80	0.84	.	0.648	0.359
Scale	0.80	0.85			

Reliability. Cronbach's alpha and the polychoric alpha are measures of internal consistency within a particular scale. The polychoric alpha accounts for the ordinal nature (e.g., Likert-scale) of the variables (Gadermann & Zumbo, 2012). A satisfactorily large alpha (i.e., >.80) implies that individuals responded sufficiently similarly across the items to produce a stable overall score. Both the raw (Cronbach's) and polychoric alpha coefficients using all eight of the Engagement items were found to be good (.80 and .85, respectively). All items contribute positively to the reliability of the scale, implying that all items contribute to the cohesiveness of the scale.

Exploratory Factor Analysis. Exploratory factor analysis is used to identify an underlying latent factor among the measured items in the scale. Adequate fit to a unidimensional model is determined by a satisfactory visual inspection of the scree plot, sufficiently large factor loadings on each item (>.30), and satisfactory fit statistics (RMSEA<.06, CFI>.90, TLI>.90) (e.g. Costello & Osborne, 2005; Hu & Bentler, 1999; Byrne, 2010).

The *Engagement scale* was first subject to exploratory factor analysis using a forced one-factor solution. The factor loadings were sufficiently large (>.30). However, in terms of fit statistics, the CFI and TLI resulting from the one-factor solution were below satisfactory (.881 and .833, respectively), and the RMSEA was found to be larger than the set conventions (.207). A two-factor solution indicating with two scales (Scale 1: E01, E02, E03, E06; Scale 2: E04, E05, E07, E08) showed satisfactory model fit (CFI=0.997, TLI=0.993, RMSEA=0.044). The first factor was comprised of items pertaining to *affective* indicators of engagement, while the second factor was comprised of items pertaining to *cognitive* or *behavioral* indicators of engagement. Correlation between the two factors was estimated at 0.380.

A series of follow-up analyses using structural equation modeling tested model fit based on a two-factor confirmatory factor analysis model and a bi-factor model. The model fit statistics were not satisfactory for the two-factor confirmatory factor analysis model (CFI=0.959, TLI=0.940, RMSEA=0.124), but were generally satisfactory for the bi-factor model (CFI=0.992, TLI=0.982, RMSEA=0.069). The factor loadings for the bi-factor model are shown in Table 1. Of the three factors present in the bi-factor model (i.e., *affective*, *cognitive/behavioral*, and *overall* engagement), the factor scores from the overall engagement scale (comprised of all eight items) was found to have the strongest positive association[1] with the four *Measuring Activation* scales (i.e., Fascination, Values, Competency Beliefs, and Sensemaking) related to science. Thus, the full eight-item engagement scale was selected for Rasch modeling since it was considered the most relevant for most analytical purposes.

Consideration was given to the option of combining of behavioral and cognitive items into a single factor rather than keeping those separate: We maintain that behavioral and cognitive engagement are conceptually two

[1] Association was tested via regression using a dataset that included two administrations of all four *Measuring Activation* scales (as a pre- and post-survey) and a survey including engagement items being administered between the two *Measuring Activation* administrations. Regression was performed in two ways: 1) engagement predicted by the pre-survey *Measuring Activation* scale, and 2) post-survey *Measuring Activation* scale predicted by engagement while including the pre-survey *Measuring Activation* scale as a covariate. Tests were performed on each *Measuring Activation* scale separately using the same engagement data while also using both factor scores generated from the bi-factor model, and average scores (performed in separate analyses). The full eight-item scale was consistently found to have the largest (and positive) significant regression coefficients (both standardized and unstandardized) compared to the *affective* and *cognitive/behavioral* engagement sub-factors. All analyses were conducted using Mplus (Muthén & Muthén, 1998).

separate aspects of engagement. However, when measured using the self-report survey data we describe herein, we are not able to disentangle these two aspects due to very high co-occurrences. As of now, it is not known if this entanglement is related to the age of the survey-taker or specific to particular kinds of science activities (i.e., tasks involving coordinated cognitive and behavioral elements).

Rasch Model Fit. The Rasch model is used to provide estimates of the "ability" of survey participants, and the "difficulty" of each of the items. In this case, "ability" pertains to the amount of the *Engagement* factor in each participant, and "difficulty" pertains to the hesitation to agree with or endorse the statements provided in each of the items. Thus, the Rasch model can account for the varying difficulties of the items in generating estimates of participant ability (something notably missing in factor analysis).

The eight *Engagement* items were fit to the partial credit Rasch model (Masters, 1982) using ConQuest and examined for significant deviations in the unweighted (outfit) and weighted (infit) mean square error statistics (Wu, Adams, Wilson, & Haldane, 2007). Infit and outfit levels between 0.6 and 1.4 are generally considered satisfactorily fitting the Rasch model for rating scales (Wright & Linacre, 1994). Another indicator of scale validity is the person-separation reliability statistic, which is used to determine the inter-item reliability of the construct (Wright & Stone, 1979). As with Cronbach's alpha, values of .80 and above are considered sufficient (Andrich, 1982).

The model fit statistics of the *Engagement* scale are shown in Table 2. Both infit and outfit statistics were satisfactory for all items. The person-separation reliability statistic was satisfactory (EAP/PV=.822).

A Wright map depicting the difficulty levels of thresholds for each of the items is shown on the following page. As depicted in the map, the item with the lowest threshold of moving from the lowest response option to the next lowest response option is "I was focused on the things we were learning most of the time (NO!, no, yes, YES!)" (E05). The most difficult threshold to endorse is to respond "YES!" to the item "I felt happy" (E02). Of note is that each item has correctly ordered thresholds meaning that moving from non-endorsement to endorsement of the item is indicative of moving higher on along the latent trait.

Table 2: Engagement Rasch Model Fit

	Unweighted		Weighted	
Item	MNSQ	*t*	MNSQ	*t*
E01	0.82	-6.9	0.82	-7.1
E02	0.88	-4.5	0.89	-4.5
E03	0.94	-2.0	0.94	-2.2
E04	0.99	-0.3	0.98	-0.8
E05	1.03	1.1	1.02	0.6
E06	1.25	8.4	1.21	7.4
E07	1.13	4.3	1.08	2.5
E08	1.23	7.7	1.19	6.9

Research reported in this publication was supported by the National Science Foundation under Award Number 1348666. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Science Foundation.

Suggested Citation:

Chung, J., Cannady, M. A., Schunn, C., Dorph, R., & Bathgate, M., (2016) Measures Technical Brief: Engagement in Science Learning Activities. Retrieved from: http://www.activationlab.org/wp-content/uploads/2016/02/Engagement-Report-3.1-20160331.pdf

```
================================================================================
ConQuest: Generalised Item Response Modelling Software     Tue Dec 02 14:42 2014
MAP OF LATENT DISTRIBUTIONS AND THRESHOLDS
================================================================================
                              Generalised-Item Thresholds
--------------------------------------------------------------------------------
                |
                |
   4            |
                |
                |
                |
                |
                |
                |
   3            |
                |
               X|
               X|
               X|
              XX|
              XX|
   2           X|
               X|E02.3
              XX|E03.3
             XXX|E01.3
              XX|
              XX|E05.3
            XXXX|
   1        XXXX|E08.3
            XXXX|E04.3  E06.3
           XXXXX|E07.3
          XXXXXX|
         XXXXXXX|E03.2
         XXXXXXX|
   0     XXXXXXX|
          XXXXXX|
         XXXXXXX|
        XXXXXXXX|
        XXXXXXXX|E02.2
      XXXXXXXXXX|E01.2
         XXXXXXX|E08.2
  -1      XXXXXX|E05.2  E06.2
            XXXX|
             XXX|E04.2
            XXXX|
              XX|
               X|E07.2
               X|E01.1
  -2           X|E03.1
               X|E04.1  E06.1
                |E08.1
                |E02.1
                |E07.1
                |
                |E05.1
  -3            |
                |
                |
                |
                |
                |
                |
  -4            |
                |
                |
================================================================================
Each 'X' represents  19.5 cases
The labels for thresholds show the levels of
    item, and step, respectively
```

==

References

Ames, C. (1992). Classrooms: Goals, structures and student motivation. *Journal of Educational Psychology*, 84(3), 261-271.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, 47(1), 105-113.

Baram-Tsabari, A., & Yarden, A. (2005). Characterizing children's spontaneous interests in science and technology. *International Journal of Science Education*, 27(7), 803-826.

Byrne, B. M. (2010). *Structural Equation Modeling with Amos: Basic Concepts, Applications, and Programming* (2nd ed.). New York: Routledge.

Cai, L., Thissen, D., & du Toit, S. (2011). *IRTPRO 2.1 for Windows* [computer software]. Chicago, IL: Scientific Software International.

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: Testing the linkages. *Research in higher education*, *47*(1), 1-32.

Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7).

Dorph R, Schunn C, Cannady M, Crowley K, Shields P, Science learning activation: Positioning youth for persistent success in science learning, literacy and careers. Structured Poster Session at the American Education Research Association (AERA) Annual Meeting, San Francisco, CA. Apr 2013.

Finn, J. D., Pannozzo, G. M., & Voelkl, K. E. (1995). Disruptive and inattentive-withdrawn behavior and achievement among fourth graders. *The Elementary School Journal*, 421-434.

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), 59-109.

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments. Issues & Answers. REL 2011-No. 098. *Regional Educational Laboratory Southeast*.

Gadermann, A. M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3).

Gardner, H. (1987). Comments on "Towards the development of a childen's science curiosity scale". *Journal of Research in Science Teaching*, 24(2), 175-176.

Girod, M. (2001). *Teaching 5th grade science for aesthetic understanding*. Michigan State University, East Lansing, MI.

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41(2), 111-127. doi: 10.1207/s15326985ep4102_4

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation Modeling*, 6(1), 1–55.

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science*, 326(5958), 1410-1412.

Kiefer, T., Robitzsch, A., Wu, M., & Robitzsch, M. A. (2013). *TAM: Test analysis modules* (Version 1.1) [computer software]. <http://cran.r-project.org/package=TA>.

Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education*, 29(7), 871–893.

Linacre, J. M. (2008) *Winsteps* (Version 3.67.0) [computer software]. Winsteps.com, Chicago.

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment*, 80, 75-86.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulleti*n, 116(1), 75-98.

Mair, P., & Hatzinger, R. (2007) Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software*, 20(9):1–20.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Muthén, L. K., & Muthén, B. O. (1998). *MPlus* (Version 7.11) [Computer Software]. Los Angeles: Muthén & Muthén.

Osborne, J. F., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049–1079.

Reid, N. (2006). Thoughts on attitude measurement. *Research in Science & Technological Education*, 24(1), 3-27.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8:3, 370.

Wright, B. D., & Stone, M. H. (1979). Best test design. *Rasch Measurement*. Chicago: Mesa Press.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest Version 2.0: Generalised Item Response Modelling Software* [computer software]. Melbourne: Australian Council for Educational Research.