Enhancing consent forms to support participant decision making in multimodal learning data research

Marc Beardsley, Judit Martinez-Moreno, Milica Vujovic, Patricia Santos, and Davinia Hernández-Leo

Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona

Marc Beardsley *is a PhD student researching the Science of Learning and Learning Design at Universitat Pompeu Fabra, Barcelona.*

Judit Martinez-Moreno *is a PhD student researching Learning Design and the development of 21st Century Digital Skills at Universitat Pompeu Fabra, Barcelona.*

Milica Vujovic *is a PhD student researching Multimodal Learning Analytics and Smart Learning Environments at Universitat Pompeu Fabra, Barcelona.*

Patricia Santos *is a senior researcher in the ICT Department of Universitat Pompeu Fabra, Barcelona.*

Davinia Hernández-Leo *is Full Professor and ICREA Academia Fellow in the ICT Department of Universitat Pompeu Fabra, Barcelona.*

Address for correspondence: Mr. Marc Beardsley, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona. E-mail: marc.beardsley@upf.edu

Abstract

Advances in the field of multimodal learning analytics (MMLA) research is often accomplished by actively exploring new technologies and techniques related to the collection and analysis of data. Exploration of ethical principles and procedures for governing the use of new technologies and techniques, however, is not as readily pursued. As collected data grows in complexity, and invasiveness, potentially, a growing need is arising to scrutinize ethical aspects of MMLA research. In our study, we introduce an informed consent comprehension test for educational technology research and assess the effects of enhancing MMLA consent forms on comprehension of informed consent and on rates of enrollment in a MMLA study. One form is written from a researcher perspective and the other from a participant perspective. Results of the study involving first year undergraduate students suggest that the overall level of comprehension did not differ between conditions. Yet, the participant-oriented consent form resulted in significantly lower rates of enrollment. Implications for MMLA researchers are discussed.

Keywords

Multimodal learning analytics, ethics, data risks, informed decision making, informed consent

Practitioner notes

What is already known about this topic

- Data collected in multimodal learning analytics (MMLA) research is growing in complexity and invasiveness.
- Informed consent is a process to enable individuals to make voluntary decisions about participating in research based on an understanding of a study's purpose, procedures, risks, and benefits.
- Studies in related fields such as Bioethics show that many informed consent processes do not adequately support participant comprehension and decision making.

What this paper adds

- A discussion of the need to scrutinize the ethical aspects of MMLA research with a focus on supporting adequate participant understanding without discouraging participation in research.
- An approach to measure the effects of enhancing MMLA consent forms on comprehension and rates of enrollment in a MMLA research study.

Implications for practice and/or policy

- MMLA researchers may need to determine what level of learner comprehension is necessary for ethical participation in research studies.
- More research is needed to discover a balanced approach that can adequately inform participants without significantly affecting rates of enrollment.
- Further work is needed to establish adequate ethical protocols that can be applied by researchers, policy makers and institutional managers to facilitate a trusted implementation of MMLA.

Introduction and Related Work

Advances in the field of multimodal learning analytics (MMLA) research is often accomplished by actively exploring new technologies and techniques related to the collection and analysis of data. However, the exploration of the ethical principles and procedures for governing the usage of the new technologies and techniques is not as readily pursued. For example, a recent paper examining the body of research emerging from MMLA workshops, published proceedings, and journals did not mention topics related to ethics (Worsley & Martinez-Maldonado, 2018). As MMLA data grows in complexity, and potential invasiveness, a need is arising to scrutinize the ethical aspects of MMLA research.

The process of informed consent may be an appropriate starting point as the results of bioethical research suggest that many informed consent processes do not adequately support participant comprehension of the studies they consent to (Flory & Emanuel,

2004; Falagas, Korbila, Giannopoulou, Kondilis & Peppas, 2009; Nishimura *et al.,* 2013; Tamariz, Palacio, Robert & Marcus, 2013; Tam *et al.*, 2015) and may fail to fulfill the requirements of valid consent. Further, recent societal shifts indicate that a greater onus is being placed on data collectors to adequately support the autonomous decision making of individuals with regard to the giving of consent and sharing of data. These shifts are evidenced by recent revisions to the Common Rule (Department of Health and Human Services et al., 2017) and the enactment of the General Data Protection Regulation (EU General Data Protection Regulation, 2016). Both of which are described later in the paper.

The field of MMLA research could benefit from demonstrating its efforts toward better supporting potential research participants in autonomously making decisions about sharing their data; especially, as both the complexity of the work being done, and potential invasiveness of the data being collected increases. Studies such as this one, attempt to contribute to the demonstration of such efforts. We introduce an informed consent comprehension test for educational technology research and assess the effects of enhancing MMLA consent forms on participant comprehension of informed consent and on rates of enrollment in a MMLA study.

Multimodal Complexity and Invasiveness

MMLA is an elaboration of learning analytics (Bilkstein & Worsley, 2016). Corrin et al. (2019) write that learning analytics "aims to provide meaningful ways of using data to support student learning within learning environments" (p. 7). Whereas learning analytics materialized in online learning environments, MMLA extended the tracking and quantification of learning to offline environments (Ochoa & Worsley, 2016). MMLA can be thought of as the learning traces extracted from log-files or digital documents combined with data from "recorded video and audio, pen strokes, position tracking devices, biosensors, and any other modality that could be useful to understand or measure the learning process" (Ochoa, Lang & Siemens, 2017, p. 129-141).

The adoption of techniques such as machine learning, and text mining paired with the increasing accessibility of collecting and storing massive amounts of data (Blikstein, 2011) has shifted the collection of student data from more discrete, activity-focused exchanges of data to ongoing monitoring both within and outside of the classroom (Beardsley, Santos, Hernández-Leo & Michos, 2019). Along with the increased complexity of the data being collected, the data is becoming more invasive as the collection of electrophysiological data in MMLA research grows. Measures of electrodermal activity (Pijeira-Díaz, Drachsler, Järvelä, Kirschner, 2019), heart rate (Larra *et al.*, 2014) and neural oscillations via an electroencephalogram (EEG) (Sun & Yeh, 2017) are more frequently being used. This biometric data is considered sensitive health data (Chassange, 2017).

Informed or Uninformed Consent

As MMLA research adds layers of complexity to participant understanding of studies; and begins to converge with health research, one can look to the field of bioethics for

ideas about how to face the forthcoming challenges via informed consent. Obtaining informed consent is a key component of upholding the ethical value of participant autonomy (Nishimura *et al.*, 2013) and is a process to enable individuals to make voluntary decisions about participating in research based on an understanding of the purpose, procedures, risks, benefits, and alternatives (Beskow, 2016). It is grounded in the ethical principle of respect for persons (Kass, Taylor, Ali, Hallez & Chaisson, 2015) and aims to respect and promote participants' autonomy and protect them from potential harm (Jefford & Moore, 2008). Obtaining informed consent is "widely regarded as central to ethical social science research practice" (Heath, Charles, Crow & Wiles, 2007, p. 403). In MMLA research, informed consent is regularly obtained via a written consent form signed by the potential research participant. The participant's signature is a visible act of signifying the decision of participating in the research (Alderson & Morrow, 2004). The presentation and signing of the consent form enable research participants to "express their own agency within the research process" – an agency which "arises from their competency at decision making" (Heath, Charles, Crow & Wiles, 2007, p. 404).

Kass, Taylor, Ali, Hallez and Chaisson (2015) argue that informed consent "rests on an assumption that individuals considering research participation have adequately understood the information provided to them" (p. 2). The requirement of understanding is echoed by many (Young, Hooker & Freeberg, 1990; Joffe *et al.*, 2001; Tait, Voepel-Lewis, Robinson & Malviya, 2002; Buccini, Iverson, Caputi, Jones & Gho, 2009; Hallinan, Forrest, Uhlenbrauck, Young & McKinney Jr, 2016; Hadden *et al.*, 2017; Muravyeva, Janssen, Dirkx & Specht, 2018). Buccini, Iverson, Caputi, Jones and Gho (2009) write that "to treat potential research participants as autonomous agents, it is imperative to ensure understanding of the consent information has actually occurred, thereby, enabling them to make autonomous decisions about participation" (p. 7). Wendler and Grady (2008) add that individuals need to comprehend the information that is needed "to determine whether participation in a given study is consistent with their interests" (p. 205). In other words, it is critical to understand "how their prospective experience will differ if they choose to enroll" (p. 207) and unless they do, their consent is unlikely to be valid.

Meta-analyses and systematic reviews from bioethics suggest that many research participants struggle to understand what they are consenting to (Flory & Emanuel, 2004; Falagas, Korbila, Giannopoulou, Kondilis & Peppas, 2009; Nishimura *et al.,* 2013; Tamariz, Palacio, Robert & Marcus, 2013; Tam *et al.*, 2015). In a recent survey related to collecting and storing of biospecimens, Beskow, Lin, Dombeck, Gao and Weinfurt (2017), found that one-third of their survey sample failed to demonstrate adequate comprehension. MMLA research may not reach the level of complexity involved in biomedical research. However, the collection and storage of biospecimens is already underway with the use of biological samples to measure changes in stress response in educational contexts (Schonert-Reichl *et al.*, 2015) and MMLA research is increasingly making use of electrophysiological data. Further, studies on informed consent comprehension related to MMLA research are few and far between but suggest that

participant understanding could be better supported. For example, a recent study by Muravyeva, Janssen, Dirkx and Specht (2018) on informed consent regarding an e-assessment system that used biometric data for identity verification found that up to one quarter of participants did not find the information presented clear enough. Moreover, the results of a study by Beardsley, Santos, Hernández-Leo and Michos (2019) suggest that teacher and learner knowledge of data sharing risks may be deficient thus limiting the effectiveness of commonly used consent forms in being used alone to communicate such risks. Jefford and Moore (2008) observe that current informed consent practices in research seem "to have been shaped by emphasis on the legal duty of disclosure" (p. 486) rather than the ethical duty to inform potential participants. Thus, comprehension may be costly to achieve in terms of effort, time, resources and, possibly, rates of participation in research studies, as improving comprehension may require making changes to how things are currently being done in MMLA research.

Enhancing MMLA Consent Forms

Bioethical studies have shown that enhancing consent forms can improve participant comprehension and contribute to validating the consent received. In a systematic review of informed consent interventions, Flory and Emanuel (2004) found that 6 out of 15 trials of enhanced consent forms showed significant improvements in understanding but the authors raised concerns about the quality of the trials. In a more recent meta-analysis of informed consent interventions, Nishimura et al. (2013) found that 41% of trials of enhanced consent forms led to significant improvements in understanding. Table 1 presents recommendations from biomedical literature toward enhancing consent forms for comprehension.

Based on these bioethical studies, enhancing MMLA consent forms may offer an unburdened approach to improving understanding as it requires few changes to the current practices of MMLA researchers. However, efforts are needed to improve on the success rate of enhanced consent forms and overcome certain challenges presented by MMLA research. For example, grasping what data is being collected and how it can be used often requires a basic knowledge of human psychology, physiology, and even signal processing. As a result, potential research participants may underestimate the risks associated with the data they agree to share as they are unaware of how such data could be used to potentially identify them, their traits (e.g. race, gender, age), and medical conditions (Mordini & Ashton, 2012; Wanlund & Schuurman, 2018). Further, MMLA research not only incorporates terminology from diverse fields but also from new technologies it adopts. As a result, the language used to explain a study can be unfamiliar to participants. To address these issues, efforts should be made to simplify the language used, avoid acronyms and specialized terms commonly used in MMLA, and offer further clarification of concepts that cannot be presented in simpler forms. Finally, the sequence of the information presented should be logical from the point of view of the interests of the receiver (Bjørn, Rossel & Holm, 1999) – this can help participants reach an understanding of what interpretations can be made from their data and better assess the obligations, benefits, and risks of their participation.

Table 1

Recommendations for enhancing consent forms from biomedical literature

Item	Suggestion	References
Reduce required reading level	Target a 9th grade level and use readability checkers to estimate reading level (Jefford & Moore, 2008).	Young, Hooker & Freeberg, 1990; Villafranca, Kereliuk, Hamlin, Johnson & Jacobsohn, 2017
Use simple language	"Modify the vocabulary used, making it more familiar, short, and easy to visualize" (Villafranca, Kereliuk, Hamlin, Johnson & Jacobsohn, 2017).	Young, Hooker & Freeberg, 1990; Bjørn, Rossel & Holm, 1999; Wittenberg & Dickler, 2007; Jefford & Moore, 2008; Hallinan, Forrest, Uhlenbrauck, Young & McKinney Jr, 2016; Kadam, 2017
Use shorter and simpler sentences	Break longer sentences that contain several ideas into shorter sentences that contain only one (Jefford & Moore, 2008).	Young, Hooker & Freeberg, 1990; Bjørn, Rossel & Holm, 1999; Wittenberg & Dickler, 2007
Shorten blocks of text and explanations	Keep paragraph length below seven lines (Kadam, 2017).	Bjørn, Rossel & Holm, 1999; Jefford & Moore, 2008; Lorenzen, Melby & Earles, 2008; Villafranca, Kereliuk, Hamlin, Johnson & Jacobsohn, 2017
Bold section headings	Describe information on types of data in a separate paragraph, under a separate header, to attract proper attention (Muravyeva, Janssen, Dirkx & Specht, 2018).	Bjørn, Rossel & Holm, 1999; Lorenzen, Melby & Earles, 2008; Manta, Ortiz, Moulton & Sonnad, 2016
Include bulleted lists, graphics, lists, summaries	Use bullet points to break-up long explanations (Jefford & Moore, 2008).	Wittenberg & Dickler, 2007; Lorenzen, Melby & Earles, 2008; Kass, Taylor, Ali, Hallez & Chaisson, 2015; Manta, Ortiz, Moulton & Sonnad, 2016
Use more white space and line spacing	"To make the forms more readable, both high and low literacy patients asked for more white space" (Lorenzen, Melby & Earles, 2008).	Wittenberg & Dickler, 2007; Villafranca, Kereliuk, Hamlin, Johnson & Jacobsohn, 2017
Organize information based on relevance to participant	Restructure information into a sequence that is logical as seen from the point of view of the receiver (Bjørn, Rossel & Holm, 1999).	Young, Hooker & Freeberg, 1990; Tait, Voepel-Lewis, Robinson & Malviya, 2002; Kass, Taylor, Ali, Hallez & Chaisson, 2015; Hallinan, Forrest, Uhlenbrauck, Young & McKinney Jr, 2016; Dranseika, Piasecki & Waligora, 2017; Karbwang *et al*., 2018

Ethical Responsibilities

Recent societal shifts, as evidenced by revisions to the Common Rule and enactment of the General Data Protection Regulation (GDPR), suggest a greater onus is being placed on data collectors to adequately support the autonomous decision making of individuals with regard to giving consent and sharing data. For example, the Common Rule which is the "overarching policy governing research with human subjects conducted and supported by most federal agencies and departments in the United States" (p. 22) strongly emphasizes "efforts to promote understanding and comprehension during the consent process" (Sugarman, 2017, p. 23). The GDPR defines consent as any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she signifies to the processing of personal data relating to him or her. Further, the notion of imbalance between the controller and the data subject is also taken into consideration by the GDPR (Article 29 Working Party 2018, p. 5). In the field of educational technology research, recent articles have discussed ethical and privacy issues related to the usage of learning analytics at various scales from teacher-led classroom usage (Rodríguez-Triana, Martínez-Monés & Villagrá-Sobrino, 2016) to institutional usage of learning analytics (Pardo & Siemens, 2014). In a European Commission publication (2014) on learning and teaching in higher education, the requirement of student consent with regard to learning analytics was put forth as "the full and informed consent of students must be a requirement" (p. 50).

This study attempts to demonstrate efforts in MMLA research toward better supporting potential research participants in autonomously making decisions about sharing their data; and is situated in the understanding that researchers face challenges and costs in having an acceptable number of participants for their research. Our research question is: **How can MMLA researchers comply with the obligation of ensuring adequate participant understanding without discouraging participation in research?**

To explore this question, we assess the effects of enhancing MMLA consent forms on comprehension of informed consent and on rates of enrollment in a MMLA study. Two enhanced consent forms are used. One consent form is written from a researcher perspective and the other from a participant perspective. We hypothesize that:

(1) *The rates of enrollment will be the same for both enhanced MMLA consent forms* as in the previously mentioned meta-analysis by Nishimura *et al.* (2013), the authors concluded that "there is little evidence that a participant's satisfaction or a study's accrual rates would be negatively altered by attempts to improve the informed consent process, which should be reassuring to investigators" (p. 12).

(2) *The participant-oriented consent form will lead to better comprehension of informed consent* as the information is presented in a manner that better aligns with the decision under consideration by the potential participant.

Methods

Participants

In total 201 first-year university students enrolled in a computer engineering course at a Spanish university were eligible to participate in the study. The course was offered either in English or Spanish/Catalan. Of the 201 eligible participants, 13 were absent, and 6 arrived late to the session and did not sign a consent form. Their data has been removed. Thus, a total of 182 students were potential participants in the study, 97 in the researcher-oriented condition; and 85 in the participant-oriented condition.

Materials

A Classroom Lesson on Data Sharing Risks

A two-hour classroom lesson on data sharing benefits and risks was used as formative material to present the main topics covered by the study to the students. A previous publication (Beardsley, Santos, Hernández-Leo & Michos, 2019), provides an outline of the lesson content.

Enhanced MMLA Consent Forms

The enhanced MMLA consent forms were constructed following best practices from bioethics with one form written from a researcher-orientation and the other from a participant-orientation utilizing a question and answer format. The best practices were derived from research articles and fall into two broad categories: (1) improving readability via the formatting of the text, the language used, and the length of the form; (2) improving the relevance of the text to the participant via the selection of content and perspective from which the text is written. Both enhanced MMLA consent forms comprised a two-page information sheet and a single consent page (see materials in Zenodo; Beardsley, Vujovic, Martinez-Moreno, Santos & Hernández-Leo, 2019). The information sheet but not the consent page differed between conditions. With regard to improving readability, both consent forms were enhanced by applying the items highlighted in Table 1. Readability was evaluated using the Flesch reading ease (Kincaid, Fishburne Jr, Rogers & Chissom, 1975) and the Flesch–Kincaid grade level (Flesch, 1948) measures which are available in Microsoft® Word and have been used in similar studies evaluating readability of consent forms (Paasche-Orlow *et al.*, 2003; Fortun, West, Chalkley, Shonde, & Hawkey, 2008; Jefford & Moore, 2008). The readability of the Spanish versions of the documents was assessed with a scale proposed by Barrio-Cantalejo *et al.* (2008) and is an adaptation of the scales proposed by Flesch and Szigriszt (Flesch-Szigriszt Index). A comparison of the information sheets is shown in Table 2 (English) and Table A1 (Spanish).

The researcher-oriented form included brief summaries under each subheading title. Whereas, the participant-oriented form used questions as subheadings (see Table 2). With regard to improving the relevance of the text to the potential participant in the participant-oriented consent form, the information was restructured into a sequence that is logically seen from the point of view of the interests of the receiver (Bjørn, Rossel & Holm, 1999; Dranseika, Piasecki & Waligora, 2017). For example, elements that participants identified as being most important such as major risks (Karbwang *et al.*, 2018) and

obligations (Wendler & Grady, 2008) were moved to appear earlier in the sequence of information presented.

Table 2

A comparison of the English MMLA consent form information sheets

Variables	A. Researcher-oriented consent form	B. Participant-oriented consent form
Word count	1,505	1,394
Flesch Reading Ease Score	54.1	54.4
Flesch-Kincaid Grade Level	9	8.9
Passive Sentences	35.2%	32.9%
Perspective	Researcher	Participant
Subheadings in order	1. Motivation and Objectives 2. Methodology 3. Collected Data 4. Risks and Privacy 5. Benefits 6. Voluntary Participation 7. Data Subject Rights	1. If I participate in the study, what will I be asked to do? 2. If I participate in the study, what are the risks? 3. Can I trust you with my data? 4. Why do you want to collect my data? 5. Do I benefit from participating in the study? 6. What if I change my mind? 7. What if I feel my rights have been violated?

An Informed Consent Comprehension Test for Educational Technology

To assess comprehension of informed consent in the MMLA study, we created a comprehension test for potential participants. The test is an adaptation of similar tests in bioethics such as the Quality of Informed Consent (QuIC) measure (Joffe *et al*., 2001), Informed Consent Questionnaire (Guarino *et al.*, 2006), and the Brief Informed Consent Evaluation Protocol (Sugarman *et al.*, 2005). Our approach follows closely that of Joffe et al. (2001) in assessing both subjective and objective understanding of participants but differs in that it integrates questions from all three tests previously mentioned and adds questions to comply with GDPR requirements on disclosure related to data collection. The GDPR disclosure requirements incorporated include clearly identifying the purpose, type of data, risks of data transfer, the identity of the data controller, and the existence of the right to withdraw consent (Article 29 Working Party 2018, p. 13). The approach to measure both subjective and objective understanding was taken as bioethical researchers have argued that individuals contemplating participation "should both *be* well informed and *feel* well informed about the study under consideration" (Guarino *et al.*, 2006, p. 140).

Part A of the test measures subjective understanding of informed consent (see Table B1), whereas Part B measures participants' objective understanding (see Table B2). The grouping of questions in Part B are based on constructs taken from GDPR's definition of consent (EU General Data Protection Regulation, 2016). The B1 grouping of items relates to consent being freely given (e.g. consent is given without coercion or pressure); B2 relates to consent being specific (e.g. the obligations, expectations, and procedures are clear); B3 relates to consent being informative (e.g. the purpose, risks, and benefits of participation are clear); and B4 relates to the identity of the controller (e.g. the participant is aware of who to contact for the research and their subject rights).

As with the QuIC measure, a 3-point scale was used for both the subjective (Part A) and objective (Part B) sections of the test. In the pilot testing of the QuIC a 3-point scale was deemed more appropriate as "intensity of agreement did not seem meaningful for statements of fact" (Joffe *et al.*, 2001). Further, best practices from similar measures created in biomedical research were followed such as varying the direction of the statements to avoid agreement bias (Joffe *et al.*, 2001; Beskow *et al.*, 2017); including a neutral option of 'not sure' to reduce participant guessing (Joffe *et al.*, 2001; Beskow *et al.*, 2017); and devising objective scoring algorithms to prevent investigator bias (Joffe *et al.*, 2001). For the scoring of Part A, 1 point was assigned for a positive subjective evaluation (i.e. Yes), 0 points for a neutral answer (i.e. Not sure), and -1 points for a negative subjective evaluation (i.e. No). For the scoring of Part B, 1 point was assigned for a correct answer, 0 points for a neutral answer (i.e. I'm not sure), and -1 point for an incorrect answer. Participants completed the test online via a Google Form. Both Spanish and English versions of the test were created and reading levels assessed (English: Flesch reading ease 61%, Flesch–Kincaid grade level of 8.1; Spanish: Flesch-Szigriszt grade level 54.87).

Design

A quasi-experimental design with two conditions was used to assess the effects of the type of enhanced MMLA consent form on rates of enrollment in the study and on participant comprehension of informed consent. This study was conducted in conjunction with a separate MMLA research study which involved the use of an online collaborative learning application and the collection of different types of data (observations, online artefacts, survey responses, video and audio recordings). Six separate groups (i.e. classes) were eligible to participate in the study, 2 of which were in English, and 4 of which were in Spanish. Half of the groups (2 Spanish and 1 English) were given an enhanced MMLA consent form written from a researcher perspective (researcher-oriented). The other half of the groups were given an enhanced MMLA consent form written from a participant perspective (participant-oriented).

Procedure

As the study took place in an educational setting, we adopted a formative approach that treated consent as an ongoing process. Heath, Charles, Crow & Wiles (2007) write, "process consent provides a useful mechanism for updating participants involved in

studies with emergent research designs, and allows existing participants to decide whether or not to remain involved" as consent "should not be assumed on the basis of initial consent only" (p. 409). At the start of the class, students were presented with the consent form and indicated their initial consent decision, they then took a comprehension test on informed consent, and finally completed a classroom lesson on data sharing risks (Figure 1). At the end of the course, once the study had been completed, students read a debriefing of the study and reviewed their consent. Further, on each data collection instrument, students were asked to explicitly mark whether they wanted the data to be shared with researchers or not.

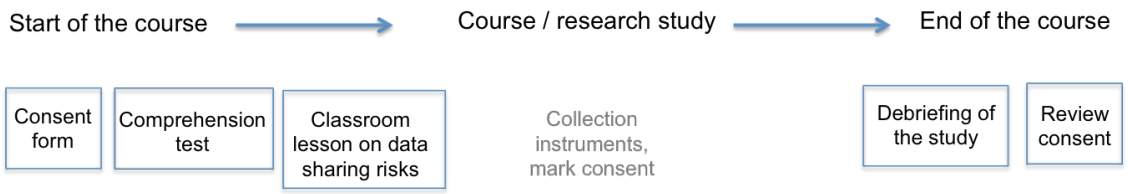

Figure 1. Schema of the procedure

Analyses

To assess the effects of the type of consent form on study enrollment rates, the rate of enrollment was calculated by using the number of consenting participants and number of eligible participants per condition. To assess the effects of the type of consent form on participant comprehension of informed consent, two steps were taken. The first step involved conducting a factor analysis to assess the validity of the variables (i.e. the questions in the comprehension test). The second step of the analysis consisted of running t-tests to investigate the difference in comprehension test performance between conditions. The factor analysis was performed on the grouping of questions (A1, A2, B1, B2, B3 and B4 as shown in Appendix B). All statistical analyses were performed using SPSS v.23.0 (IBM Corp., Armonk, NY, USA).

Results

(1) The rates of enrollment will be the same for both enhanced MMLA consent forms. In total 182 participants completed the MMLA consent forms with 134 consenting to the study (M = 73.63%). In the researcher-oriented condition, 83 out of 97 (M = 85.57%) eligible participants consented to participating in the study compared to 51 out of 85 (M = 60%) in the participant-oriented condition. There was a statistically significant difference between conditions ($p < 0.001$, two-tailed Fisher's exact test, Cramer's V =0.289).

(2) The participant-oriented consent form will lead to better comprehension of informed consent. Firstly, the factors analysis could not be run on the B1 grouping of questions due to the nature of the answers within it but was run on the other groupings. The analysis (see Appendix C) showed that 61.7% of variance or less was explained by components (i.e. the questions in the grouping A1, A2, B2, B3, and B4). As the size of R^2 coefficients did not differ greatly among questions in each grouping the initial groupings could be

maintained (Tables C1.2, C2.2, C3.2, C4.2, C5.2) and a comparison of comprehension scores using all questions could be conducted.

As a process consent approach was followed, participants were afforded the opportunity to reconsider their consent decision within the informed consent test form. In total, 162 participants consented to sharing the data from their tests with 7 participants withdrawing their consent and 35 changing from not consenting to consenting. In the researcher-oriented condition, 89 (80 yes, 9 no) consented to sharing their results. In the participant-oriented condition, 73 (47 yes, 26 no) consented to sharing their results. In comparing comprehension test scores by condition, no difference was detected in the scores on the subjective portion (Part A) of the informed consent comprehension test ($M_{researcher}$ = 5.64, SD = 1.89, 95% CI [−1, 7] ; $M_{participant}$ = 5.29, SD = 2.34, 95% CI [−5, 7]; U = 3013, p = .404, two-tailed Mann-Whitney U Test, 95% CI [−5, 7], d = 0.272). Further, no difference was detected between conditions in the scores for the objective portion (Part B) of the informed consent comprehension test ($M_{researcher}$ = 10.97, SD = 3.54, 95% CI [4, 19]; $M_{participant}$ = 11.4, SD = 3.61, 95% CI [1, 18]; U = 2928.5, p = .279, two-tailed Mann-Whitney U Test, 95% CI [1, 19], d = 0.314).

Per question comparison between conditions

In exploring the data further, we compared the responses to each question between conditions and found that there was a significant difference on the following questions: *B2.3. The study will not collect my personal data.* ($M_{researcher}$ = -0.067, SD = 0.94; $M_{participant}$ = 0.52, SD = 0.80; p < .001, Test, 95% CI [-1, 1]). *B2.4. Because I am participating in a research study, it is possible that others not directly involved in my education may access my data from this class* ($M_{researcher}$ = 0.00, SD = 0.96; $M_{participant}$ = 0.53, SD = 0.78; p < .001, Test, 95% CI [-1, 1]). The results of the comparison of each question can be found in Table A1 and Table A2.

Discussion

Contrary to the findings of Nishimura et al. (2013) and Hallinan, Forrest, Uhlenbrauck, Young & McKinney Jr (2016) the type of consent form affected the rates of enrollment in the MMLA study. The participant-oriented forms which reflected our efforts to better support potential participants' decision making via the consent form resulted in a much lower rate of enrollment in the study. The cause of the lower rate of enrollment in the participant-oriented group is unclear. A possible explanation is that those receiving the participant-oriented forms were more aware of the risks of the study as reflected by the differences between conditions on the questions related to risk (B.2.3 and B.2.4). Question B2.3 related to whether participants understood that personal data would be collected. This question is a gauge for participant understanding of risk as personal data is commonly understood to be riskier than non-personal data for individuals. Question B3.4 referred to whether identifiable data would be collected and serves a similar purpose. Even though a statistically significant difference between conditions was not found for question B3.4 (p = .105), the results trended in a similar direction to those of B2.3 with the participant-oriented condition showing a greater understanding of the riskiness of the

data. Question B2.4 related to data access and is an indirect gauge for participant understanding of risk as an interpretation can be made that the more people who have access to the data the greater the risk. Question B2.5 served a similar purpose but no statistically significant difference was found on B2.5 (p = .56). It is possible that the participant-oriented condition had a greater awareness of the risks of the study which dissuaded them from enrolling in the study, but further investigations are needed.

Consistent with past studies (Flory & Emanuel, 2004; Stunkel *et al.*, 2010; Beskow *et al.*, 2017) the type of consent form did not affect participant comprehension of informed consent. Also, consistent with the findings of Stunkel *et al.* (2010) we did not find a correlation between previous research participation and greater comprehension (See Appendix D). However, these past findings typically came out of comparisons between short and long consent forms and did not compare two types of enhanced consent forms. As noted above, differences were identified on specific questions on the objective measure related to understanding of risk.

Limitations of the study

Participant subjective and objective understanding of informed consent was assessed by a test created for the study. We modeled the test after similar measures created in biomedical research and followed suggested practices. Nonetheless, the comprehension test should be further refined and validated. The results of the factor analyses suggest that while questions should not be removed, they could be rephrased to better explain the differences among the questions that appear to be in the same component. Moreover, all questions have been weighted equally which may not aptly reflect the comparative importance of the questions. It may also be worthwhile to have criterion experts and potential participants assess the accuracy and readability of measure; and to assess its test-retest reliability (Joffe *et al.*, 2001; Guarino *et al*., 2006).

Kass, Taylor, Ali, Hallez and Chaisson (2015) noted that most informed consent intervention studies used simulated research conditions. We assessed individuals' understanding of informed consent in an actual MMLA research study. Despite having the study begin in 6 different groups and in 2 different languages, we managed to hold all groups in the same week and with the same teacher. However, we were not able to access all of the data relevant to the study as we had to omit the data from those who did not enrol in the study. We attempted to mitigate the data lost by following an ongoing process of consent which enabled potential participants to reflect upon and change their decision regarding consent in each data collection instrument. Still we ended up with less data from participants that did not give consent on the initial consent form (72.92%) compared to those that gave consent on the initial consent form (94.78%). Moreover, as the study was conducted in a first-year course of a computer science bachelor's degree programme, the generalizability of the findings is not clear and could be improved by collecting more qualitative data via focus groups and interviews to gain deeper insights into participants' understanding of informed consent and the rationale for their decisions.

Implications for multimodal learning analytics researchers

Recent MMLA research is showing a greater interest in ethical issues related to user/participant understanding of data acquisition and use. For example, Cowling and Birt (2020) discuss ethical concerns related to data storage, privacy, and security when applying MMLA to innovative learning scenarios such as those involving mixed reality. Schneider, Reilly and Radu (2020) point out concerns related to the increasing amount of fine-grained data that is infiltrating educational environments. However, recent MMLA studies collecting physiological data (e.g. Pijeira-Díaz, Drachsler, Järvelä, Kirschner, 2019; Echeverria, Martinez-Maldonado, Buckingham Shum, 2019) may promote the use of responsible practices, but do not make explicit the importance of informed consent and ensuring adequate understanding of the study and its risks by participants. The results of our study point to deficiencies in our informed consent process for MMLA research. The average score on the subjective portion of the test was 78.29% (M = 5.48, SD = 2.10) whereas the average score on the objective portion of the test was only 53.17% (M = 11.17, SD = 3.57). Further, the scores for a number of questions identify specific concerns related to a lack of participant understanding of the type and riskiness of the data being collected (B2.3, B2.4, B2.5, B3.4). As there is a lack of publications related to informed consent and ethics in the MMLA research, it is unclear how common researcher-oriented consent forms are and whether our results could be indicative of the field. If comprehension is required for valid consent, then those conducting MMLA experiments may need to determine what level of comprehension is deemed adequate, what instruments are appropriate for measuring it, and what steps need to be taken if adequate comprehension is not demonstrated. For example, Beskow et al. (2017) had participants review the sections of the consent form that corresponded to the items they answered incorrectly, and then had them complete a retest on the same topics.

Beyond the issues related to understanding, additional ethical concerns emerged out of participant test responses. A number of participants noted feeling pressure to participate in the study (A2.6) – marking they were unsure (18 out of 162) or felt pressure (7 out of 162) to participate in the study. Further, some participants were not sure if their participation in the study would appear on their student records (B1.2: 30 out of 162) nor if their teacher would be disappointed in them if they did not join the study (B1.3: 20 out of 162). Such results could invalidate consent in formal educational settings in which there is an unequal distribution of power among parties – and much of MMLA research takes place in such settings. A sharing of best practices among researchers could help address these concerns but an initial step involves collecting data to see if such a problem exists or not. Buccini, Iverson, Caputi, Jones & Gho (2009) suggest that instruments such as informed consent comprehension tests are useful in identifying gaps in knowledge and pointing to where additional education is necessary.

Conclusions

We introduced an informed consent comprehension test for educational technology research and assessed the effects of enhancing MMLA consent forms on comprehension of informed consent and on rates of enrollment in a MMLA study. The MMLA consent

form written from a participant perspective resulted in higher levels of comprehension on test questions related to risk, yet lower rates of enrollment. Tait, Voepel-Lewis, Robinson and Malviya (2002) write that "it is every investigator's goal to maximize recruitment rates in order to provide a representative sample of sufficient size to achieve statistical power" but the authors add that investigators must achieve this goal "through the design of ethically sensitive protocols involving complete and honest disclosure" (p. 335). Our study suggests that more work is needed to discover a balanced approach that can adequately inform participants about risks and benefits without significantly affecting rates of enrollment, especially as MMLA research data increases in complexity and invasiveness. Overall, our work highlights potential weaknesses in the informed consent process of MMLA research conducted in a formal educational setting (e.g. participant understanding of risk, feeling of pressure, feelings of not being adequately informed); provides evidence that the manner in which studies are communicated to participants via consent forms can affect enrollment rates; and introduces an approach for MMLA research, derived from bioethics, to evaluate participant understanding of consent.

Acknowledgements

This work was partially funded by the Spanish Ministry of Economy and Competitiveness under MDM-2015-0502, TIN2017-85179-C3-3-R. D. Hernández-Leo acknowledges financial support by ICREA under the ICREA Academia programme.

Statements on open data, ethics and conflict of interest

The study was approved by the ethics committee of Universitat Pompeu Fabra, Barcelona. Consent was obtained from participants. Anonymized data excerpts are available in Zenodo (http://doi.org/10.5281/zenodo.3557272). There are no potential conflicts of interest in the work.

References

Alderson, P., & Morrow, V. (2004). Ethics, social research and consulting with children and young people. *Bulletin of Medical Ethics*, *139*, 5-7.

Article 29 Working Party (2018) Guidelines on consent under Regulation 2016/679. 17/EN, WP259 rev.01

Barrio-Cantalejo, I. M., Simón-Lorda, P., Melguizo, M., Escalona, I., Marijuán, M. I., & Hernando, P. (2008). Validación de la Escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes. In *Anales del Sistema Sanitario de Navarra. 31*(2), 135-152. Gobierno de Navarra. Departamento de Salud.

Beardsley, M., Vujovic, M., Martinez-Moreno, J., Santos, P. & Hernández-Leo, D. (2019). Enhancing consent forms to support participant decision making in multimodal learning research [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3557272

Beardsley, M., Santos, P., Hernández-Leo, D., & Michos, K. (2019). Ethics in educational technology research: Informing participants on data sharing risks. *British Journal of Educational Technology*, *50*(3), 1019-1034.

Beskow, L. M., Lin, L., Dombeck, C. B., Gao, E., & Weinfurt, K. P. (2017). Improving biobank consent comprehension: a national randomized survey to assess the effect of a simplified form and review/retest intervention. *Genetics in Medicine*, *19*(5), 505.

Beskow, L. M. (2016). Lessons from HeLa cells: the ethics and policy of biospecimens. *Annual Review of Genomics and Human Genetics*, *17*, 395-417.

Bjørn, E., Rossel, P., & Holm, S. (1999). Can the written information to research subjects be improved?--an empirical study. *Journal of Medical Ethics*, *25*(3), 263-267.

Blikstein, P. (2011, February). Using learning analytics to assess students' behavior in open-ended programming tasks. *In Proceedings of the 1st international conference on learning analytics and knowledge* (pp. 110-116). ACM.

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*(2), 220-238.

Buccini, L. D., Iverson, D., Caputi, P., Jones, C., & Gho, S. (2009). Assessing clinical trial informed consent comprehension in non-cognitively-impaired adults: a systematic review of instruments. *Research Ethics*, *5*(1), 3-8.

Chassang, G. (2017). The impact of the EU general data protection regulation on scientific research. *ecancermedicalscience*, *11*.

Corrin, L., Kennedy, G., French, S., Shum, S.B., Kitto, K., Pardo, A., West, D., Mirriahi, N. & Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education. *A Discussion Paper*. Retrieved from https://melbourne-cshe.unimelb.edu.au/__data/assets/pdf_file/0004/3035047/LA_Ethics_Discussion_Paper.pdf

Cowling, M. A., & Birt, J. R. (2020). Mixed Reality Multimodal Learning Analytics. *In Encyclopedia of Educational Innovation*. Springer

Department of Health and Human Services et al. (2017). Final Rule: Federal Policy for the Protection of Human Subjects. *Federal Register*, *82*(12), 7149-7274.

Dranseika, V., Piasecki, J., & Waligora, M. (2017). Relevant information and informed consent in research: in defense of the subjective standard of disclosure. *Science and Engineering Ethics*, *23*(1), 215-225.

Echeverria, V., Martinez-Maldonado, R., & Buckingham Shum, S. (2019, May). Towards collaboration translucence: Giving meaning to multimodal group data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).

EU General Data Protection Regulation (2016), "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union*, 119 (1), http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN.

European Commission (2014). New modes of learning and teaching in higher education. Luxembourg: Publications Office of the European Union 2014, 68 pp. ISBN 978-92-79- 39789-9 DOI:10.2766/81897 http://portal3.ipb.pt/images/icm/1.2_2014_modernisation_en.pdf

Falagas, M. E., Korbila, I. P., Giannopoulou, K. P., Kondilis, B. K., & Peppas, G. (2009). Informed consent: how much and what do patients understand?. *The American Journal of Surgery*, *198*(3), 420-435.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221.

Flory, J., & Emanuel, E. (2004). Interventions to improve research participants' understanding in informed consent for research: a systematic review. *JAMA*, *292*(13), 1593-1601.

Fortun, P., West, J., Chalkley, L., Shonde, A., & Hawkey, C. (2008). Recall of informed consent information by healthy volunteers in clinical trials. *QJM: An International Journal of Medicine*, *101*(8), 625-629.

Guarino, P., Lamping, D. L., Elbourne, D., Carpenter, J., & Peduzzi, P. (2006). A brief measure of perceived understanding of informed consent in a clinical trial was validated. *Journal of Clinical Epidemiology*, *59*(6), 608-614.

Hadden, K. B., Prince, L. Y., Moore, T. D., James, L. P., Holland, J. R., & Trudeau, C. R. (2017). Improving readability of informed consents for research at an academic medical institution. *Journal of Clinical and Translational Science*, *1*(6), 361-365.

Hallinan, Z. P., Forrest, A., Uhlenbrauck, G., Young, S., & McKinney Jr, R. (2016). Barriers to change in the informed consent process: A systematic literature. *IRB*, *38*(3).

Heath, S., Charles, V., Crow, G., & Wiles, R. (2007). Informed consent, gatekeepers and go-betweens: negotiating consent in child and youth-orientated institutions. *British Educational Research Journal*, *33*(3), 403-417.

IBM Corp. Released 2015. IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.

Jefford, M., & Moore, R. (2008). Improvement of informed consent and the quality of consent documents. *The Lancet Oncology*, *9*(5), 485-493.

Joffe, S., Cook, E. F., Cleary, P. D., Clark, J. W., & Weeks, J. C. (2001). Quality of informed consent: a new measure of understanding among research subjects. *Journal of the National Cancer Institute*, *93*(2), 139-147.

Kadam, R. A. (2017). Informed consent process: A step further towards making it meaningful!. *Perspectives in Clinical Research*, *8*(3), 107.

Karbwang, J., Koonrungsesomboon, N., Torres, C.E., Jimenez, E.B., Kaur, G., Mathur, R., Sholikhah, E.N., Wanigatunge, C., Wong, C.S., Yimtae, K. & Malek, M. A. (2018). What information and the extent of information research participants need in informed consent forms: a multi-country survey. *BMC Medical Ethics*, *19*(1), 79.

Kass, N. E., Taylor, H. A., Ali, J., Hallez, K., & Chaisson, L. (2015). A pilot study of simple interventions to improve informed consent in clinical research: feasibility, approach, and results. *Clinical Trials*, *12*(1), 54-66.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel" (1975). *Institute for Simulation and Training*. 56.
https://stars.library.ucf.edu/istlibrary/56

Larra, M. F., Schulz, A., Schilling, T. M., de Sá, D. S. F., Best, D., Kozik, B., & Schächinger, H. (2014). Heart rate response to post-learning stress predicts memory consolidation. *Neurobiology of Learning and Memory*, *109*, 74-81.

Lorenzen, B., Melby, C. E., & Earles, B. (2008). Using principles of health literacy to enhance the informed consent process. *AORN Journal*, *88*(1), 23-29.

Manta, C. J., Ortiz, J., Moulton, B. W., & Sonnad, S. S. (2016). From the patient perspective, consent forms fall short of providing information to guide decision making. *Journal of Patient Safety*.

Mordini, E., & Ashton, H. (2012). The transparent body: Medical information, physical privacy and respect for body integrity. In *Second generation biometrics: the ethical, legal and social context* (pp. 257-283). Springer, Dordrecht.

Muravyeva, E., Janssen, J., Dirkx, K., & Specht, M. (2018, December). Students' Attitudes Towards Personal Data Sharing in the Context of e-Assessment: Informed Consent or Privacy Paradox?. In *International Conference on Technology Enhanced Assessment* (pp. 16-26). Springer, Cham.

Nishimura, A., Carey, J., Erwin, P. J., Tilburt, J. C., Murad, M. H., & McCormick, J. B. (2013). Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. *BMC Medical Ethics*, *14*(1), 28.

Ochoa, X., Lang, A. C., & Siemens, G. (2017). Multimodal learning analytics. *The Handbook of Learning Analytics, 1*, 129-141

Ochoa, X., & Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics*, *3*(2), 213-219. doi: 10.18608/jla.2016.32.10

Paasche-Orlow, M. K., Taylor, H. A., & Brancati, F. L. (2003). Readability standards for informed-consent forms as compared with actual readability. *New England Journal of Medicine*, *348*(8), 721-726.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438-450.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2019). Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members?. Computers in Human Behavior, 92, 188-197.

Rodríguez-Triana, M. J., Martínez-Monés, A., & Villagrá-Sobrino, S. (2016). Learning Analytics in Small-Scale Teacher-Led Innovations: Ethical and Data Privacy Issues. *Journal of Learning Analytics*, *3*(1), 43-65.

Schneider, B., Reilly, J., & Radu, I. (2020). Lowering Barriers for Accessing Sensor Data in Education: Lessons Learned from Teaching Multimodal Learning Analytics to Educators. *Journal for STEM Education Research*, 1-34.

Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., & Diamond, A. (2015). Enhancing cognitive and social–emotional development through a simple-to-administer mindfulness-based school program for

elementary school children: A randomized controlled trial. *Developmental Psychology*, *51*(1), 52.

Stunkel, L., Benson, M., McLellan, L., Sinaii, N., Bedarida, G., Emanuel, E., & Grady, C. (2010). Comprehension and informed consent: assessing the effect of a short consent form. *IRB*, *32*(4), 1.

Sugarman, J., Lavori, P. W., Boeger, M., Cain, C., Edson, R., Morrison, V., & Yeh, S. S. (2005). Evaluating the quality of informed consent. *Clinical Trials*, *2*(1), 34-41.

Sugarman, J. (2017). Examining provisions related to consent in the revised common rule. *The American Journal of Bioethics*, *17*(7), 22-26.

Sun, J. C. Y., & Yeh, K. P. C. (2017). The effects of attention monitoring with EEG biofeedback on university students' attention and self-efficacy: The case of anti-phishing instructional materials. *Computers & Education*, *106*, 73-82.

Swanlund, D., & Schuurman, N. (2018). Second Generation Biometrics and the Future of Geosurveillance: A Minority Report on FAST. A*CME: An International E-Journal for Critical Geographies*, 17(4).

Tait, A. R., Voepel-Lewis, T., Robinson, A., & Malviya, S. (2002). Priorities for disclosure of the elements of informed consent for research: a comparison between parents and investigators. *Pediatric Anesthesia*, *12*(4), 332-336.

Tam, N. T., Huy, N. T., Thoa, L. T. B., Long, N. P., Trang, N. T. H., Hirayama, K., & Karbwang, J. (2015). Participants' understanding of informed consent in clinical trials over three decades: systematic review and meta-analysis. *Bulletin of the World Health Organization*, *93*, 186-198H.

Tamariz, L., Palacio, A., Robert, M., & Marcus, E. N. (2013). Improving the informed consent process for research subjects with low literacy: a systematic review. *Journal of General Internal Medicine*, *28*(1), 121-126.

Villafranca, A., Kereliuk, S., Hamlin, C., Johnson, A., & Jacobsohn, E. (2017). The appropriateness of language found in research consent form templates: a computational linguistic analysis. *PLOS ONE*, *12*(2), e0169143.

Wendler, D., & Grady, C. (2008). What should research participants understand to understand they are participants in research?. *Bioethics*, *22*(4), 203-208.

Wittenberg, K. M., & Dickler, H. B. (2007). Universal use of short and readable informed consent documents: How do we get there? Creating Informed Consent

documents that inform: A Literature Review. *Association of American Medical Colleges Appendix C*.

Worsley, M., & Martinez-Maldonado, R. (2018, January). Multimodal Learning Analytics' Past, Present, and Potential Futures. In *CrossMMLA@ LAK*.

Young, D. R., Hooker, D. T., & Freeberg, F. E. (1990). Informed consent documents: increasing comprehension by reducing reading level. *IRB*, *12*(3), 1-5.