2020

Informing Makerspace Outcomes Through a Linguistic Analysis of Written and Video-Recorded Project Assessments

Kevin M. Oliver

Jennifer K. Houchins

Robert L. Moore
Old Dominion University

Chuang Wang

Original Publication Citation

Oliver, K. M., Houchins, J. K., Moore, R. L., & Wang, C. (2020). Informing Makerspace Outcomes Through a Linguistic Analysis of Written and Video-Recorded Project Assessments. *International Journal of Science and Mathematics Education*, 22 pp. doi:10.1007/s10763-020-10060-2

International Journal of Science and Mathematics Education
https://doi.org/10.1007/s10763-020-10060-2

Informing Makerspace Outcomes Through a Linguistic Analysis of Written and Video-Recorded Project Assessments

Kevin M. Oliver[1] · Jennifer K. Houchins[1] · Robert L. Moore[2] · Chuang Wang[3]

Received: 4 September 2018 / Accepted: 17 January 2020

Abstract

A growing body of research focuses on what outcomes to assess in makerspaces, and appropriate formats for capturing those outcomes (e.g. reflections, surveys, and portfolios). Linguistic analysis as a data mining technique that holds promise for revealing different dimensions of learning exhibited by students in makerspaces. In this study, student reflections on makerspace projects were gathered in 2 formats over 2 years: private written assessments captured in the 3D GameLab gamification platform, and semi-public video-recorded assessments posted in the more social FlipGrid platform. Transcripts of student assessments were analyzed using Linguistic Inquiry Word Count (LIWC) to generate 4 summary variables thought to inform makerspace outcomes of interest (i.e. analytical thinking, authenticity, clout, and emotional tone). Comparative findings indicate that written assessments may elicit more analytical thinking about maker projects compared with less analytical conversation in videos, while video assessments may elicit somewhat higher clout scores as evidence of social scaffolding along with a much more positive emotional tone. Recommendations are provided for layering assessment approaches to maximize the potential benefits of each format, including reflective writing for social spaces, in social groups, and about design processes and procedures.

Keywords Assessment · Data mining · Learning analytics · Linguistics · LIWC · Makerspace · Reflection

✉ Kevin M. Oliver
kmoliver@ncsu.edu

[1] Department of Teacher Education and Learning Sciences, College of Education, North Carolina State University, 402T Poe Hall, Box 7801, Raleigh, North Carolina 27695-7801, USA

[2] Old Dominion University, Nortfolk, USA

[3] University of North Carolina at Charlotte, Charlotte, USA

The increasing availability of inexpensive and open source tools for consumer making along with forums for sharing and remixing makes has led to the increased prevalence of makerspaces in communities, homes, and schools (Hagel, Brown & Kulasooriya, 2014). Makerspaces have considerable historical antecedents and similar principles to Montessori schools, Dewey's Progressive Education movement, Piaget's constructivist learning, Papert's constructionist learning, and Kolb's experiential learning (Herr-Stephenson, Rhoten, Perkel & Sims, 2011; Hira, Joslyn & Hynes, 2014; Sheridan, Halverson, Litts, Brahms, Jacobs-Priebe & Owens, 2014). For example, in makerspaces, students take advantage of shared tools, resources, and expertise to promote interest-driven creation and play with supportive communities (Educause, 2013; Kurti, Kurti & Fleming, 2014).

Informal learning traditionally places less emphasis on assessment, yet researchers have begun to ask what outcomes can and should be assessed in making. School systems and agencies providing funds for makerspaces likewise are interested in outcomes supported to justify their investment. Assessment of informal learning is a known challenge given divergent social-cognitive outcomes one could study (see, for example, Lemke, Lecusay, Cole & Michalchik, 2015) and the restrictive nature of tapping into developing ideas, questioning, and interests (Brody, Bangert & Dillon, 2007). Lemke et al. (2015) found that effective documentation and assessment of informal learning activities should not only include content knowledge but also social, emotional, and developmental outcomes. Regarding makerspace outcomes in particular, the Tinkering Learning Dimensions Framework (TLDF) likewise recommends looking at not only development of understanding but also initiative and intentionality, social scaffolding, and engagement (Bevan, Gutwill, Petrich & Wilkinson, 2015). The TLDF framework is designed to be applied as a "guide" to design robust makerspace activities capable of producing the noted outcomes, or as a "reflective tool" to determine if collected evidence supports the presence of a given outcome (Bevan, Ryoo, Vanderwerff, Wilkinson & Petrich, 2017, p. 5). Without an instrument to assess TLDF dimensions, we studied analog linguistic measures in this study.

Another question of interest is how to assess these recommended outcomes in the context of a makerspace with researchers "not providing a firm answer on how makerspace learning can be measured" (Peppler, Keune, Xia & Chang, 2018b, p. 11). Assessment in makerspaces can be a challenge as pausing for documentation disrupts the flow of making (Litts, Kafai, Fields, Halverson, Peppler, Keune et al., 2016). In their survey reflecting 28 out-of-school makerspaces, Peppler et al. (2018) found that 64% did attempt to assess learning despite the challenges. Student self-reflection at the end of a project was the most common form of assessment, along with exit surveys, peer assessments, and portfolios (Peppler et al., 2018). Portfolios in particular have been touted for their ability to support documentation, sharing, and reflection on learning in makerspaces (Keune & Peppler, 2017).

Revealing Informal Outcomes Through Linguistic Analysis

Given desired makerspace outcomes can be reflected in linguistic elements and the primary forms of makerspace assessment generate words or transcriptions (i.e. reflections, surveys, and portfolios), it is worth considering the value of linguistic analysis as a

data mining technique to reveal outcomes of interest. Linguistic Inquiry Word Count (LIWC) is one popular data mining package that outputs different linguistic variables to represent a text (Moore, Oliver & Wang, 2019; Pennebaker, Boyd, Jordan & Blackburn, 2015). We analyzed the presence of four LIWC summary variables in makerspace assessments (analytical thinking, authenticity, clout, and emotional tone) as these variables were seen as related to the four TLDF dimensions advocated for makerspaces and might efficiently provide evidence of the dimensions since a TLDF instrument was not otherwise available. LIWC summary variable scores for a given text are derived from algorithms based on combinations of individual LIWC variables (Pennebaker, Chung, Frazee, Lavergne & Beaver, 2014; Pennebaker et al., 2015). For example, the analytical thinking summary variable is based on a factor in which positively loaded LIWC variables (i.e. articles and prepositions) are added, and negatively loaded LIWC variables are subtracted (i.e. personal/impersonal pronouns, auxiliary verbs, conjunctions, adverbs, and negations) (Pennebaker et al., 2014). Summary variables are "standardized scores that have been converted to percentiles … ranging from 0 to 100" (Pennebaker Conglomerates, 2018). LIWC analyzes each text separately, with its score based on a match with words identified in previously analyzed texts.

The first summary variable analyzed in this study is analytical thinking which reflects "the degree to which people use words that suggest formal, logical, and hierarchical thinking patterns" (Pennebaker Conglomerates, 2018; Pennebaker et al., 2014). Texts that contain many articles ("a, an, the") and prepositions ("all, below, much") "reveal relatively formal and precise descriptions of categories (e.g. objects, events, goals, and plans)" and would score higher on the 100-point scale (Pennebaker et al., 2014, p. 6). In contrast, texts that contain more pronouns ("I, us, you") and auxiliary verbs ("are, did, have") "have been associated with more time-based stories and reflect a dynamic or (personal) narrative language style" (Pennebaker et al., 2014, p. 6). Analytical thinking may be a helpful measure to inform the TLDF tinkering dimension development of understanding in which learners are expected to "express claims/realizations," "offer explanations," and "apply prior knowledge" (Bevan et al., 2015, pp. 7–8). These tasks are indeed analytic and formal in nature compared with writing a story or narrative; thus, one might expect a student who is justifying claims and explaining to score higher on the LIWC analytical thinking measure relative to students writing in a more personal style.

The second summary variable analyzed in this study was authenticity. This variable can help to reveal inauthentic persons who are modifying what they are saying or filtering their talk for an audience on the low end of the scale (Bulkeley & Graves, 2018) versus authentic persons who are being more "personal, humble, and vulnerable" at the high end of the 100-point scale (Pennebaker Conglomerates, 2018). The authenticity algorithm was derived from honesty and deception studies with deceptive talk found to include "fewer self-references" to dissociate one's self from lies, more negative words (hate and sad) owing to the discomfort felt when lying, fewer exclusive words such as "except, but, and without" that honest people include in language to describe "what they did and what they did not do," and more simple motion verbs such as "walk, move, and go" that those telling falsehoods find easier to weave into stories (Newman, Pennebaker, Berry & Richards, 2003, p. 666–667). The authenticity variable may be helpful to inform the TLDF dimension initiative and intentionality in which students are expected to "take intellectual risks or show intellectual courage," "persist in the face of

setbacks or frustration," "set goals/pose problems," "plan steps," and "seek/respond to feedback" (Bevan et al., 2015, pp. 7–8). One might expect a student who is expressing authentic language to self-reference themselves in detailing what steps they took in a process, with the humble/vulnerable nature of authenticity reflected in people willing to describe setbacks with regard to the intellectual risks they have taken on.

The third summary variable analyzed in this study was clout which reflects "the relative social status, confidence, or leadership that people display through their writing or talking" (Pennebaker Conglomerates, 2018). The algorithm for clout was derived from an examination of studies focusing on the interactions between people with a finding that people with high status tend to use more first-person plural pronouns ("we") and second-person pronouns ("you") that are "other-oriented," compared with those with low status who tend to use more first-person singular pronouns ("I") that reflect self-attention and would score lower on the 100-point scale (Kacewicz, Pennebaker, Davis, Jeon & Graesser, 2014, p. 137). Clout may help to inform the TLDF dimension social scaffolding in which students are expected to "request/offer help/ideas," "notice/talk about others' work," "use/modify others' ideas," and "connect work" (Bevan et al., 2015, pp. 7–8). Students who are referring to others' work or how they have used/connected to others' work may have a higher clout score as they would use more "we" and "you" pronouns, compared with those whose work was not social using "I" first-person singular pronouns. The role of community in makerspaces is important and through this variable we may identify students who need to be encouraged to work more closely with peers and connect with peers' work.

The fourth and final summary variable analyzed in this study was emotional tone which combines positive and negative emotion words into a single variable using an algorithm that generates a higher score for using more positive words ("happy, good, and nice") and a lower score below 50 when using more negative words ("kill, ugly, and guilty") (Cohn, Mehl & Pennebaker, 2004, p. 689; Pennebaker Conglomerates, 2018). Engagement is the TLDF dimension that best aligns with emotional tone in which students spend time, "try something over and over," "display motivation or investment through affect/behavior," and "show emotions such as joy, pride, or disappointment" (Bevan et al., 2015, pp. 7–8). Students invested in maker projects who enjoy their work should describe it more positively with a higher-related LIWC emotional tone score.

In summary, the four LIWC summary variables computed in this study and their proposed alignment with four TLDF dimensions are presented in Table 1.

Two Tested Approaches to Makerspace Assessment

We analyzed two approaches to makerspace assessment (i.e. writing and video) to determine how they reflected LIWC summary variables and how they differed by format. The two assessment formats tested were selected for their purported motivational affordances that might increase student willingness to reflect in the informal space where reflection is difficult to elicit: private, written assessment between a student and mentor captured in a gamification platform (year one, case one); and semi-public, video-based assessment captured in a social media space (year two, case two).

To encourage and capture assessment in year one (2016–17), we populated more than 40 makerspace "quests" into the 3D GameLab (now Rezzly) gamification

t1.1 **Table 1** Proposed alignment of LIWC summary variables with related TLDF dimensions

LIWC Summary variables	Texts scoring closer to 100 include more ...	Related TLDF dimensions	Description
t1.3 Analytical thinking	Articles, prepositions, indicative of "formal, logical, and hierarchical thinking"	Development of understanding	Making claims and explanations using prior knowledge
t1.4 Authenticity	"Personal, humble, and vulnerable" language, self-references, exclusive words	Initiative and intentionality	Goal setting, planning, persisting, taking intellectual risks
t1.5 Clout	"Other-oriented" first-person plural and second-person pronouns	Social scaffolding	Request/offer help, connect work, notice or modify others' work
t1.6 Emotional tone	Positive words ("happy," "good," "nice")	Engagement	Display motivation, show emotions

platform and assigned each student an account. Students completed quests in three project areas to earn points, levels, and badges consistent with gamification principles. A public player card and leader board showed quests completed by students and who had earned the most points or badges. To earn a badge, students had to complete both core (directed) quests and a self-directed quest addressing a problem of interest. Students documented quest completion by answering prompted questions in writing and uploading photos of their project for verification by club mentors.

Gamification systems are assumed to be intrinsically motivating on the basis of applying game mechanics that people associate with fun to learning (Marti-Parreno, Mendez-Ibanez & Alonso-Arroyo, 2016). Critics, however, argue that "reward-based" gamification systems are extrinsically motivating, advocating for "meaningful" gamification that uses non-point elements (challenges, narratives, play, and choice) to encourage personal connections to material (Becker & Nicholson, 2016). 3D GameLab not only reflected a hybrid approach with points, levels, badges, and leader boards, but also offered choice of quests. Students did not document many quests in this platform, and reflections were quite brief. Students commented that the platform did not seem like a game, just "more work," suggesting it was perceived as extrinsically motivating.

To address these perceptions of written assessment, in year two (2017–18), we opted to test the affordances of a more open, social platform for capturing reflections. FlipGrid is a web-based tool that prompts and collects video responses from any digital device with a camera (e.g. laptop and smart phone). In this platform, students again responded to prompted questions, but this time using selfie-style videos in which they held up, demonstrated, and discussed their work.

Social media platforms have received attention as a tool to engage learners and support identity development through articulation and community negotiation. Holloway (2015) tested Twitter to engage African-American males, noting that "it is hard to get students to check email or to respond to emails, but students do not mind sending text messages, tweeting, or using Facebook or other social media" (p. 104). Craig-Hare, Rowland, Ault and Ellis (2017) used social media for students to present evidence-based arguments with questioning, noting that the approach "emulates how scientists collaborate on their own research and share research findings" (p. 81). Pinkard, Erete, Martin and McKinney de

Royston (2017) applied the iRemix "social learning network" for posting maker work, receiving feedback, and peer-critiquing work (p. 488). Girls were able to cultivate "a desired social reputation" where "they were viewed, critiqued, and positioned by others, gaining recognition ...constructing their own narratives" (p. 488).

Predicting which Assessment Format Best Elicits Different Informal Outcomes

Some prior research has compared written and videoed assessments that help to predict in which assessment format we might expect to elicit desired informal learning outcomes from TLDF and their matched LIWC variables. First, research results are mixed in terms of which assessment format can best document development of understanding. In their comparison of video and written judgment tests, Lievens and Sackett (2006) reported video tests to be "more predictive of an interpersonal criterion" and written tests to be "more predictive of cognitive aspects of the criterion space," suggesting that written tests may better capture developing cognitive understanding (p. 1186). However, other studies illustrate how video can be used to capture cognitive-procedural outcomes when students are tasked with documenting their completion of some experiment or exercise in the video rather than simply answering reflection questions (Erdmann & March, 2014; Speed, Lucarelli & Macaulay, 2018; Tierney, Bodek, Fredricks, Dudkin & Kistler, 2014). Speed et al. (2018) asked biochemistry students to film video reports about experiments, with evidence of creative and critical thinking in videos and higher grades when compared with written lab reports. Tierney et al. (2014) asked organic chemistry students to record themselves using a molecular model kit in response to prompted exercises with videos posted on VoiceThread. This process was helpful to determine "a student's higher order thinking processes" and to visualize misconceptions (p. 984). Assessment videos that support "monitoring (of) student technique" may be particularly valuable in makerspaces where procedural skills are emphasized in addition to cognitive skills (Erdmann & March, 2014, p. 655).

In terms of the informal outcome social scaffolding, video tests being "more predictive of an interpersonal criterion" in the Lievens and Sackett (2006) study suggest that video might better elicit social elements. This effect would likely be amplified if students were prompted to notice their own and others' work as part of the video. Beyond creating video as an assessment artifact, created videos have also been studied as the basis for reflection which ties in with portfolios recommended for makerspaces and typically would include both artifacts and reflections (Oliver, Moore & Evans, 2017). Barry (2012) found that playing back video-recorded oral business presentations helped students to develop personal awareness of faults and areas for improvement.

Existing research has not informed the initiative and intentionality outcome; however, it would be helpful to know if the public video platform open to wider scrutiny encourages or discourages students from being honest and descriptive of their design process and from expressing vulnerability in describing any challenges or setbacks in design. Finally, with regard to the informal outcome engagement, a few studies have hinted that video assessment may be more engaging to students (Speed et al., 2018; Tierney et al., 2014). Speed et al. (2018) reported that filming video reports was more engaging and enjoyable than writing reports. Likewise, Tierney et al. (2014) reported

that students were "more at ease" filming video presentations compared with in-class presentations, and that they "enjoyed completing the exercises" (p. 985).

In summary, research suggests that written assessment may best support the capture of developing understanding, but video assessment may as well when designed to capture a process and related student thinking. Video assessment may provide an advantage to capture evidence of social scaffolding and engagement. It is unclear if video assessment may elicit or hinder the capture of initiative and intentionality. In the current study, LIWC was used to generate linguistic variables present in makerspace assessments with those elements compared by assessment format controlling for project type and grade level. Results illustrate how LIWC variables can inform the presence of desired TLDF tinkering dimensions and the conditions under which different dimensions may be elicited (i.e. assessment format, project type, and grade level).

Methodology

Design and Research Question

Since this study's change in assessment format (written to video) was emergent, the exploratory case study design was selected to inquire into changing conditions, to provide insight into the phenomenon of prompted student reflections in a physical makerspace, and to inform future research (Yin, 2014). LIWC generated linguistic profiles to represent assessments in these two unique cases. Profiles in cases were subsequently compared through multiple analysis of variance to reveal differences and to generate hypotheses regarding relative affordances of each format to be confirmed in future research. Two research questions were addressed and these are as follows:

1. How are four LIWC summary variables reflected in student makerspace assessments (i.e. analytical thinking, authenticity, clout, and emotional tone)?
2. Do summary variables differ on the basis of assessment format (written/video-recorded), makerspace project type (paper craft/paper circuit/soft circuit), or grade level (middle/high school, ages ~ 11–13 and ~ 14–18)?

Participants and Participation

Participants included both middle and high school students at a combined, public, all-girls school in an urban city in the southeastern United States. The school is populated by lottery but strives for racial balance and serving families without a prior college graduate. In 2016–17 (year one, case one), 34 students participated in the maker club with a mean attendance of 15.3 out of 25 meetings (std dev 6.9, average 30.7 contact hours). In 2017–18 (year two, case two), 37 students participated with a mean attendance of 16 out of 27 meetings (std dev 8.3, average 31.9 contact hours). In 2016–17, the mean number of written project assessments per club member was 4.8 (std dev 4.5), and in 2017–18, the mean number of video-recorded project assessments was 3.2 (std dev 2.3).

In 2016–17, participating 6th through 10th graders were more evenly distributed than in 2017–18 when a large group of 22 new 6th graders joined the club, and rising 10th

graders moved out of the school to an early college campus per this school's structure. Twenty-seven new students joined the club in year two, with ten continuing students participating and documenting projects in both years. The racial makeup of the two cases reflected both white ($n = 20/n = 21$) and black/Asian students ($n = 14/n = 16$) both years.

Procedures

Students worked on makerspace projects in three categories every Tuesday after school (September–May): circuitry (e.g. paper circuits, soft circuits, and LittleBits); programmed robotics (e.g. Hummingbird, Sphero, K'Nex, and MicroBits); and fabrication (e.g. paper crafts, MakeDo construction, and 3D pens/printers). In year one, students submitted written assessments and photos of completed projects to the 3D GameLab gamification platform. This assessment was private between the student and club mentors only, but students could see a leader board showing who had earned the most points and levels, as well as player cards showing projects completed and badges earned. Assessment was monitored in year one with club mentors reviewing submitted work and either approving the work with relevant gamification points/badges earned or returning the work to the student if they failed to answer any of the prompted questions.

Since students were resistant to assessment in year one, a more open and social system was employed in year two that might prove more motivational. Students were asked to submit a video documenting and visually displaying their work in FlipGrid. A different "grid" or response board was created for each project area. These grids were semi-public to anyone with the club's FlipGrid password and students could see and reply to others' assessments, although replying was rare when unprompted (and not included in this study's data set). Assessment in year two was not monitored or approved by club mentors, as points were not being assigned.

In both years, students were prompted to answer questions written by club mentors to prompt thinking: (1) What worked and did not work so well in completing your project? (2) What was the most challenging part of this project and how did you overcome that challenge? and (3) How would you change your process the next time? Steier and Young (2016) employed similar prompts in makerspace journals about challenges met and strategies to overcome them.

Data Sources and Projects Documented

The procedures resulted in 164 project assessments (written) in year one and 74 project assessments (video) in year two. The type of projects documented each year differed with three exceptions where students documented a given project in both years: paper craft (25 in Y1, 22 in Y2); paper circuits (27 in Y1, 21 in Y2); and soft circuits (24 in Y1, 13 in Y2). To explore differences between written and video-based assessments, comparisons were based on these three projects only to control for any influence of project type on student reflections.

To prepare data for analysis, written documentation was copied from 3D GameLab and video recordings were transcribed from FlipGrid then added to a combined spreadsheet with categorical codes to note how a given text was associated with assessment type, project type, and a responding student's grade level. All documentation was cleaned with misspellings corrected and missing punctuation added to better

reflect a student's intent when texts were read by LIWC. Thirteen video assessments were removed from the final video data set where students replied to question prompts in small groups rather than individually, as this group response could affect linguistic scores (i.e. 10 Spin Bot projects, 1 Coding Card project, and 2 Hummingbird projects).

Data Analysis

To probe for any differences between written and video-based assessment, each text was separately analyzed by LIWC (Pennebaker et al., 2015) with output summary variables for each text imported into SPSS for further analysis. Since only three project sets had enough captured assessments to compare between years (paper craft, paper circuit, and soft circuit), other assessments were removed from the analysis of variance. This resulted in a reduction of sample size from 238 to 132, of which 76 were from year one (written format) and 56 were from year two (video format). Nested in these reflections were 24 unique students in Y1 and Y2, and four students with reflections in both years. Y1 and Y2 samples were not entirely independent given these four students, although responses were spaced apart temporally. Three-way multivariate analysis of variance (MANOVA) was employed to examine the mean differences in a linear combination of the summary variables (analytical thinking, authenticity, clout, and emotional tone) between the year one (written format) and year two (video format) students. MANOVA is appropriate to decrease Type I error and to determine if "independent variables are related to combinations of dependent variables" (Warne, 2014, p. 3). Partial eta squared (η^2) values of .01, .06, and .14 were regarded as small, medium, and large effect sizes, respectively (Cohen, 1988).

Limitations

The assessment conditions in this study emerged between years one and two. Since key differences in conditions could impact on findings, future research should be used to confirm reported trends as consistent with exploratory case study research. In terms of monitoring, submitted written assessments in year one were reviewed, approved by club staff, and rewarded with points, levels, and badges, while video assessments in year two were not monitored or approved, nor did they result in any reward other than intrinsically sharing with the peer group. The dip in year two assessments may have resulted from students feeling less pressure to submit reflections to an unmonitored system despite being regularly encouraged to do so. A second difference of note between years one and two relates to social recognition. Written assessments in year one were viewable only by the submitting student and club staff (private), while video assessments in year two were viewable by peers (public). Finally, the students in this study were all female, and it is unclear if findings would generalize to mixed-gender settings.

Findings

Findings are divided into two sections. First, a general summary of the linguistic characteristics present in the overall data set is provided for context ($n = 238$ assessments). Second, findings specific to linguistic summary variables are provided between

matched project groups ($n = 132$ assessments) to inform the presence or absence of desired TLDF dimensions and the conditions under which they were elicited (i.e. by assessment type, project type, and grade level).

Overall Linguistic Characteristics

As shown in Table 2, video assessments included a much higher mean word count while written assessments included slightly more mean words per sentence and mean words greater than six letters. This finding might suggest that video assessments were more conversational and flowing in nature and writing was more analytic with longer sentences and bigger words.

The mean percentage of words in a text that reflected different linguistic types is shown in Table 3 (e.g. a mean of 7.82% of the words across all written assessments were "personal pronouns" as matched to the LIWC dictionary). In general, video assessments included a higher mean percentage of pronouns of every type, while written assessments included a higher mean percentage of articles (e.g. "the"). In a study by Pennebaker et al. (2014), the use of more pronouns was associated with a narrative, personal style by dynamic thinkers, contrasted with the use of more articles by categorical thinkers who ended up with higher GPAs. In the current study then, it is possible that video assessment elicited more narrative-style writing and written assessment more academic-style writing. A higher use of adjectives and quantifiers in written assessment lends strength to this argument as students were possibly more descriptive in a written mode where they had to detail their project for readers compared with a video mode where a project was more simply shown to the camera. Finally, the higher presence of negation words (e.g. "no") in written assessment may suggest a more negative tone in that mode.

Linguistic Summary Variables

Descriptive mean values of LIWC summary variables in the $n = 132$ assessments of matched year one/two project groups are shown in Table 4. In general, the variable clout had the lowest mean scores, perhaps owing to students engaging primarily individually rather than socially with assigned maker tasks. The variables analytical

t2.1 **Table 2** Independent samples *t* test comparing mean number of words across all written and video-recorded assessments

		Assessment source								
t2.3		Written (GameLab)			Video (FlipGrid)					
Q8		*M*	*SD*	n	*M*	*SD*	n	*t*	*p*	ES
t2.5	Word count	73.98	30.72	164	186.86	122.28	74	*− 7.83	.001	− 1.27
t2.6	Words per sentence	16.94	5.12	164	14.68	4.72	74	3.23	.001	.46
t2.7	Words > 6 letters	15.65	5.69	164	12.47	3.57	74	*5.23	.001	.67

*Equal variances not assumed; more conservative values reported

t3.1 **Table 3** Independent samples *t* test comparing mean percentage of words of a linguistic type across all written and video-recorded assessments

		Assessment source								
t3.3		Written (GameLab)			Video (FlipGrid)					
Q9		*M*	*SD*	n	*M*	*SD*	n	*t*	*p*	ES
t3.5	Total pronouns	15.87	4.59	164	20.09	3.10	74	*− 8.3	.001	− 1.08
t3.6	Personal pronouns	7.82	3.22	164	10.87	2.96	74	− 6.9	.001	− .99
t3.7	1stPersSng (I)	6.31	3.42	164	7.39	3.19	74	− 2.3	.02	− .33
t3.8	1stPersPlural (We)	.76	1.77	164	1.23	1.86	74	− 1.86	.06	− .26
t3.9	2ndPers (You)	.35	.99	164	1.37	1.37	74	*− 5.76	.001	− .85
t3.10	3rdPerSng (She He)	.10	.42	164	.33	.80	74	*− 2.41	.02	− .36
t3.11	3rdPersPlur (They)	.30	.78	164	.54	.72	74	− 2.3	.02	− .32
t3.12	Impers pronouns	8.10	3.11	164	9.23	2.89	74	− 2.74	.007	− .38
t3.13	Articles	9.75	3.44	164	5.80	2.22	74	*10.58	.001	− .42
t3.14	Prepositions	11.80	3.99	164	11.24	3.60	74	1.04	.30	.15
t3.15	Auxiliary verbs	10.40	3.76	164	10.77	2.69	74	*− .85	.40	− .11
t3.16	Common adverbs	5.59	3.14	164	7.83	2.59	74	− 5.38	.001	− .78
t3.17	Common adjectives	6.12	2.87	164	4.57	2.26	74	*4.50	.001	.60
t3.18	Conjunctions	6.76	3.16	164	10.48	2.90	74	− 8.64	.001	− 1.23
t3.19	Quantifiers	4.41	2.60	164	3.54	2.16	74	*2.70	.008	.36
t3.20	Negations	2.23	1.85	164	1.46	1.10	74	*3.97	.001	.51

*Equal variances not assumed; more conservative values reported

thinking and authenticity were modestly represented, averaging around the mid-point of the scale at 50. The variable emotional tone had the highest mean scores, suggesting that students were generally positive about their work.

Results of Box's test of equality of covariance matrices showed that the observed covariance matrices of the dependent variables were equal across the groups, F (20, 1519) = 1.38, p = .12. The assumption of multivariate normality was met based on

t4.1 **Table 4** Mean LIWC summary variable values, scale = 0 (low) to 100 (high)

		Analytical thinking	Authenticity	Clout	Emotional tone
t4.3	Format	Written (Y1)	55.32 (28.87)	55.06 (28.43)	22.70 (17.98)	46.31 (32.45)
t4.4		Video (Y2)	23.69 (17.70)	47.47 (21.19)	38.47 (21.12)	79.79 (20.20)
t4.5	Projects	Paper craft	39.15 (27.19)	48.99 (26.59)	38.10 (24.67)	71.81 (28.60)
t4.6		Paper circuit	41.96 (25.94)	50.74 (26.95)	24.19 (13.93)	59.02 (32.29)
t4.7		Soft circuit	45.35 (35.57)	56.90 (23.04)	25.08 (19.83)	48.10 (32.95)
t4.8	Grade	Middle school	36.01 (27.98)	53.77 (23.95)	31.71 (20.69)	63.87 (31.86)
t4.9		High school	62.86 (23.79)	45.01 (31.02)	21.15 (19.49)	48.58 (32.03)

Numbers in parentheses are standard deviations

BHEP test results (Baringhaus & Henze, 1988). MANOVA results suggested no statistically significant two-way or three-way interaction effects between assessment format, project type, and grade levels ($p > .05$) but statistically significant main effects of format, Wilk's lambda = 0.82, F (4, 118) = 6.65, $p < .001$, partial $\eta^2 = .18$ (large effect size), project, Wilk's lambda = 0.83, F (8, 236) = 2.80, $p = .006$, partial $\eta^2 = .09$ (medium effect size), and grade, Wilk's lambda = 0.85, F (4, 118) = 5.22, $p = .001$, partial $\eta^2 = .15$ (large effect size), on the dependent variables (i.e. summary variables).

Effect of Assessment Format

Tests of between-subjects effects showed that, when grade level and project were controlled, the differences between written and video formats were in the outcome of clout, F (1, 121) = 4.75, $p = .03$, partial $\eta^2 = .04$ (small effect size), emotional tone, F (1, 121) = 21.50, $p < .001$, partial $\eta^2 = .15$ (large effect size), and analytical thinking, F (1, 121) = 6.53, $p = .01$, partial $\eta^2 = .05$ (medium effect size), but not on authenticity, F (1, 121) = 1.50, $p = .22$, partial $\eta^2 = .01$ (small effect size). Specifically, students in year two (video format) scored higher on clout and emotional tone but lower on analytic thinking.

The data suggest an advantage for video assessment to capture a positive emotional tone, while written assessments about the same projects skewed negative. We suspect the private nature of the written assessments in 3D GameLab between a student and mentor made students more comfortable to talk about problems encountered, whereas students were more likely to project positivity for public consumption in the video format. The aforementioned differences in year one and two assessment monitoring could factor into results as well, since students who were struggling in year one may have sensed pressure to submit a written assessment under monitored conditions, lowering emotional tone scores for that format. In contrast, students expressing lower emotional tone may not have been represented in year two video data if they lacked the engagement to submit an unmonitored assessment to FlipGrid.

The data suggest an advantage for written assessment to capture analytical thinking, adding to the aforementioned findings of longer sentences and bigger words in writing. The lower analytical scores for video suggest that students may be more apt to discuss their experiences, which adds to the aforementioned finding of more narrative-style pronoun use in video. Since students may be more descriptive of experiences in video, and perhaps any help received from peers, this may reflect in the higher clout score reported for video. Video may be a better choice to capture evidence of students working socially ("we," "you") despite the assessment task being an individual reflection, while the written format may better elicit evidence of thinking.

The following excerpts from the same paper circuit project reflect the noted statistical differences with the student in the video saying much more in a conversational and positive tone, but actually analyzing electrical concepts minimally. In contrast, the student in the written assessment is less conversational but describes project specifics. Pseudonyms are applied:

> So there's my card. It says, "great to see you again little unicorn girl." Now let's see what's inside. Wow look at all that! … Yes, so I learned today … get excited about is how the copper actually connects as a circuit and makes the LED light actually work. It's insane. It's amazing too. I am most proud of actually being

successful with this because I didn't think I would do it. But you got to believe. Her, I believe, helped me a lot and the teacher encouraged me some. I'm happy about that and our neighbor for sure. Not my actual next door neighbor but my neighbor in class. I would like to learn more about how the circuitry actually flows through the copper. That's what I'm wondering. I helped a few people by getting them some objects that they needed and explaining how or where they go. Thank you for listening. Goodbye! (Video Assessment, Paper Circuit Project, Analytical Score 11.05, 6th grader/~11 years old, Elaine)

You had to carefully peel the copper tape because if not you would have your tape all curled up and messed up. For this project to work you need one side of the tape to not touch the other side of the tape. This project was simple to understand and works great. In the future I would like to try something new like at home a while back I made a traffic light design. There is an endless possibility and that is what I love about engineering. (Written Assessment, Paper Circuit Project, Analytical Score 65.41, 6th grader/~11 years old, Rabia)

Effect of Project Type

When format and grade level were controlled, the differences between projects were found to be on clout, $F\,(2, 121) = 6.27$, $p = .003$, partial $\eta^2 = .09$ (medium effect size), and emotional tone, $F\,(2, 121) = 4.77$, $p = .01$, partial $\eta^2 = .07$ (medium effect size), but not on analytical thinking, $F\,(2, 121) = 0.65$, $p = .52$, partial $\eta^2 = .01$ (small effect size), or authenticity, $F\,(2, 121) = 1.74$, $p = .18$, partial $\eta^2 = .03$ (small effect size). Post hoc multiple comparisons revealed that the differences between projects on clout and emotional tone were between paper craft and paper or soft circuit ($p < .05$). There was no statistically significant difference between paper circuit and soft circuit projects. Specifically, students who worked on paper craft scored higher on clout and emotional tone than their counterparts who worked on paper circuit or soft circuit. No differences were noticed on analytical thinking or authenticity.

The higher clout score for paper craft projects likely reflects students working together more closely on those activities compared with other projects. The higher emotional tone score for paper craft likely corresponds to their lower difficulty level compared with more challenging paper/soft circuit projects that likewise required modest construction but with the addition of wiring. One might expect students to express more positivity around easier projects that are less frustrating. Emotional tone may help to reveal project types students find approachable and projects that may require support. The following excerpts illustrate a paper craft reflection with a high emotional tone score compared with a soft circuit reflection with a low emotional tone score:

Something that worked well was scoring all lines carefully to ensure neat folding. The project was pretty easy so I didn't find anything that didn't work well. The most difficult part of the project was taping the pieces of paper together to actually form the shape. Sometimes taping one piece would restrict taping the other piece. I overcame that difficulty by planning out how I would tape it before I did, so that I could make sure I had easy access to tape it. I wouldn't change it at all except for the fact that I would like more options to create! (Written

Assessment, Paper Mask Project, Emotional Tone Score 99.0, 10th grader/~15 years old, Lucy)

When completing the quest, some ways I tried to sew the light and battery onto the flowers didn't work so I was a bit perplexed as to how to sew it so I re-read the instructions and did it correctly. The most difficult part of this quest was trying to get the pin through the flowers correctly without poking myself. If I could change my process in the future, I would probably read the instructions more thoroughly. (Written Assessment, Soft Circuit Project, Emotional Tone Score 9.85, 10th grader/~15 years old, Nora)

Effect of Grade Level

When format and projects were controlled, the differences between grade levels were found on analytical thinking only, $F\ (1, 121) = 9.50$, $p = .003$, partial $\eta^2 = .07$ (medium effect size). No statistically significant differences were noted for clout, $F\ (1, 121) = 0.53$, $p = .47$, partial $\eta^2 = .004$ (small effect size), emotional tone, $F\ (1, 121) = 0.14$, $p = .71$, partial $\eta^2 = .001$ (small effect size), or authenticity, $F\ (1, 121) = 2.37$, $p = .13$, partial $\eta^2 = .02$ (small effect size). Specifically, high school students scored higher on analytical thinking than middle school students but not on clout, emotional tone, or authenticity. These findings suggest that the middle and high school groups were somewhat similar, but high school student reflections were more analytical with perhaps more description of maker processes undertaken. The following excerpts illustrate the difference between a high school and middle school student reflection on the same soft circuit project, with more analysis by the high school student:

When completing the quest sewing the light on correctly and making sure the sides matched up worked. Having too much thread did not work well. The most difficult part of the quest was trying to get the light to work after tying it to the battery. I would probably sew the light to a flower or something decorative before starting. Also, space the thread out more. (Written Assessment, Soft Circuit Project, 10th grader/~15 years old, Roberta)

The most difficult part of this quest was finding how to sew everything in. I overcame this by asking for help, and tried different ways. What worked in completing this was trying different ways to get everything on it. In the future I would change the way I did this by paying closer attention to the directions.(Written Assessment, Soft Circuit Project, 6th Grader/~11 years old, Allie)

Discussion and Future Directions

This study applied linguistic analysis of makerspace assessments to inform the presence of TLDF tinkering dimensions with statistical evidence of the conditions under which these dimensions may be elicited. In discussion, we reflect on the four summary variables and the implications of findings for structuring more ideal makerspace assessment.

Reflections on the Four Summary Variables and Related TLDF Dimensions

LIWC analytical thinking means were generally below 50 (on a scale to 100) with large standard deviations suggesting that not all students exhibited analytical thinking around maker projects, and the related TLDF dimension "development of understanding" was somewhat limited. We see some evidence in this study that experience may influence analytical scores since high school students had significantly higher scores than middle school students. Also, the significantly greater presence of analytical thinking in written assessments compared with video might suggest that writing is a more appropriate choice if makerspace leaders wish to elicit more formal and analytical student reflections. The video format likely elicited a more narrative style found to be associated with lower analytical thinking scores (Pennebaker et al., 2014). The following examples illustrate two students who scored high and low on analytical thinking on the soft circuit project, with the high scorer describing a realization about aligning positive/negative ends of batteries/LEDs while the low scorer avoids discussing project specifics:

> The first 4 times I tried this quest I didn't realize that I needed to sew the light to the flower and battery, while matching the negative sides of the battery. Also to match the positive sides. The most difficult part of this quest was having patience, I almost gave up on this project after my third try, but I guess 5's the charm. And I kept trying to complete my quest, and I finished. I will read through all instructions, before starting my quest. (Written Assessment, Soft Circuit Project, Analytical Thinking Score 73.69, 6th Grader/~11 years old, Sia)
>
> I think that this is the easiest thing I have done in my life very easy and simple. I like it. I think understanding everything it was good and easy but the description for me was almost not understandable. I think I should work on keep reading until I can understand it but other than that I am great everything was very smooth and quick. (Written Assessment, Soft Circuit Project, Analytical Thinking Score 2.01, 6th Grader/~11 years old, Patricia)

LIWC authenticity means fell near the middle of the summary variable scale with large standard deviations, suggesting that linguistics in some texts were reflective of being "personal, humble, and vulnerable" and linguistics in other texts were not (Pennebaker Conglomerates, 2018). No significant differences were noted in authenticity scores between assessment formats, project types, or grade levels, so authentic/inauthentic talk was just as likely to be found in written or video assessments, different maker projects, and different age groups. This authenticity measure may be useful to identify individual students with consistently lower scores across a set of maker projects whose more inauthentic language in assessment (e.g. fewer self-references, more negative words, and fewer exclusive words used in description) (Newman et al., 2003) could suggest that they are attempting to pass off understanding where it is lacking, or they lack initiative and intentionality per the aligned TLDF dimension and require further support. The following examples illustrate two students who scored high and low on authenticity on the paper circuit project. The high scorer humbly referenced project difficulties and help seeking behavior compared with the low scorer who postured impersonally that everything "worked well":

> I did good with actually making what I wanted to put on it, but I had trouble cutting out the hole for the light. At first I really did not know how I was going to make the hole for the light, so I asked around on how to do it and they said to use a scorer. I would say that you could plan everything out in pencil and then make sure it's how you like till you cut the hole or do the marker. (Written Assessment, Paper Circuit, Authenticity Score 87.51, 8th Grader/~13 years old, Seema)
> Drawing in pencil worked well and tracing it with marker worked, but the paint did not. The most difficult was trying to use the paint because it was so thick so I had to be very careful. I would say to make it as a poster, and also use a bigger battery. (Written Assessment, Paper Circuit, Authenticity Score 5.76, 8th Grader/~13 years old, Harriet)

LIWC clout means were the lowest of all the summary variables, suggesting that language in assessment was generally very self-oriented which is perhaps not surprising since students were reflecting individually on their projects in these assessments and would be expected to use first-person singular pronouns "I." A small effect of assessment type favored video where students exhibited more clout, suggesting that video may be preferable for eliciting any evidence of the "other-oriented" TLDF dimension social scaffolding since students were describing individual projects in both written/video platforms yet still managed more "we" and "you" pronoun use in video (Kacewicz et al., 2014, p. 137). The finding that students expressed more clout when describing paper craft projects likely reflects students working together more closely on those projects. While clout scores were generally low indicating the TLDF tinkering dimension, social scaffolding was limited in this space ("requesting help," "using/modifying others' ideas," "connecting work") (Bevan et al., 2015, pp. 7–8), words such as "we" and "peer" in the following excerpts with medium clout scores signal students working with peers and mentors in the space:

> ... We thought the light was broken and that wasn't the problem, but when we changed the battery and that was the problem the battery was dead. (Written Assessment, Soft Circuit, Clout Score 43.48, 6th Grader/~11 years old, Anna)
> ... It was also pretty difficult to make the letters interconnected and smoothly connected. I overcame it by asking my peers and the leader for recommendations and choosing the best one to fix it. (Written Assessment, Paper Circuit Project, Clout Score 30.71, 10th Grader/~15 years old, Kate)

LIWC emotional tone means were among the highest across the summary variables suggesting that students generally had a positive reaction to their makerspace projects, or engagement was relatively high per the aligned TLDF dimension. A large effect size was noted for assessment type on emotional tone with positive tone found to be much higher in video. This finding could suggest that students were led to express more positivity in their public, social space, relative to their private, written space that only they and the teacher could access. Students may be more willing to detail project challenges and setbacks in private or when directed at adult mentors rather than their peers to whom they might rather convey competence. It is also possible that students who were more successful and positive were more likely to post in the social FlipGrid space, as reflection in this space was encouraged but not required, and students who

were struggling or more negative about projects in general may not have been as motivated to reflect in FlipGrid. The finding that students expressed more positivity around paper craft projects compared with paper/soft circuit projects likely reflects the easier nature of the paper craft tasks. Also, since our prompts for assessment asked students to detail what worked and did not work, and to describe challenges faced, it is reasonable to assume that more critical language was generated about more complex projects which may have decreased tone scores for the more challenging project types. The following examples illustrate students exhibiting positive and negative emotional tone on the same soft circuit project. The lower-scoring reflection refers several times to what "unfortunately" "did not work" and "challenges:"

> This is my pumpkin. What I'm most excited about with starting this project is the little face on it because I like things like that. And I would like to learn more about how circuits work, big lights at the barn, the street lights they hang over and they're super big. I would like to know how they hook up. Do they go underground? Do they above ground? And what worked well and not so well was when I did the battery. I kept on trying to tape it and the tape wasn't insulated so it didn't work so well or it didn't read the circuits. And so I had to take the tape off and get a new battery and use tiny pieces of tape and fix it.... (Video Assessment, Soft Circuit Project, Emotional Tone Score 96.09, 6th Grader/~11 years old, Rachel)
>
> Today we made little LED light circuitry LED lights. So, unfortunately mine didn't work. So it's supposed to light up but it didn't. So, how these are made is there's a battery and this is just to put on to your clothes or whatever. And what you're, what's happening here is the thread is actually thread that's circuit so it acts almost like the copper foil that we used the last time we did a circuitry project. So, and then it's attached to the light and then it makes it to light up. ... I want to learn about how you can make light up dresses which I thought were cool. So, what really didn't work? What didn't work for me was actually trying to sew. So, what didn't work was trying to sew these together. And what did work was asking someone else to help and, um, actually trying to sew it. What, um, what was challenging was getting the light to work. And I overcame that by asking for help. (Video Assessment, Soft Circuit Project, Emotional Tone Score 41.71, 6th Grader/~11 years old, Dray)

In summary, the LIWC software provided an efficient means of examining transcribed student documentation to reveal embedded characteristics or summary variables. We noted that students were largely positive regarding their maker projects, although documentation would indicate not everyone was analytical about their work, some inauthentic in attempting to pass off understanding, and project talk very self-oriented. The LIWC summary variables helped to provide indirect evidence of the TLDF dimensions including good engagement but limited development of understanding, a mix of initiative and intentionality with some students more motivated to complete project work than others, and limited evidence of social scaffolding. Future researchers might look to develop an instrument to more directly measure TLDF dimensions, although certain dimensions such as initiative and intentionality and social scaffolding may be challenging to capture outside of observation or detailed examples captured as artifacts or audio or video recordings.

Leveraging the Relative Advantages of Different Assessment Formats

Given the noted linguistic analytics and what they were able to inform about makerspace outcomes of interest, is one assessment format (written or video-recorded) preferable over the other? The study findings are mixed with some evidence that written assessment may elicit more formal, analytical thinking coupled with some indicators of that analytical thinking: more words per sentence, more words greater than six letters, more usage of descriptive adjectives and categorical articles. Lievens and Sackett (2006) similarly reported written tests to be more predictive of cognitive abilities compared with video-based tests. In contrast, video assessment was found to elicit more clout or "other-oriented" talk related to social scaffolding, more positive tone suggestive of engagement, and posts that included higher word count and greater use different pronouns suggestive of perhaps less analytical and more conversational speak.

Makerspace researchers and funders are likely to be interested in all of these outcomes, so there is merit to the two approaches. Given recent interest in promoting computational thinking, assessment formats that can elicit and document student thinking will be of interest. Given continuing interest among informal learning leaders to foster a positive, supportive climate and encourage persons to enter STEM-related college and career tracks, video assessment formats that may promote social ties and engagement will also be of interest. Halverson, Kallio, Hackett and Halverson (2016) note that makerspaces are an affiliation-type participatory culture through which one can "grow new interests based on the social connections made in the culture" and "interact with others who share interests" (p. 3). Hence, video assessment is a good fit for community-oriented makerspaces, particularly when encouraging students to not just post but also to review and reply to others' videos. There is value in students using social media to post claims and arguments, but also to question and collaborate (Craig-Hare et al., 2017). Reviewing peer assessments in project areas one has not attempted yet could generate questions or encourage students to try new projects.

Given the relative advantages to the written and video-recorded assessment formats, what opportunities exist to combine the two? The solution could be as simple as asking students to write a script about their makerspace project to record in a video. Murphy and Barry (2016) captured student video presentations on a group wiki and then required them to self-reflect on that video in writing. Portfolio systems touted for makerspaces could likely support similar steps (Peppler et al., 2018). The challenge, however, is student resistance to assessment in informal settings as we experienced in both years of this study, and "balancing automated and manual documentation with (the) least disruption of making" (Litts et al., 2016, p. 1046). It is difficult enough to encourage one form of assessment when students resist breaking from their making, let alone asking for two. If students were working on computers, the occasional prompting electronic scaffold might "stimulate reflection" or provide "peripheral information" on interaction in a system as a reminder to reflect (Glahn, Specht & Koper, 2008), but given most making happens offline, such reminders may be out of reach.

An approach that would not double the assessment load is to retain reflective writing that may be more analytical as desired, and layer in a social element. A discussion forum, for example, would still prompt individual written reflections, but with the added visual affordance of attachments, and the added social affordance of a public space for peers to see and comment on reflections. Socially oriented assessment would

allow students to begin to articulate their "narrative" (Pinkard et al., 2017), present evidence, and ask and respond to questions. For makerspace assessment where students often present solutions to design challenges of personal interest, they could be prompted to talk about their backgrounds and what led them to address a particular problem. As Sias, Wilson-Lopez and Mejia (2016) note, teachers can draw on students' non-academic "funds of knowledge" in connecting their backgrounds to engineering design in such areas as recreation, workplace skills, and household maintenance (p. 31).

Another approach that would not double the assessment load is to task student teams with preparing written assessments together, enhancing social interaction. Keune and Peppler (2017) analyzed different makerspace portfolio entries and reported that portfolios capturing "shared projects and documentation" were "richer" and showcased better "social engagement" (p. 547). Socially engaged approaches may help to foster interest in Science, Technology, Engineering and Mathematics (STEM), a goal of many informal programs including our own. Pinkard et al. (2017) note that "interest development, particularly for marginalized and stereotyped youth, is not simply an individual accomplishment or discrete activity but a social and interactional process that is often mediated by how students perceive the valued ways of knowing and being of a given practice or discipline" (p. 481). Future research could investigate these two modified approaches to written assessment with a social element to determine how moving written assessment into a public space or encouraging group writing impacts the linguistic results seen in this study.

One final recommendation for future research is to consider the approach to assessment itself. Our prompted approach (i.e. show your finished product and answer questions) is one option, but students could also be tasked with reflecting on their making in progress and lessons learned as they worked toward a finished product. For example, Steier and Young (2016) had students write about their processes in makerspace journals. If a makerspace had enough mentors to support assessing processes, "one-on-one teacher-student feedback conversations" could be recorded as employed by Van der Kleij, Adie and Cumming (2017, p. 1096). Peppler, Keune, and Chang (2018a) tout open portfolios where "what is documented is the journey–the makers' process and the outcomes of that process" (p. 16). Students can reflect on process at the end of a task, but "turns they took, decisions they made, challenges they faced, and mistakes they confronted" may be more fully captured in a recurring manner (p. 18). While not completed in the context of a makerspace, Speed et al. (2018) and Tierney et al. (2014) used video reporting to document students working on prompted science experiments and exercises. All of these assessment types better capture processes over time compared with a one-time reflection at the end of a project, and may better reveal cognitive processes and developing understanding.

The findings of this context-dependent, exploratory case study hint at potential advantages to elicit both written and video-recorded assessments in informal learning settings. Tentative hypotheses should be confirmed by future research: written assessments better support the analytical deconstruction of maker project specifics, while video assessments shared in a social forum better support social, positive expression of developing STEM understanding. As noted in limitations, it is unclear what impact if any of the private, monitored nature of written documentation versus the semi-public, unmonitored nature of video-based documentation might have had on linguistics with a need to untangle these effects in a non-exploratory study. Future research could also

investigate impacts on linguistics given reflective writing for social spaces or in social groups, or reflections captured over the course of a particular design process.

Funding Information This research was funded by a grant from the Burroughs-Wellcome Fund's Student Science Enrichment Program (SSEP).

References

Baringhaus, L., & Henze, N. (1988). A consistent test for multivariate normality based on the empirical characteristic function. *Metrika, 35*(1), 339–348. https://doi.org/10.1007/s00362-002-0119-6.

Becker, K., & Nicholson, S. (2016). Gamification in the classroom: Old wine in new badges. In K. Schrier (Ed.), *Learning, education and games* (pp. 61–86). Pittsburgh, PA: ETC Press.

Bevan, B., Gutwill, J. P., Petrich, M., & Wilkinson, K. (2015). Learning through STEM-rich tinkering: Findings from a jointly negotiated research project taken up in practice. *Science Education, 99*(1), 98–120. https://doi.org/10.1002/sce.21151.

Bevan, B., Ryoo, J. J., Vanderwerff, A., Wilkinson, K., & Petrich, M. (2017). *Making deeper learners: A tinkering learning dimensions framework v. 2.0*. San Francisco, CA: Exploratorium and Research + Practice Collaboratory.

Brody, M., Bangert, A., & Dillon, J. (2007). *Assessing learning in informal science contexts*. Washington, DC: Center for the Advancement of Informal Science Education (CAISE).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science, 15*(10), 687–693.

Craig-Hare, J., Rowland, A., Ault, M., & Ellis, J. D. (2017). Practicing scientific argumentation through social media. In I. Levin & D. Tsybulsky (Eds.), *Digital tools and solutions for inquiry-based STEM learning* (pp. 82–111). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-5225-2525-7.

Educause. (2013). Seven things you should know about... makerspaces. Educause Learning Initiative. Retrieved from: https://net.educause.edu/ir/library/pdf/ELI7095.pdf.

Erdmann, M. A., & March, J. L. (2014). Video reports as novel alternate assessment in the undergraduate chemistry laboratory. *Chemistry Education Research and Practice, 15*(4), 650–657. https://doi.org/10.1039/C4RP00107A.

Glahn, C., Specht, M. & Koper, R. (2008). Supporting reflection in informal learning. In *Proceedings of the ECTEL Doctoral Consortium '08*. Maastricht, The Netherlands.

Hagel, J., Brown, J. S. & Kulasooriya, D. (2014). A movement in the making. Westlake, TX: Deloitte Press. Retrieved from: http://dupress.com/articles/a-movement-in-the-making/

Halverson, R., Kallio, J., Hackett, S., & Halverson, E. (2016). Participatory culture as a model for how new media technologies can change public schools. In *WCER working paper 2016–7*. Madison, WI: Wisconsin Center for Education Research.

Herr-Stephenson, B., Rhoten, D., Perkel, D., & Sims, C. (2011). Digital media and technology in afterschool programs, libraries, and museums. In *The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning*. Cambridge, MA: The MIT Press.

Hira, A., Joslyn, C. H. & Hynes, M. M. (2014). *Classroom makerspaces: Identifying the opportunities and challenges*. Paper presented at the Frontiers in Education Conference, Madrid, Spain.

Holloway, C. (2015). Accelerated engagement of African-American males through social media. In M. Simonson (Ed.), *38th annual proceedings of the Association for Educational Communications and Technology* (pp. 103–104). Bloomington, IN: AECT.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology, 33*(2), 125–143.

Keune, A., & Peppler, K. (2017). Maker portfolios as learning and community-building tools inside and outside makerspaces. In *Proceedings of the 2017 Computer Supported Collaborative Learning (CSCL) Conference* (pp. 545–548). Philadelphia, PA: ISLS.

Kurti, R. S., Kurti, D. L., & Fleming, L. (2014). The philosophy of educational makerspaces: Part 1 of making an educational makerspace. *Teacher Librarian, 41*(5), 8–11.

Lemke, J., Lecusay, R., Cole, M., & Michalchik, V. (2015). *Documenting and assessing learning in informal and media-rich environments*. Cambridge, MA: MIT Press.

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology, 91*(5), 1181–1188.

Litts, B. K., Kafai, Y. B., Fields, D. A., Halverson, E. R., Peppler, K., Keune, A., et al. (2016). Connected making: Designing for youth learning in online maker communities in and out of schools. In *Proceedings of the 12th International Conference of the Learning Sciences (ICLS), Vol. 2* (pp. 1041–1047). Singapore: ISLS.

Marti-Parreno, J., Mendez-Ibanez, E., & Alonso-Arroyo, A. (2016). The use of gamification in education: A bibliometric and text mining analysis. *Journal of Computer-Assisted Learning, 32*(6), 663–676. https://doi.org/10.1111/jcal.12161.

Moore, R. L., Oliver, K. M., & Wang, C. (2019). Setting the pace: Examining cognitive processing in MOOC discussion forums with automatic text analysis. *Interactive Learning Environments, 27*(5–6), 655–669. https://doi.org/10.1080/10494820.2019.1610453.

Murphy, K., & Barry, S. (2016). Feed-forward: Students gaining more from assessment via deeper engagement in video-recorded presentations. *Assessment and Evaluation in Higher Education, 41*(2), 213–227. https://doi.org/10.1080/02602938.2014.996206.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin, 29*(5), 665–675.

Oliver, K., Moore, R. L., & Evans, M. A. (2017). Establishing a virtual makerspace for an online graduate courses: A design case. *International Journal of Designs for Learning, 8*(1), 112–123. https://doi.org/10.14434/ijdl.v8i1.22573.

Pennebaker Conglomerates. (2018). Interpreting LIWC Output. Retrieved from: http://liwc.wpengine.com/interpreting-liwc-output/.

Pennebaker, J., Chung, C., Frazee, J., Lavergne, G., & Beaver, D. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS One, 9*(12), 1–10. https://doi.org/10.1371/journal.pone.0115844.

Pennebaker, J. W., Boyd, R. L., Jordan, K. & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin. Retrieved from https://repositories.lib.utexas.edu/handle/2152/31333.

Peppler, K., Keune, A. & Chang, S. (2018a). Introducing phase 2 of the open portfolio project: Assessment in makerspaces (Research Brief #11). Retrieved from http://makered.org/opp/publications/.

Peppler, K., Keune, A., Xia, F. & Chang, S. (2018b). Survey of assessment in makerspaces. MakerEd open portfolio project (Research Brief #17). Retrieved from: http://makered.org/opp/publications/.

Pinkard, N., Erete, S., Martin, C. K., & McKinney de Royston, M. (2017). Digital youth divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities. *Journal of the Learning Sciences, 26*(3), 477–516. https://doi.org/10.1080/10508406.2017.1307199.

Sheridan, K. M., Halverson, E. R., Litts, B. K., Brahms, L., Jacobs-Priebe, L., & Owens, T. (2014). Learning in the making: A comparative case study of three makerspaces. *Harvard Educational Review, 84*(4), 505–531. https://doi.org/10.17763/haer.84.4.brr34733723j648u.

Sias, C., Wilson-Lopez, A., & Mejia, J. (2016). Connecting students' background experiences to engineering design. *Technology and Engineering Teacher, 76*(1), 30–35.

Speed, C. J., Lucarelli, G. A., & Macaulay, J. O. (2018). Student produced videos: An innovative and creative approach to assessment. *International Journal of Higher Education, 7*(4), 99–109. https://doi.org/10.5430/ijhe.v7n4p99.

Steier, L. P., & Young, A. W. (2016). Growth mindset and the makerspace educational environment. In *Masters of arts in education action research papers (#196)*. St. Paul, MN: St. Catherine University.

Tierney, J., Bodek, M., Fredricks, S., Dudkin, E., & Kistler, K. (2014). Using web-based video as an assessment tool for student performance in organic chemistry. *Journal of Chemical Education, 91*(7), 982–986. https://doi.org/10.1021/ed400195c.

Van der Kleij, F., Adie, L., & Cumming, J. (2017). Using video technology to enable student voice in video feedback. *British Journal of Educational Technology, 48*(5), 1092–1105. https://doi.org/10.1111/bjet.12536.

Warne, R. T. (2014). A primer on multivariate analysis of variance (MANOVA) for behavioral scientists. *Practical Research & Evaluation, 19*(17), 1–10.

Yin, R. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.