Published in final edited form as:
Educ Technol Res Dev. 2020 December ; 68(6): 3143–3163. doi:10.1007/s11423-020-09813-1.

Exploring Indicators of Engagement in Online Learning as Applied to Adolescent Health Prevention: A Pilot Study of REAL media

Anne E. Ray, PhD[1], **Kathryn Greene, PhD**[2], **Teja Pristavec, PhD**[2], **Michael L. Hecht, PhD**[3], **Michelle A. Miller-Day, PhD**[3], **Smita C. Banerjee, PhD**[4]

[1]University of Kentucky

[2]Rutgers University

[3]REAL Prevention LLC

[4]Memorial Sloan Kettering Cancer Center

Abstract

Engagement is central to the effectiveness of online health messages and the related educational programs that aim to deliver these messages to the intended audience (Li, Won, Yang et al. 2019: Lin, Hung, Kinshuk et al. 2019). Drawing from health communication and social learning theories, the Theory of Active Involvement (TAI) (Greene, 2013) posits that an online prevention program's impact depends on how engaged participants are. In practice, measuring engagement in this context has relied primarily on self-report measures (e.g., Hamutoglu, Gemikonakli, Duman et al. 2019). However, the emergence and growth of online learning platforms to deliver health-specific information offers other options for assessing engagement. This includes program analytics that capture interaction with content and facilitate examination of patterns via multiple indicators such as responses to interactive questions and time spent in the program (Herodotou, Rienties, Boroowa, et al. 2019; Li, Wong, Yang et al. 2019; van Leeuwen, 2019). However, little is known about the relationships between these different indicators of engagement as it applies to health curricula. This study uses self-report, observational, and program analytic data collected on a small (N = 38) sample using REAL media, an online substance use prevention program, to examine relationships among various indicators of engagement. Findings suggest a cluster of indicators across the three modalities that provide a useful way of measuring engagement. A cluster centered around complexity suggests a separate factor to be considered when designing engaging interventions.

Correspondence should be addressed to Anne E. Ray (anne.ray@uky.edu).

Conflicts of Interest: Kathryn Greene, Smita Banerjee, Michael Hecht, and Michelle Miller-Day disclose intellectual property interests in the REAL media curriculum.
Human rights: All procedures involving human participants were in accordance with institutional review board ethical standards and with the 1964 Helsinki declaration and its later amendments. IRB protocol approval #15–544Rc (Rutgers University).
Informed Consent: Informed consent from parents and youth assent was obtained from all study participants.

Keywords

engagement; e-learning; health messages; substance use; media literacy; complexity

Much like in more traditional educational contexts, engagement with online health prevention messages and related curricula is a central construct in health message design and evaluation (Dusenbury, Brannigan, Hansen, Walsh, & Falco, 2010; Low, Ryzin, Brown, Smith, & Haggerty, 2013; Tobler, Roona, Ochshorn, Marshall, Streke, & Stackpole, 2000). Rather than conceptualizing participants in prevention programs as passive recipients of content with fidelity in implementation the primary concern, emerging theoretical frameworks point to the audience's active role in the process (Pettigrew, Graham, Miller-Day, Hecht, Krieger, & Shin, 2015; Pettigrew & Hecht, 2015; Low et al., 2013; Weissberg & Pachan, 2010; Durlak & DuPre, 2008; Dusenbury et al., 2005; Lee & Hannafin, 2016). In other words, the audience "co-creates" the experience with the program designer and implementer (Pettigrew et al., 2015), and has the autonomy to make their own choices throughout the learning process (Lee & Hannafin, 2016). This is particularly true of digital messages and online learning programs, in which individuals interact and communicate with an online program (e.g., Henrie, Halverson, & Graham, 2015), and where research seeks to identify factors influencing uptake and adoption (Ma & Lee, 2018).

As a result, engagement is conceptualized as a key to effective digital messaging (Davis, Sridharan, Koepke, Singh, & Boiko, 2018; Dicheva, Dichev, Agre, & Angelova, 2015; Li, Won, Yang et al. 2019: Lin, Hung, Kinshuk et al. 2019) and is a key to message and program impact (Davis et al., 2018). The level of engagement may differ based on psychographics (Hidi & Renninger, 2006; Plass & Kinzer, 2015). In some cases, engagement is seen as so central to the process of effective messaging that is it identified as the key outcome (Davis et al., 2018; Hidi & Renninger, 2006, Girard & Magnan, 2012).

In this paper, we explore indicators of learner engagement as it applies to REAL media, an online health prevention curriculum, with a focus on how these indicators relate to one another. Given the growth of online learning generally, and within the field of health prevention more recently, understanding the interplay between indicators of engagement is an important and not well understood area of inquiry.

Conceptualizing Engagement

The term engagement has several meanings within and across disciplines. For example, in their study of online learning environments, Kahn et al. (2016) referenced a few conceptualizations of engagement: one, as a general idea of the effort and commitment put forth by learners, and two, as learners' own agency, or intentional action, related to learning. In an investigation of a curriculum focused on integrating technology in rural classrooms, Reynold and Caperton (2011) described engagement as an environment where learners work together, communicate with one another, and create a product. Oh, Bellur, and Sundar (2018) describe engagement as a continuum, with four components: physical interaction, interface assessment, absorption, and digital outreach. In this paper, we utilize the Theory of Active Involvement (TAI), a health message theory that conceptualizes engagement as central to the effectiveness of health prevention curricula (Greene 2013). We focus on the theory of TAI as our goal is to explore engagement as it relates REAL media, which utilizes an online learning format to deliver content to high-school aged youth that aims to prevent onset of and/or reduce use of alcohol, tobacco, and other drugs among youth (Ray et al., 2019).

TAI was developed to provide a conceptual framework to understand how health curricula that involve youth message planning and production related to substance use subsequently impacts both cognitive and behavioral outcomes of program participants (Greene, 2013). The theory posits that programs that include active involvement components (e.g., having youth create anti-substance messages) fosters youth engagement in curriculum, which in turn leads to increases in knowledge and skills. The knowledge and skills acquired by youth lead to a period of reflection on their own beliefs and behaviors, which then impacts substance-related cognitions (e.g., expectancies, normative beliefs, and intentions to use) and ultimately the target behavior itself (e.g., actual youth substance use). Notably, engagement is a critical construct and starting point in this cascading conceptual model.

According to TAI, engagement is marked by two linked components or processes that facilitate the recipient's information processing following program exposure: arousal and involvement. Arousal is indicated by one's perceived novelty of and attention to program activities, whereas involvement is indicated by one's interest in the topic, perception of utility or gain of program content, and reflectiveness related to one's own behavior. Drawing on social cognitive theory, TAI argues that there are both cognitive and behavioral elements to engagement. The cognitive element refers to how participants process information following arousal and involvement. That is, arousal and involvement provide the necessary condition for an individual to subsequently process the information received in the program. The behavioral element requires that participants actively perform some task that is part of the intervention strategy. These cognitive and behavioral elements are both central to the engagement process.

The REAL media Curriculum

REAL media is a substance abuse prevention program for youth developed based on a TAI approach to media literacy (Greene, Banerjee, Ray, & Hecht, 2017). REAL media aims to increase youth media literacy skills through increasing their awareness of advertising reach and costs, building knowledge of common sales techniques, and building skills related to counterarguing and critical thinking. The program culminates with skill application in the form of youth-created anti-substance prevention messages targeting peers. It is a self-paced, online learning curriculum adapted from an evidence-based curriculum designed for face-to-face delivery (Banerjee, Greene, Magsamen-Conrad, Elek, & Hecht, 2015; Greene, Catona, Elek, Magsamen-Conrad, Banerjee, & Hecht, 2016); this foundational curriculum has been independently rated "effective" by the Substance Abuse and Mental Health Services Administration's National Registry of Evidence-Based Programs (Banerjee et al., 2015; SAMHSA NREPP, 2017). REAL media was developed through a partnership model including the developers of the foundational program, a health message development company, and community partners. The name REAL media is consistent with the branding of the development company and details of the adaptation process are reported elsewhere (Ray et al., 2019). The program was implemented in 10 states through the 4-H organization, the latter which is undergoing a randomized control trial to evaluate program efficacy, and it was just released for national distribution through SHOP 4-H. The program is also distributed nationally through the D.A.R.E. program under the title REAL messages.

Following TAI, REAL media incorporates both cognitive and behavioral aspects of engagement. First, REAL media produces cognitive engagement by having participants perform both critical analysis of media messages as well as the planning of messages to stimulate systematic processing. TAI argues that these strategies encourage youth to more critically consider perspectives in both media and health promotion messages, as well as to contrast these messages with their own and their peers' perspectives. As a result, this reflexive process shapes the youth's cognition and behavior, including substance use behavior.

Second, REAL media engages youth behaviorally through message production and sharing processes. The efficacy of these strategies is supported by studies demonstrating that giving youth the opportunity to create anti-substance messages increases the likelihood that they will subsequently report factors that are predictive of lower substance use levels (SAMHSA NREPP, 2017; Banerjee & Greene, 2006; Banerjee & Greene, 2007; Hecht & Miller-Day, 2009; Greene, 2012; Greene, Carpenter, Banerjee, Magsamen-Conrad, Hecht, & Elek, 2015). After learning about the media and media influence strategies, youth produce and share via social media counter-messages, created during participation in REAL media. REAL media is designed to use TAI's cognitive and behavioral engagement strategies in the belief that changes in youth substance use behavior depend on the extent of youth's engagement with the intervention. Together, the analysis, planning, and production activities are designed to engage youth, thereby facilitating arousal and involvement processes which TAI argues is essential in leading to skill building, cognitive change, and ultimately, desired behavioral outcomes.

Measuring Engagement

Understanding and accurately measuring youth engagement is thus crucial for designing and evaluating online learning programs, particularly those derived from theories such as TAI, which posit that participants' active involvement serves a critical role in change. In the context of substance use, this goal is pressing given that youth who are most at-risk are also frequently dismissive of health education (Arnett, 1992), as well as the general finding that interventions actively involving the audience are more effective (Durlak & DuPre, 2010; Dusenbury et al., 2010; Tobler et al., 2000). Additionally, better understanding program engagement and its relationship with health outcomes is vitally important to advancing prevention science, and that requires valid and reliable measurement of engagement (Soffer & Nachmias, 2018). This means an integration of objective and subjective measures.

In practice, measuring engagement is challenging and has relied primarily on self-report measures. Some rely on teacher reports about implementation (e.g., Low et al., 2013), while others explicitly ask youth to rate and report their levels of engagement, solicit their opinion on the intervention program, or directly inquire about reasons for (non-) engagement (Hamutoglu, Gemikonakli, Duman et al. 2019; Mason, 2011; Tran, Nguyen, Nong, Maher, Nguyen, Nguyen, & Le, 2015). Given differences in how engagement is conceptualized within and across fields, self-report measures are advantageous in that they can be tailored to any given definition or framework. They are also potentially problematic due to social desirability bias (Lillehoj, Griffin, & Spoth, 2004) as well as participants' absence of comparative bases for ratings. In addition, it is difficult to measure the engagement of participants who do not, in fact, engage with the material (i.e., fail to complete program activities). Further, engagement is multidimensional and encompasses behavioral as well as psychological and cognitive factors (Appleton, Christenson, & Furlong, 2008; Finn & Zimmer, 2012; Fredricks, Blumenfeld, & Paris, 2004). Multiple facets of engagement may be difficult to capture when relying solely on participants' self-reports.

Another approach to measure engagement while mitigating potential social desirability is to observe participant behaviors. That is, rather than asking participants about their engagement directly, researchers may collect raw observational data and apply coding or rating schema to arrive at more abstract indicators of involvement. For example, the complexity of responses to program material may be one such abstract engagement indicator (Burleson & Waltman, 1988; O'Keefe & Sypher, 1981). For example, researchers may code or rate open-ended responses pertaining to an intervention content to measure the complexity of the response as a proxy for depth of involvement. Similarly, at the most basic level, it is possible to derive a word count or length of response to arrive at a potential engagement measure.

A third, less intrusive approach that has emerged in recent years using analytic information from e-learning platforms and online learning management systems (LMS) (Herodotou, Rienties, Boroowa, et al. 2019; Li, Wong, Yang et al. 2019; van Leeuwen, 2019). These platforms offer new opportunities for measuring engagement in online learning. LMS collect activity data as individuals progress through the program, and data from these user interactions and behavior can provide new insights into learning and involvement (Macfadyen & Dawson, 2012). For example, several studies investigated the relationship between online behavior and youth achievement using analytic or computer log data (Morris, Finnegan, & Wu, 2005; Rafaeli & David, 1997; Peled & Rashty, 1999; Zaiane & Luo, 2001). Researchers interested in identifying youth at risk of poor academic performance analyzed online LMS behavior and identified students who failed or succeeded in a course based on these interaction patterns (Mason, 2011; Macfadyen & Dawson, 2012). Others focused on online course material use, such as the frequency of accessing or viewing online study resources, and its relationship to student performance. For example, analyzing server data that included information on students' online resource use, studies found that students who accessed these materials subsequently achieved higher exam scores (Giesbers, Rienties, Tempelaar, & Gijselaers, 2013; Grabe & Christopherson, 2005; Grabe & Christopherson, 2008; Stewart, Stott, & Nuttall, 2011). Some studies employ program analytic data such as time, or the duration of program use, to gauge engagement; knowledge acquisition and processing require time, and more time spent on the program may indicate greater engagement with its content (De Boer & Collis, 2005; Kolari, Savander-Ranne, & Viskari, 2008; Van den Brande, 1994). Longer time spent on the program or using a greater number of online resources may also suggest persistence and greater motivation to learn

(Macfadyen & Dawson, 2012; Dev, 1997). A different argument, however, could be made that longer time spent in program was indicative of less engagement because a user was more distracted and completing other tasks while using a program, but this perspective is less common. It could also be indicative of cognitive, or information, overload on the part of the user (e.g., Jong, 2010; Kohan et al., 2017; Mayer & Moreno, 2003). Further complicating the use of time, it is difficult for programs to assess when the user has stayed in the program (e.g., it is active on computer or mobile device) but turned to other tasks.

Indicators collected in LMS and through program analytics can therefore be conceptualized as indicators of program engagement. However, the relationship between measures such as time spent on learning tasks or LMS interaction time with subsequent performance and traditionally collected self-reports of engagement is unclear. Although some studies lend support to the assertion that program analytic factors are associated with later performance, others suggest a weak or negative correlation between indicators such as time spent on tasks and self-reported engagement or learning (Allen. Lerner, & Hinrichsen, 1972; Greenwald & Gilmore, 1997; Wagner, Schober, & Spiel, 2008; Wagstaff & Mahmoudi, 1976; Wellman & Marcinkiewicz, 2004).

Current Study

The goal of the current study is to explore relationships between traditional, theory-driven self-report measures of engagement in relation to observational and analytic data collected during a pilot study of REAL media. Understanding how these different approaches to measuring engagement relate to each other allows for a better understanding of how best to evaluate the impact of similar health prevention curricula moving forward. Additionally, an in-depth understanding of engagement advances prevention science through identifying the mechanisms underlying health behavior change rather than focusing on amorphous constructs such as “attitudes” “persuasive effectiveness”, and “intentions” that may correlate with behavior but tell us very little how to create behavior change.

Consistent with our conceptualization of engagement, TAI guides the self-report measurement. Greene and colleagues (2015) developed the Audience Engagement Scale (AES) to capture the cognitive processes (i.e., arousal and involvement) that are hypothesized to underlie engagement. More specifically, the AES captures one’s personal reflection, critical thinking, and perceived novelty as related to the curriculum. In addition, we also include self-report measures of usability (e.g., Brooke, 1996, Nokelainen, 2006). Usability is generally meant to captures the user’s satisfaction (e.g., effectiveness, ease of use) with a given system such as an online learning program. Although such measures were developed with the goal of assessing one’s satisfaction with a given system, there is arguably some overlap with one’s attention and involvement as described by TAI. Further, usability measures are commonly implemented when evaluating online programs, thus it is worth exploring the extent to which these often used self-report scales correlate with the more nuanced AES, which is tailored to TAI. Given that TAI was developed to correspond to face-to-face curricula, prior measurement papers on this theory do not incorporate observational and analytic measures that are readily available given the online format of REAL media. As discussed earlier, it is possible that observational measures like word count serve as indicators of involvement. For example, responses with longer word count and/or one's choice to click on optional content could be reflective of increased personal interest or reflectiveness. If that is the case, we would expect to see strong correlations between self-report, observational, and analytic measures. It is also possible the available observational and analytic measures have less overlap with self-report measures, which would suggest they may reflect another dimension of engagement or construct entirely (e.g., these indicators could be reflective of cognitive ability or overload). Accordingly, we address the following research question: What are the relationships among self-report, observational, and analytic measures of engagement in relation to REAL media, a youth media literacy substance prevention program delivered via an online learning platform?

Method

Sample

Participants were recruited among New Jersey 4-H youth club members in Fall 2015 to participate in a pilot study of the REAL media program. 4-H clubs are part of a national, United States Department of Agriculture (USDA)-sponsored network of youth organizations using positive development and experiential learning to cultivate youth citizenship, leadership, responsibility, and life skills. 4-H members have substance use patterns that mirror the general population despite the protective influence of their club involvement (Lerner & Lerner, 2013). Participants provided assent after research staff obtained parental consent and were compensated with a $30 gift card on completion of the single research session. A university Institutional Review Board approved the study procedures.

The final analytic sample includes 38 4-H youth (35% male, 65% female) who were between 14 and 17 years old ($M = 14.85$, $SD = 0.97$) at the time of the study. Youth identified as being European-American or white (n = 14), African-American or Black (n = 4), Asian or Pacific Islander (n = 2), Hispanic or Latino (n = 3), American Indian or Alaskan Native (n = 1), or some other ethnicity (n = 2). The clubs to which they belonged were in urban areas or large cities (n = 9), suburban areas or near a large city (n = 8), smaller cities (n = 2), small towns (n = 6), or in rural areas (n = 1). On average, participating youth have been involved with 4-H for 4.2 years ($SD = 3.38$, min = 0.7, max = 10.5).

Intervention: REAL media

REAL media is a self-paced, online 90-minute curriculum designed to decrease substance use in adolescents. It consists of 5 levels or lessons covering topics such as media reach, media ethics, influence strategies, advertising claims and evidence, and production techniques, consistent with the overall aim of the program as described above. Level 1 introduces concepts of media reach and cost, as well as media ethics. Level 2 focuses on target audience and persuasion strategies used in advertising. Level 3 identifies arguments or claims used in advertisements as well as missing information and counter-arguing. Level 4 focuses on attention-getting tactics and major production techniques used in advertising. In the 5th and final level, youth plan, produce and share a counter-message (i.e., drug prevention message) targeting peers. In addition to including informational content (e.g., definitions of key concepts), the levels are designed to be highly interactive, using "drag and drop," multiple choice, sliders, fill-in the blank, hover and reveal, and other engaging features that require users to make decisions or participate in the program. Through these interactive features afforded by the e-learning platform, youth are asked to demonstrate their understanding of concepts as well as apply skills. For example, to elucidate the concept of counter-arguing in level 3, users are shown ads and asked to identify counterarguments. Given the difficulty of this concept, they are first presented with multiple choice questions. Later in the level, they are asked to demonstrate this on their own via an open-ended response box. In addition, each level has at least one "optional depth" feature, which allows participants to explore a topic in greater detail, as well as a final "challenge" that tests their knowledge of key program constructs. For purposes of this study, youth did not produce or share messages (Level 5) given the restrictions of time and resources. This study is based on data collected from youth using the REAL media program as a part of a pilot study that followed program development.

Procedures

Data were collected during two to two half-hour-sessions in which 4-H youth navigated the REAL media program and were asked to provide feedback. Research staff provided MAC and PC computers and headphones for program narration, with computers spaced at the location to ensure some privacy. Staff provided technical assistance with the program as needed. Data were collected at 4-H meeting locations known to participants. Youth progressed through the program levels at their own pace and took breaks as needed (snacks provided).

Measures

We collected self-report, observational, and program analytic data related to the program. As participants progress through REAL media modules, they provided self-reported engagement ratings of the program overall and by level while the LMS logged a number of participant interactions, as well as completed self-reported measures of program usability. Participants also responded to open-ended questions and situations through textboxes (e.g., describe your ethical standards for using social media). These inclusions allowed us to collect observed and program analytic data, in addition to the self-report measures, and thus examine the relationships between these indicators of engagement with the REAL media program. These measures are described in detail next.

Self-reported engagement measures.—The project uses three self-reported engagement measures. The first, the Audience Engagement Scale (AES, Greene et al. 2015), measured perceived engagement with each program level (1–4), while the other two, the Pedagogically Meaningful Learning Quality (PMLQ; Nokelainen, 2006), and System Usability Scales (SUS; Brooke, 1996) provide two measures of overall program evaluations (i.e. after completion of the program).

Audience Engagement Scale (AES) is a multi-dimensional construct that measures self-reported audience engagement in each level of the program Greene et al. 2015. Each dimension—perceived novelty, personal reflection, and critical thinking—was measured after a participant completed a level (i.e., prior to moving to next level) using two agree-disagree items with response categories ranging from 1 (“*strongly disagree*”) to 5 (“*strongly agree*”).

Perceived novelty indicates the youth’s perception of the newness or originality in the curriculum (“This level was different from regular school classes” and “This level was just like what we normally do in school”). It is reasonable to expect that novelty would be an indicator of engagement for adolescents because it compares how a program is like/different from usual educational experiences encountered.

Personal reflection indicates the youth’s perception of the degree to which program content lead them to evaluate their own personal behavior. This indicates engagement because it means that they use knowledge gained through the curriculum to reevaluate their actions and behaviors (“This level made me think a lot about the impact of advertising on me” and “This level made me think a lot about my substance use (drugs, alcohol, tobacco)”.

Critical thinking indicates the youth’s perception of the degree to which the curriculum encouraged their critical media message evaluations (“This level made me think about the ads that I see” and “This level made me think about the truthfulness of ad claims).

Negatively worded items were reverse-coded for directionality and items for each dimension were averaged, with higher scores indicating higher *perceived novelty* (L1 $M = 3.9$, $SD = 0.87$, range = 2–5; L2 $M = 4.1$, $SD = 0.82$, range = 2.5–5; L3 $M = 3.4$, $SD = 1.04$, range = 2–5; L4 $M = 3.9$ $SD = 0.88$, range = 2–5), *reflection* (L1 $M = 3.3$, $SD = 0.87$, range = 1.5–5; L2 $M = 2.99$, $SD = 0.84$, range = 1.5–4; L3 $M = 3.3$, $SD = 0.94$, range = 1.5–5; L4 $M = 3.2$, $SD = 0.80$, range = 1–5), or *critical thinking* (L1 $M = 4.2$, $SD = 0.70$, range = 2.5–5; L2 $M = 3.9$, $SD = 0.96$, range = 1–5; L3 $M = 4.3$, $SD = 0.83$, range = 1–5; L4 $M = 4.0$, $SD = 0.78$, range = 1.5–5), respectively. For analytic purposes, scores for overall AES and each subscale were calculated across levels to put them on the same level of analyses as other measures.

The overall AES scale demonstrated good reliability (Cronbach’s $\alpha = 0.86$) with each subscale also demonstrating good reliability, likely influenced by the inclusion of item pairs across levels 1–4 (Novelty $\alpha = 0.79$; Critical thinking $\alpha = 0.76$; Personal reflection $\alpha = 0.76$). Item pair correlations within level were more variable, which is not surprising given the low sample size and ranged from 0.24–0.56 for Novelty, 0.21–0.65 for Critical thinking, and 0.25–0.40 for Personal reflection.

Pedagogically Meaningful Learning Quality (PMLQ) provides a global or overall measure of the technical and pedagogical evaluation of the learning platform and learning materials (Nokelainen, 2006). PMLQ provides a measure of involvement with the curriculum as well as tapping perceptions that the program is meaningful to participants. The involvement component captures immersion in the program, a key element in engagement. The other components are less direct indicators. Our logic is that youth will not engage in an educational program that is not meaningful to them. The construct was measured using 14 selected agree-disagree items with response categories ranging from 1 (“*strongly disagree*”) to 5 (“*strongly agree*”).

Items measure factors such as learners' *perceived memory burden* (i.e., "it was easy to learn in this program" and "the amount of information was right for me") and *depth of involvement with the curriculum* (i.e., "when using the program, I forgot what was happening around me and how much time I spent on it" and "I liked the answers I gave while going through the program"). Other items tapped the youth's *perceived clarity of curriculum goals and objectives* (i.e., "program told me clearly what I'm expected to know (or learn)" and "I understand why the material was useful"), *perceptions of the transferability of curriculum knowledge to a new context* (i.e., "I will be able to use the skills and knowledge I learned in the future" and "I feel that this program was designed for me"), and the *perceived degree to which the online learning system provided encouragement and commentary* (i.e., "the program gave me motivating feedback" and "this program gave me immediate feedback on my work"). Items were averaged, with higher scores indicating higher usability ($M = 4.08$, $SD = 0.50$, range = 2.64–5). The overall PMLQ scale reliability was good (Cronbach's $\alpha = .85$).

System Usability Scale (SUS) provided a second measure of the program's overall or global engagement (Brooke, 1996). This measure was obtained after participants completed the entire program and provides an indirect measure of engagement. Our logic was that youth would not engage with a program that they found difficult to use. The SUS is a common overall measure evaluating digital programs (Brooke, 2013; Bangor, Kortum, & Miller, 2008).

The scale used in this study consists of six selected agree-disagree items with response categories ranging from 1 ("*strongly disagree*") to 5 ("*strongly agree*"). Items ask about agreement with the statements "program was too complex" (reverse-coded), "this program was easy to use," "the parts of this program were well integrated," "this program was very hard to use" (reverse-coded), "I felt very confident using this program," and "I liked the narrator in the program." Items were averaged, with higher scores indicating better reported system usability ($M = 4.18$, $SD = 0.58$, range = 2.80–5). The scale demonstrated adequate reliability (Cronbach's $\alpha = 0.70$).

Analytic engagement measures.—The project captured two forms of program analytic data to measure engagement: optional depth and challenge question accuracy.

Optional depth is a continuous indicator of the number of optional depth segments that the respondent chose to explore, with a higher score indicating that a participant engaged in more segments (L1 $M = 0.7$, $SD = 0.84$, range = 0–2; L2 $M = 1.5$, $SD = 1.20$, range = 0–4; L3 $M = 0.8$, $SD = 0.65$, range = 0–2; L4 $M = 0.5$, $SD = 0.65$, range = 0–2). Conceptually, the choice to participate in optional content reflects engagement with materials. Disengaged participants do not voluntarily explore optional content; they seek to get through the material as quickly as possible. On average, participants completed 3.5 optional depth segments in REAL media ($SD = 2.51$, range = 0–8).

Challenge correct count is a continuous indicator of the number of correct answers the respondent gave to questions posed in the final section of each level, or the level "challenge." It is a mean score of correct responses on challenge items (L1 $M = 0.4$, $SD =$

0.31, range = 0–1; L3 M = 0.6, SD = 0.26, range = 0–1; L4 M = 0.6, SD = 0.31, range = 0–1). The challenges reflect knowledge of the curriculum, an outcome we deem reflective of engaging with the materials. Since the levels were short (e.g., approximately 20 minutes) and many of the challenge questions were not particularly difficult, youth who engaged the materials were likely to get many of the questions correct. On average, participants correctly answered 43.4% of challenge questions across all levels (SD = 0.24, range = 0–0.86).

Observed engagement measures.—Two forms of observed engagement measures were used in the present study, both derived from participant responses to open-ended items.

Word count.: Word count is a continuous measure of the number of words that respondents wrote in response to six open-ended program-content related items: 1) What's one thing that would be unethical or wrong to do on social media?; 2) Note the different settings where advertisers show people having fun with the group using their product. Which setting works the best for you?; 3) Look at this next ad. Then answer three questions. What is fun about it?; 4) What claim is this ad making?, 5) What is the missing claim?; and 6) Think about the ad we just talked about. Why are they on a raft in the ocean and not somewhere else?." Conceptually, word counts were chosen as a proxy for engagement since participants did not need to write anything let alone lengthy responses to these queries. We argue that more engaged youth will provide longer answers. Word counts for each item were averaged, with higher scores on the resulting measure indicating that the respondent wrote more words in answering open-ended items (M = 12.1, SD = 6.01, Mdn = 10.6, range = 4.8–31).

Complexity.: Complexity is a multi-dimensional construct reflecting an individual's ability to differentiate, integrate, and abstract phenomena. The measure was derived from the construct of cognitive complexity (Burleson & Waltman, 1988; O'Keefe & Sypher, 1981) and conceptualized as a proxy for depth of involvement (i.e., more complex responses reflect greater depth of involvement). Two undergraduate research assistants were trained to code six open-ended questions that appear in REAL media for the three complexity dimensions: differentiation (the number of constructs a respondent used in evaluating an issue or ad), integration (the quality of construct organization and connectedness in the response), and abstraction (the degree of concreteness versus generality in construct description in the response). Inter-rater reliability between the two assistants was calculated by complexity dimension for each item and across all six scored items, counting adjacent scores as agreement. Average inter-rater reliability for all dimensions across all items was very high at 0.96 and exceeded 0.84 for every individual dimension. A graduate research assistant independently scored any dimensions on items where undergraduate coders' scores differed by more than 1 point. A mean score was derived for overall complexity (M = 2.2, SD = 0.81, range = 0.88–3.72) and for each dimension from scores across all six open-ended items, as described below.

Differentiation is a continuous measure of the number of constructs a respondent used in evaluating an issue or ad (O'Keefe & Sypher, 1981). As with complexity, differentiation provides a deeper analyses or engagement with materials. Constructs in each open-ended response were counted, discounting descriptives and modifiers. A higher number of constructs used indicates a higher differentiation score. Scores from individual items were averaged to obtain a mean differentiation score ($M = 1.7$, $SD = 0.58$, range = 0.83–2.60).

Integration is a continuous measure of construct organization, connectedness, and systematic relation (Burleson & Waltman, 1988). Again, more integration reflects a higher level of engagement since it takes more time and effort to construct an organized and systematic response. Integration for each open-ended response was scored on a scale from 0 (no integration) to 5 (response elaborates on contradictions, variability, and construct relations). Scores from individual items were averaged to obtain a mean integration score ($M = 1.8$, $SD = 0.56$, range = 0.66–2.75).

Abstraction is a continuous measure of the concreteness versus generality in construct description (Burleson & Waltman, 1988). Abstract responses reflect higher order thinking and thus should be related to engagement with materials. Constructs in each open-ended response were scored on a scale from 0 (no abstraction) to 4 (abstract, higher-level concepts) and scores were summed for each item. Scores from individual items were then averaged to obtain a mean abstraction score ($M = 3.2$, $SD = 1.49$, range = 1.17–6.83).

Analytic Plan

We calculated descriptive statistics for self-reported, observed, and program analytic indicators, and computed zero-order pairwise correlations between the measure sets. The present analyses cut across levels (i.e., summary scores were used for all variables). We chose this correlational approach given the nature of the data, the exploratory goal, and the small sample size. More powerful and sophisticated analyses are not supported by these current data. Similarly, results were presented as significant if they reached a level of $p < .10$ or less. The .10 level is used due the exploratory nature of the study and low sample size (Poitevineau & Lecoutre, 2001; Royall, 1986). We recognize the possibility of Type 2 error but given the exploratory nature of this study and the sample size we chose a more liberal criterion than the standard .05 level.

Results

Table 1 presents the zero-order Pearson’s correlation coefficients from pairwise correlation matrices of self-reported, observational, and program analytic measures derived from the REAL media program.

Correlations Among Program Analytic Engagement Indicators

Optional depth count.—The number of optional depth segments that youth engaged in is negatively correlated with complexity ($r = -.33$) and abstraction ($r = -.38$), but positively related to personal reflection ($r = .28$), PMLQ ($r = .38$), and SUS ($r = .59$). This program analytic indicator reveals that youth who provided more complex examples engaged less with program materials through optional depth, a puzzling and perhaps counter-intuitive finding, but they also provided more positive, overall program ratings on PMLQ and SUS.

Correct challenge count.—The mean number of correct responses to challenge questions is positively correlated with complexity ($r = .38$), integration ($r = .43$), and abstraction (r = .40). Unlike optional depth, youth who correctly answered more of the challenge questions tended to provide more complex open-ended responses. Unlike the optional depth segments, the challenge items were required to move forward in the program.

Correlations Among Observed Engagement Indicators

Word count.—Word count is positively correlated with complexity, with correlation coefficients for associations between mean complexity score and complexity dimensions with word count ranging from r = 0.59 to r = 0.70. Word count is also positively correlated with personal reflection (r = .40). It is negatively correlated with optional depth (−.33 for overall complexity and −.38 for abstraction) and SUS (r = −.46). Thus, observational metrics show that youth who wrote longer responses to open-ended items also exhibited higher complexity in their responses but were more negative in their overall perceptions of the program and less likely to explore optional depth.

Complexity.—As a function of measurement, mean complexity scores for the overall scale and subscales are strongly and positively intercorrelated as expected (r = .65-.97). Overall complexity is also positively correlated with personal reflection (r = .40), and the overall as well as each subscale are negatively correlated with SUS (r = −.47 to −.53). These findings support the construct validity of the complexity measures but also indicate that those providing complex answers may not engage as fully as those who provide less complex responses. We return to explanations for these findings in the conclusion.

Correlations Among Self-Reported Engagement Indicators

AES.—As with complexity, the intercorrelations among subscales and between subscales were all positive and generally high, consistent with prior research and theory (Greene et al., 2015). Exceptions were the correlations between novelty and personal reflection (r = .39), and novelty and critical thinking (r = .29). This may indicate that novelty operates somewhat independently, although it correlates highly with the overall scale mean (r = .71). Novelty is inherently tied to comparison between other educational programs and the current program, and the sample varied in terms of exposure in diverse school districts within one state as well as several home-schooled youths who may have different bases for comparison.

AES (overall and subscales) also exhibits strong correlations, in general, with PMLQ (r = .42-.71) and SUS (r = .38-.78), while the personal reflection subscale is also related to optional depth (r = .28) and word count (r = .40). Thus, AES demonstrates moderate to strong relationships with other engagement measures. Youth who reported the program as having higher AES personal reflection, novelty, and critical thinking reported that it was pedagogically and technically well designed and useful (PMLQ and SUS); additionally, these youth made greater use of optional features (e.g., optional depth, longer fill-in sections). This is especially promising given the lack of common scale variance.

PMLQ.—In addition to its relationship with AES and its subscales, PMLQ scores are positively correlated with SUS scores (r = .78). This strong relationship indicates the scales are measuring common usability characteristics of the program overall.

Discussion

Three types of engagement data were collected and analyzed in this study: self-report, observations, and analytic data. We urge caution in interpreting our analyses given the exploratory nature of the study as well as the smaller sample size. However, given the paucity of datasets with all three types of engagement data and the need to develop more sophisticated conceptual and methodological approaches to engagement, we argue that these findings are important for prevention and implementation science and online learning researchers to consider.

When patterns among the correlations were examined, the measures within these sets formed two separate clusters of measures with little intercorrelation between sets. The correlations are positive within each set and negative across the sets. This suggests they constitute very different aspects of engagement, a thought we return to after reviewing specific findings. We start with issues of reliability before moving to a discussion of the clustering.

Reliability of Measurement

Averaged across levels, the self-report data exhibit good reliability overall and within subscale. Notably, correlations between subscale item pairs within a given level were not as favorable; however, this was not surprising given the small sample size (most were significant and moderate in strength). Complexity, the one coded observable variable, also exhibited strong reliability for coding. Thus, even with low sample size, some measures with few items, and few coders, the measurement was generally reliable.

Clustering of Measures

In the absence of statistical data reduction approaches (factor analysis, cluster analysis, latent class analysis) we applied a conceptual or qualitative approach to clustering. As noted previously, there appear to be two engagement clusters. First, observed variables including complexity (and subscales of complexity) and word count were highly correlated, and complexity variables were also positively correlated with scores on the level challenges. Further, several of these variables were negatively related to the other self-report and analytic variables (e.g., optional depth, SUS). We had assumed that complexity and word count would tap the depth of a participant's engagement. While this perspective may be correct, the strong positive relationship between complexity and challenge responses and their negative relationship with the other variables, as well as the negative relationship between word count and the SUS, suggests that this factor appears to be related to the participant's competence or media literacy as related to program content. It is possible that some participants were already "media literate" and, as a result, produced longer and more complex answers and responded correctly to more of the challenge questions. If true, this would explain negative associations with other engagement scores, specifically the SUS total score. These participants may not have been adequately challenged by the material presented and were bored, and/or found the program to be too simplistic in its programming, an important consideration for future programming that utilizes individualized learning programs. This is consistent with previous findings the participant characteristics should be accounted for when considering engagement (Hidi & Renninger, 2006; Plass & Kinzer, 2015). This cluster can be expected to correlate highly with pretest media literacy scores in future research with this program, and if these associations hold and moderate program outcomes, findings would indicate the need to either provide more challenging material and/or present the curriculum to a younger audience.

Second, the self-report scales (AES, PMLQ, and the SUS) cluster together, with optional depth positively correlated with several of these variables. These measures appear to constitute a core engagement cluster that manifests itself across self-report, and analytic data. However, of the traditional usability scales (PMLQ and SUS), the SUS was not as strongly correlated with the theory-driven AES. Optional depth is particularly interesting as an analytic variable in this cluster, as it reflects choice on the part of the participant to proceed with non-required content. This may explain the positive association with usability measures – e.g., participants liked the experience and wanted to see more. It's also possible those who chose to see optional content were less "media literate" which would explain the negative associations with complexity – those with more competency were not interested in examining additional content.

Among the self-report measures, AES appears to be a more robust self-reported measure because it is theoretically driven and is a multi-dimensional construct that measures perceived novelty, personal reflection, and critical thinking as metrics of audience engagement with a program or intervention (Greene et al. 2015). These constructs are important indicators of (or predictors of) engagement, as defined by the TAI, and thus have a central role to play in evaluation. In addition, given the strong association between the SUS and PMLQ scales, both may not be needed in future analyses. Based on the pervasiveness of the SUS scale in the literature evaluating online programs and interventions (Brooke, 2006; Brooke, 2013; Bangor, Kortum, & Miller, 2008), its use may be more desirable than PMLQ for purposes of cross study comparison including but not limited to meta-analyses. Another added benefit of the SUS is that the original scale is much shorter than the PMLQ. However, the PMLQ arguably taps into a pedagogical aspect of usability not captured by the SUS. Notably, there is another option, the User Engagement Scale (UES), which was not assessed in the current study (O'Brien, Cairns, & Hall, 2018). The UES has both short and long forms and captures both technical usability and engagement with the platform being assessed.

Ultimately, future evaluations may consider adopting the self-report scales that best suit their needs. For example, the UES would provide a good option if the goal is to have an assessment of both functionality and the extent to which an individual is immersed in content. Overall, the advantage of self-report measures is that they offer researchers an option that can be closely matched to the theoretical underpinnings of the program being evaluated, consistent with our choice of the AES given its relation to TAI, which guided the study.

When considering program analytics and observational variables, optional depth and word count are both easily measured and useful. Optional depth allows us to assess whether participants voluntarily explore additional segments in greater depth, thereby allowing the learning to happen at the participants' chosen pace and level of immersion. Similarly, word count of open-ended program-content related items provides another objective measure, although calculating word count requires additional steps. Both of these features can be incorporated in program data capture (and may be easy to retrieve in specific LMS) and would be crucial to consider in the planning stages of programming. Whereas these indicators offer an advantage to self-report data in their ease of collection and lack of respondent burden, questions remain as to what extent these variables tap into other dimensions of engagement, and/or are reflective of other constructs altogether. Further, there are many other options for analytic and observational measures, all with their own pros and cons, which could be considered (Henrie, Halverson, & Graham, 2015).

We expect that these clusters of self-report, observational and analytic engagement indicators will be strongly related to program outcomes in future research, both at short-term and longer-term. Greater engagement in an online program or intervention may generate longer lasting influence on individual behavioral choices, and lead to longer-term favorable outcomes (Greene, 2013).

If, as these data suggest, analytics can form a core piece of outcome evaluations, other factors must also be taken into consideration. For example, it is important to note the ethical issues related to Facebook and other social media platforms' use of analytics that have recently been raised in the popular press (e.g., Singer, 2018). Facebook, for example, collects data about user behavior on the site and employs it for a variety of commercial purposes, raising questions related to privacy and transparency. In the present study, we discuss the use of unobtrusive participant data collection as they use REAL media. In both cases (REAL media and sites such as Facebook), participant consent is obtained, although one may question if either set of participants thoroughly reads the consents—in our terminology fully engages with the notion of consent—and/or if they understand the full implications of the information collected. Are these equal invasions of privacy? As analytic data become more common in our interventions, greater attention to consent and these ethical issues is needed.

In conclusion, this paper argues that engagement is a central construct in developing both prevention science and online learning programs. Too often in prior research, participants have been treated almost as inanimate objects, passively receiving prevention interventions or simply responding to ratings of "liking". For example, the implementation literature often focused on "fidelity" or the degree to which the implementer adheres to the curriculum, rather than on including participant engagement in the broader constructs of "implementation quality" (Pettigrew et al., 2015). In contrast, this paper argues that engagement should be conceptualized as central to program design, implementation, and evaluation.

The small sample size and exploratory nature of these analyses are a notable limitation of the study and conclusions must be approached with caution. Nonetheless, we believe this model suggests an approach to conceptualizing and measuring engagement that advances implementation science to a more nuanced approach often articulated by review pieces but only infrequently attempted (Pettigrew et al., 2015). This approach highlights participants' actions and perceptions while using the curriculum. By focusing on engagement, participants are cast into a more active role in the process in contrast with the more passive or receiving view that is common in implementation research and some online learning programming.

We note that stronger relationships between some of the constructs may be observed with larger samples and more powerful tests. It also may be, however, that engagement needs to be conceptualized in more discrete ways; there may, in fact, be different types of engagement (e.g., engagement with content, engagement with technology) similar to the conceptualization articulated in Greene et al., 2015. In addition, when larger datasets are available, content analyses may unpack qualitative characteristics that provide a richer view of engagement such as examination of themes and framing in free response answers. A larger sample size would also allow for more sophisticated data reduction analyses mentioned prior.

In conclusion, this paper offers a conceptualization of an engagement that lead us to collect self-report, analytic, and observational data and an exploratory study that suggests a measurement model to accomplish these purposes, an approach that can guide future research.

Acknowledgements:

We gratefully acknowledge the contributions of the 4-H clubs and their members, particularly Rachel Lyons of 4-H and Rutgers University.

Funding sources:

Supported by NIDA/NIH (R21DA027146, R41DA039595, R42DA039595) and NCI/NIH (P30CA008748) grant funding.

References

Allen GJ, Lerner WM, & Hinrichsen JJ (1972). Study behaviors and their relationships to test anxiety and academic performance. Psychological Reports, 30, 407–410. doi:10.2466/pr0.1972.30.2.407 [PubMed: 5024914]

Appleton JJ, Christenson SL, & Furlong MJ (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. Psychology in the Schools, 45, 369–386. doi:10.1002/pits.20303

Arnett J (1992). Reckless behavior in adolescence: A developmental perspective. Developmental Review, 12, 339–373. doi:10.1016/0273-2297(92)90013-R

Banerjee SC, Greene K, Magsamen-Conrad K, Elek E, & Hecht ML (2015). Interpersonal communication outcomes of a media literacy alcohol prevention curriculum. Translational Behavioral Medicine, 5:4, 425–432. doi: 10.1007/s13142-015-0329-9 [PubMed: 26622915]

Banerjee SC, & Greene K (2007). Anti-smoking initiatives: Effects of analysis versus production media literacy interventions on smoking-related attitude, norm, and behavioral intention. Health Communication, 22, 37–48. doi:10.1080/10410230701310281 [PubMed: 17617012]

Banerjee SC, & Greene K (2006). Analysis versus production: Adolescent cognitive and attitudinal responses to anti-smoking interventions. Journal of Communication, 56, 773–794. doi: 10.1111/j.1460-2466.2006.00319.x

Bangor A, Kortum PT, Miller, JT. An empirical evaluation of the system usability scale. International Journal of Human-Computer Interaction.2008; 24, 574–594. doi:10.1080/10447310802205776

Brooke J (2013). SUS: A retrospective. Journal of Usability, 8 (2), 29–40.

Brooke J (1996). SUS: A quick and dirty usability scale. In Jordan PW, Thomas B, Weerdmeester BA, & McClelland IL (Eds.), Usability evaluation in industry (pp. 4–7). London: Taylor & Francis.

Burleson BR, & Waltman MS (1988). Complexity: using the role category questionnaire measure. In Tardy C (Ed.), A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes (pp. 1–35). Norwood, NJ: Abex Publishing.

Davis KSridharan H, Koepke LSingh S & Boiko R (2018). Learning and engagement in a gamified course: Investigating the effects of student characteristics. J Comput Assist Learn, 34, 492–503.

De Boer W, & Collis B (2005). Becoming more systematic about flexible learning: Beyond time and distance. ALT-J Research in Learning Technology, 13, 33–48. doi:10.3402/rlt.v13i1.10971

Dev PC (1997). Intrinsic motivation and academic achievement: What does their relationship imply for the classroom teacher? Remedial and Special Education, 1, 12–19. doi:10.1177/074193259701800104

Dicheva D, Dichev C, Agre G, & Angelova G (2015). Gamification in education: A systematic mapping study (report). Educational Technology& Society, 18(3), 75–88.

Durlak J, Weissberg R, & Pachan M (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. American Journal of Community Psychology, 45, 294–309. doi:10.1007/s10464-010-9300-6. [PubMed: 20300825]

Durlak JA, & DuPre EP (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. American Journal of Community Psychology, 41, 327–350. [PubMed: 18322790]

Dusenbury L, Hansen WB, Jackson-Newsom J, Pittman D, Wilson C, Simley K, Ringwalt C, Pankratz M, & Giles S (2010). Coaching to enhance quality of implementation in prevention. Health Education, 110, 43–60. [PubMed: 22022672]

Dusenbury L, Brannigan R, Hansen WB, Walsh J, & Falco M (2005). Quality of implementation: developing measures crucial to understanding the diffusion of preventive interventions. Health Education Research, 20, 308–313. doi:10.1093/her/cyg134. [PubMed: 15522898]

Finn JD, & Zimmer KS (2012). Student engagement: What is it? Why does it matter? In Christenson SL, Reschly AL, & Wylie C (Eds.), Handbook of research on student engagement (pp. 97–131). Boston, MA: Springer.

Fredricks JA, Blumenfeld PC, & Paris AH (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74, 59–109. doi:10.3102/00346543074001059

Giesbers B, Rienties BTempelaar D, & Gijselaers W (2013). Investigating the relations between motivation, tool use, participation, and performance in an e-learning course using web-videoconferencing. Computers in Human Behavior, 29, 285–292. doi:10.1016/j.chb.2012.09.005

Girard C, Ecalle. J, & Magnan A (2012). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. Journal of Computer Assisted Learning, 29:3, 207–219. 10.1111/j.1365-2729.2012.00489.x

Grabe M, & Christopherson K (2005). Evaluating the advantages and disadvantages of providing lecture notes: The role of internet technology as a delivery system and research tool. Internet and Higher Education, 8, 291–298. doi:10.1016/j.iheduc.2005.09.002

Grabe M, & Christopherson K (2008) Optional student use of online lecture resources: Resource preferences, performance and lecture attendance. Journal of Computer Assisted Learning, 24, 1–10. doi:10.1111/j.1365-2729.2007.00228.x

Greene K, Banerjee SC, Ray AE, & Hecht ML (2017). Active involvement interventions in health and risk messaging. In Parrott RL (Ed.), Oxford encyclopedia of health and risk message design and processing (pp. 1–36). New York, NY: Oxford University Press. doi: 10.1093/acrefore/9780190228613.013.527

Greene K, Catona D, Elek E, Magsamen-Conrad K, Banerjee SC, & Hecht ML (2016). Improving prevention curricula: Lessons learned through formative research on the Youth Message Development Curriculum. Journal of Health Communication, 21, 1071–1078. doi: 10.1080/10810730.2016.1222029 [PubMed: 27684111]

Greene K, Yanovitzky I, Carpenter ABanerjee SC, Magsamen-Conrad K, Hecht ML, & Elek E (2015). A theory-grounded measure of adolescents' response to a media literacy intervention. Journal of Media Literacy Education, 7, 35–49.

Greene K (2013). The Theory of Active Involvement: Processes underlying interventions that engage adolescents in message planning and/or production. Health Communication, 28, 644–656. doi: 10.1080/10410236.2012.762824 [PubMed: 23980581]

Greene K (2012, 4). Active involvement: A brief media literacy approach to substance use prevention. Paper presented on a panel highlighting the value of Communication Research at the annual meeting of the Society for Behavioral Medicine, New Orleans, LA.

Greenwald AG, & Gilmore GM (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. Journal of Educational Psychology, 89, 743–751. doi:10.1037/0022-0663.89.4.743

Hamutoglu NB, Gemikonakli O, Duman Iet al. (2019). Evaluating students experiences using a virtual learning environment: satisfaction and preferences. Education Tech Research Dev (2019). 10.1007/s11423-019-09705-z

Hecht ML, & Miller-Day M (2009). The Drug Resistance Strategies Project: Using narrative theory to enhance adolescents' communication competence. In Frey L & Cissna K (Eds.), Routledge handbook of applied communication (pp. 535–557). New York and London: Routledge

Henrie CR, Halverson LR, & Graham CR (2015). Measuring student engagement in technology-mediated learning: A review. Computers & Education, 90, 36–53.

Herodotou C, Rienties B, Boroowa Aet al. (2019). A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective. Education Tech Research Dev 67: 1273. 10.1007/s11423-019-09685-0

Hidi S, & Renninger K (2006). The four-phase model of interest develop-ment. Educational Psychologist, 41(2), 111–127.

Kahn P, Everington L, Kelm K, Reid I, & Watkins F (2017). Understanding student engagement in online learning environments: The role of reflexivity. Educational Technology Research and Development, 65(1), 203–218.

Kolari S, Savander-Ranne C, & Viskari E-L. (2008). Learning needs time and effort: A time-use study of engineering students. European Journal of Engineering Education, 33, 483–498. doi:10.1080/03043790802564046

Lee E, & Hannafin MJ (2016). A design framework for enhancing engagement in student-centered learning: Own it, learn it, and share it. Educational technology research and development, 64(4), 707–734.

Lerner RM, & Lerner JV (2013). The positive development of youth: Comprehensive findings from the 4-H study of positive youth development. Available online at: http://www.4-h.org/about/youth-development-research/positive-youth-development-study/ (Accessed April 27, 2018).

Li J, Wong SC, Yang Xet al. (2019). Using feedback to promote student participation in online learning programs: evidence from a quasi-experimental study. Education Tech Research Dev (2019). 10.1007/s11423-019-09709-9

Lillehoj CJ, Griffin KW, & Spoth R (2004). Program provider and observer ratings of school-based preventive intervention implementation: agreement and relation to youth outcomes. Health Education & Behavior, 31, 242–257. Doi:10.1177/1090198103260514. [PubMed: 15090124]

Lin LC, Hung IC, Kinshuket al. (2019) The impact of student engagement on learning outcomes in a cyber-flipped course. Education Tech Research Dev 10.1007/s11423-019-09698-9.

Low S, Ryzin MJV, Brown EC, Smith BH, & Haggerty KP (2013). Engagement matters: lessons from assessing classroom implementation of steps to respect: A bullying prevention program over a one-year period. Prevention Science. doi:10.1007/s11121-012-0359-1.

Ma L, & Lee CS (2018). Investigating the adoption of MOOCs: A technology–user–environment perspective. Journal of Computer Assisted Learning, 34:6, 1–10. 10.1111/jcal.12314

Macfadyen LP, & Dawson S (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. Journal of Educational Technology and Society, 15, 149–163.

Mason RB (2011). Student engagement with, and participation in, an e-forum. Journal of Educational Technology & Society, 14, 258–268.

Morris LV, Finnegan C & Wu S-S (2005). Tracking student behaviours, persistence, and achievement in online courses. Internet and Higher Education, 8, 221–231. doi:10.1016/j.iheduc.2005.06.009

Nokelainen P (2006). An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students. Educational Technology and Society, 9, 178–197.

O'Brien HL, Cairns P, & Hall M (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies, 112, 28–39.

Oh J, Bellur S, & Sundar SS (2018). Clicking, assessing, immersing, and sharing: An empirical model of user engagement with interactive media. Communication Research, 45(5), 737–763.

O'Keefe DJ, & Sypher HE (1981). Complexity measures and the relationship of Complexity to communication. Human Communication Research, 8, 72–92. doi:10.1111/j.1468-2958.1981.tb00657.x

Peled A, & Rashty D (1999). Logging for success: Advancing the use of the WWW logs to improve computer mediated distance learning. Journal of Educational Computing Research, 21, 413–431. doi:10.2190/NLR6-K355-LAQY-U01D

Pettigrew J & Hecht ML (2015). Developing prevention curricula. In Bosworth K (Ed.) Prevention Science in School Settings: Complex Relationships and Processes (pp. 151–174). New York, NY: Springer.

Pettigrew J, Graham JW, Miller-Day M, Hecht ML, Krieger JL, & Shin Y (2015). Adherence and delivery quality: Implementation quality and outcomes of 7th grade keepin' it REAL program. Prevention Science, 16, 90–99. doi:10.1007/s11121-014-0459-1 [PubMed: 24442403]

Plass J, Homer B, & Kinzer C (2015). Foundations of game -based learning. Educational Psychologist, 50(4), 258–283.

Poitevineau J, & Lecoutre B (2001). Interpretation of significance levels by psychological researchers: The .05 cliff effect may be overstated. Psychonomic Bulletin & Review, 8, 847. 10.3758/BF03196227 [PubMed: 11848609]

Rafaeli S, & Ravid G (1997). Online, web-based learning environment for an information systems course: Access logs, linearity and performance. Available online at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.9119 (Accessed April 2018).

Reynolds R, & Caperton IH (2011). Contrasts in student engagement, meaning-making, dislikes, and challenges in a discovery-based program of game design learning. Educational Technology Research and Development, 59(2), 267–289.

Royall RM (1986) The effect of sample size on the meaning of significance tests. The American Statistician, 40:4, 313–315. doi: 10.1080/00031305.1986.10475424

SAMHSA NREPP. (2017). Youth Message Development. Available online at: https://nrepp.samhsa.gov/ProgramProfile.aspx?id=201 (Accessed April 27, 2018).

Singer N (2018, 4 11). What you don't know about how Facebook uses your data. The New York Times. https://www.nytimes.com/2018/04/11/technology/facebook-privacy-hearings.html Retrieved 28 April 2018.

Soffer T, & Nachmias R, (2018). Effectiveness of learning in online academic courses compared with face-to-face courses in higher education Journal of Computer Assisted Learning, 34:5, 534–543. 10.1111/jcal.12258

Stewart M, Stott T, & Nuttall A (2011). Student engagement patterns over the duration of level 1 and level 3 geography modules: Influences on student attendance, performance and use of online resources. Journal of Geography in Higher Education, 35, 47–65. doi:10.1080/03098265.2010.498880

Tobler NS, Roona MR, Ochshorn P, Marshall DG, Streke AV, & Stackpole KM (2000). School-based adolescent drug prevention programs: 1998 meta-analysis. The Journal of Primary Prevention, 20, 275–336. doi:10.1023/A:1021314704811.

Tran B, Nguyen Q, Nong V, Maher R, Nguyen A, Nguyen H, & Le H (2015). Student-centered outcomes of an e-learning course on public health in Hanoi and New York. European Journal of Open, Distance and E-Learning, 17, 1–14. doi:10.2478/eurodl-2014-0016

Van den Brande L (1994). Flexible and distance learning. Chichester, UK: John Wiley & Sons.

van Leeuwen A (2019). Teachers' perceptions of the usability of learning analytics reports in a flipped university course: When and how does information become actionable knowledge? Education Tech Research Dev (2019) 67: 1043. 10.1007/s11423-018-09639-y

Wagner P, Schober B, & Spiel C (2008). Time students spend working at home for school. Learning and Instruction, 18, 309–320. doi:10.1016/j.learninstruc.2007.03.002

Wagstaff R, & Mahmoudi H (1976). Relation of study behaviors and employment to academic performance. Psychological Reports, 38, 380–382. doi:10.2466/pr0.1976.38.2.380

Wellman GS, & Marcinkiewicz H (2004). Online learning and time-on-task: Impact of proctored vs. un-proctored testing. Journal for Asynchronous Learning Networks, 8, 93–104.

Zaiane OR, & Luo J (2001). Towards evaluation learners' behaviour in a web-based distance learning environment. IEEE International Conference on Advanced Learning Technologies Proceedings. 1–4. doi:10.1109/ICALT.2001.943944

Table 1.

Program Analytic, observed, and self-reported engagement indicators pairwise correlation matrix

		PAM		OM					SRM					
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Program analytic measures (PAM)														
1	Optional depth count	1.00	-	-	-	-	-	-	-	-	-	-	-	-
2	Challenge correct count	-	1.00	-	-	-	-	-	-	-	-	-	-	-
Observed measures (OM)														
3	Word count mean	-	-	1.00	-	-	-	-	-	-	-	-	-	-
4	Complexity mean	−0.33	0.38	0.70	1.00	-	-	-	-	-	-	-	-	-
5	Differentiation mean	-	-	0.68	0.91	1.00	-	-	-	-	-	-	-	-
6	Integration mean	-	0.43	0.59	0.79	0.66	1.00	-	-	-	-	-	-	-
7	Abstraction mean	−0.38	0.40	0.65	0.97	0.83	0.65	1.00	-	-	-	-	-	-
Self-reported measures (SRM)														
8	AES Total	-	-	-	-	-	-	-	1.00	-	-	-	-	-
9	AES Personal	0.28	-	0.40	-	-	-	-	0.88	1.00	-	-	-	-
10	AES Novelty	-	-	-	-	-	-	-	0.71	0.39	1.00	-	-	-
11	AES Critical	-	-	-	-	-	-	-	0.83	0.73	0.29	1.00	-	-
12	PMLQ Total	0.38	-	-	-	-	−0.36	-	0.71	0.42	0.60	0.56	1.00	-
13	SUS Total	0.59	-	−0.46	−0.53	−0.48	−0.47	−0.49	0.38	-	0.38	0.35	0.78	1.00

Notes. N=38. Only coefficients significant at p<0.10 or lower are shown.