



Performance Study for Streamlined Prototype Free Application for Federal Student Aid (FAFSA)

Task Duration, Error Rate, and User Satisfaction

January 12, 2017

Submitted to :

National College Access
Network (NCAN)

Submitted By:

Ben Harper & Iva Stoyneva
for ICF

85 prospective postsecondary education students and parent/student dyads were recruited and screened to complete one of the five main user flows through the existing FAFSA on the Web and a streamlined design that relies on automation and data sharing between federal agencies to reduce the burden of the application. In the context of a usability study, participants completed the FAFSA form using one of the two Web-based systems while their time on task, user errors, and user satisfaction were measured.

Significant main effects for the experimental conditions were found in all measures, indicating that the streamlined FAFSA performed better than the Current FAFSA across users of all flow groups. Main effects for groups were also observed in task time and errors showing the advantage of being in a group that is required to answer fewer questions. There were no significant interaction effects observed.

These results demonstrate the advantage of many of the design decisions made, while ensuring that no group was made to perform worse in the streamlined design. Future approaches to improve the FAFSA user experience are proposed.

Contents

I. Introduction.....	2
II. Methodology	3
1. Participants	3
2. Materials	4
3. Procedure	4
3.1 Timing	5
3.2 Errors.....	6
3.3 Satisfaction.....	6
4. Power Analysis	7
III. Results.....	7
1. Task Timing	7
2. Errors	8
3. Satisfaction	9
IV. Discussion	10
V. Methodological Limitations.....	13
VI. References	14



I. Introduction

The Free Application for Federal Student Aid (FAFSA) is the federal application students need to complete if they wish to apply for federal financial aid for college, career school or graduate school. This is the largest source of financial aid for higher education in the US, providing more than \$150 billion in grants, work-study funds and loans. Many states and colleges use the data from the form to determine eligibility for state and school aid. Nearly all students who apply qualify for some type of aid. However, according to the US Department of Education, 20% of undergraduate students failed to fill out the financial aid application in 2011-2012, with lower income students being heavily represented in this group. As part of a larger effort to increase FAFSA completion rates, the National College Access Network (NCAN) is interested in improving and simplifying the user experience of the form.

There are a number of factors that impact the usability and perceived burden of filling out the FAFSA in any of its forms. Gathering the correct documents and records is anecdotally the most difficult aspect of the task, with participants struggling to find the tax records for the appropriate year, either because they have been lost or because they were prepared by someone else. Documented business income can also be difficult to locate and organize.

The length of the form can also be intimidating, with long pages of densely packed questions discouraging people from completing the form. The terminology used in the form can lead users to provide incorrect data or provide the right data but in the wrong format. Users may fail to answer questions that require extensive follow-ups with applicants, or cause outright denial of aid. There are also sections in the form where some users are asked to enter the same information into the form multiple times, or provide data that applicants have already provided to other federal agencies.

The Web interface to the application, FAFSA on the Web (FOTW) has undergone usability testing in the past, and improvements are made each year. However, FOTW is showing its age and is in need of an updated review. NCAN took a fresh look at the process, using the feedback from applicants to produce an alternative, Web-based design approach for the FAFSA that:

- Increases the use of information already reported to the IRS by applicants by expanding the range of information that can be imported into the application from the IRS Data Retrieval Tool (DRT). This helps minimize the need for tax documents and increases the quality of the data provided by eliminating data entry errors.
- Ensure that information provided in the DRT is not also required to be manually entered, reducing the length of the form and the burden on applicants.
- Organizes questions so that they can be quickly read and navigated.
- Reduces the length and complexity of instructions to improve comprehension.
- Updates the User Experience with a modern look and more dynamic data entry options and validation.

ICF's usability researchers conducted in-person moderated usability testing in order to investigate how users interacted with a Streamlined FAFSA on the Web (Streamlined FAFSA) form, using the current year 2016-2017 FAFSA on the Web (Current FAFSA) as a control. In order to cover the full range of FAFSA applicants and the six paths through the application that

each take, participants were selected that meet the requirements for each path and assigned to that group. The groups were:

Receives MTB – Dependent students whose families meet the requirements for Means Tested Benefits such as Federal Free or Reduced Price Lunch programs have already proved their need to a federal agency.

Single – Dependent students living with one parent most of the time and whose parent files taxes using a Single Status.

Married filing Jointly – Dependent students living with both parents whose parents file taxes jointly

Married with Assets – Dependent students living with both parents whose parents file taxes jointly and are required to report additional information about their assets, often related to a business or farm.

Independent Students – Students who are over 24 years old, or are not supported by parents for a number of reasons.

A 6th group, Married Filing Separately was found to account for less than 1% of participants and was excluded for reasons of simplicity and cost.

The hypotheses for this study include:

Task Duration – The Streamlined FAFSA, by shifting burden to automation to answer many questions will reduce the time to complete the form for each of the five main user groups of the FAFSA.

Errors –In the face of faster performance users of the Streamlined FAFSA will see no increase in errors on the form.

Satisfaction – Users of the Streamlined FAFSA will be more satisfied than users of the Current 2016-2017 FAFSA.

Equivalence – None of the FAFSA user groups will perform more poorly in the Streamlined version than the Current FAFSA.

II. Methodology

1. Participants

20 independent prospective students and 65 dependent high school junior and senior student/parent dyads were recruited from three locations nationally and screened to ensure racial and gender diversity. All were interested in attending a postsecondary education institution in the next two years. They were also screened to ensure that they matched the requirements and profile of their group. Participants were assigned to experiment or control conditions randomly.

Table 1. Participant demographics

Gender	#	Race / Ethnicity	#
Female Parents	52	African American Students (Dependent)	22
Male Parents	13	Hispanic Students (Dependent)	18
Female Students (Dependent)	34	Other Students (Dependent)	25
Male Students (Dependent)	31	African American Students (Independent)	9
Female Students (Independent)	11	Hispanic Students (Independent)	8
Male Students (Independent)	9	Other Students (Independent)	3

2. Materials

The FAFSA Demo site (<https://fafsademotest.ed.gov/index.htm>), which is provided to the public by the Department of Education, was used for the control condition; test data was provided with the system. The Demo site also supplied the information provided in the FSA ID system, simulating the use of the FSA ID (which populates some student information provided during the registration process). No applications were submitted through this system and all information was cleared at the end of each session. The Demo site doesn't support the IRS DRT for either parents or students, so participants had to provide that information from materials they were given.

The NCAN Streamlined FAFSA (experimental condition) was also presented to participants as a highly functioning mock-up via a web browser. It simulated the FSA ID, as well as the IRS DRT for parent and student tax information. The test data in this system was equivalent to the Demo site data, but not identical because the control system could not be modified.

In order to avoid collecting personally identifiable information about participants, they were provided with materials for hypothetical families and students that was similar to their own family situation. This information included tax forms for parents and students, as well as Social Security cards and a brief profile containing many other details that participants could use to fill out the application. Participants were asked to use these materials for some personal information, but to use their own information in other areas where they were comfortable providing it.

3. Procedure

The researchers conducted data collection in Bethesda MD, Houston TX and Los Angeles CA using a software package to measure time on task and record the facilitator and secondary reviewer's coding of errors, observations, and session notes.

Dyad participants were asked which person would be filling out the form if they were completing the FAFSA at home. This person was positioned in front of the test computer and was the primary user for the session. Some dyads shared data entry responsibility between each other.

In most cases the non-data entering participant provided error checking on the fly, and frequently “wrangled” the supporting documents that they used to complete the form.

Participants were asked to complete the form in sections. Between sections when task timing stopped, they could ask questions and were asked clarifying questions by the facilitator. Any section that the participant was not able to complete successfully without help from the facilitator was considered a failure. After the form was completed, participants completed a six-question user satisfaction scale.

3.1 Timing

Segment timing was completed during the sessions by the facilitator using the Morae software on a second computer, and verified during session recording reviews. Participants were instructed before the session that they could only ask questions during breaks between timing sections. If the facilitator had to intervene during a timed segment to address technical issues or provide additional task guidance, that section of the recording was excluded from task timing. This occurred in less than 10% of sessions. Additionally, any tasks that could not be completed successfully by the participants without facilitator intervention were marked as failed tasks and excluded from timing, but not error calculations. There were 12 failed tasks in 135 observations.

Task timing and errors were recorded for subsets of the FAFSA form so that comparable tasks could be compared across groups and experimental conditions. These tasks were:

1. **Student Demographics** – Questions about students, their educational background, qualification for aid, and selection of colleges to share FAFSA results with.
2. **Parent Demographics and Financial Information** – Questions for parents about their household, income, assets, and expenses.
3. **Student Financial Information** – Questions about the student’s income and assets.
4. **State-Specific Student Questions** – These are additional questions about students required by states rather than the Department of Education.

In the Current FAFSA (control) form the State-specific student questions are combined with student demographic information, but the Streamlined FAFSA (experimental) form includes a separate section for this.

Some groups are not required to complete all tasks in the form.

- Independent Students are not required to provide Parent information in the form and were instructed not to.
- Dependent students that meet the requirements for Means Tested Benefits are not required to complete the Student Financial task, and are required to complete a smaller number of questions on Parent Demographics and Financial Information.
- Participants in groups that have assets above the defined cut-off point are not asked to complete an additional task, but are asked some additional questions in the Parent section.

Table 2. The sections of the application that each participant need to answer varied by group and by condition. A group of questions request that are required by the state are mixed into the Student Demographics section in the Current 2016-17 FAFSA, but are presented separately in a new state-specific section in the Streamlined FAFSA.

	Control				Experimental			
	Student Demographics	Parent Demo and Financial	Student Financial	State-Specific	Student Demographics	Parent Demo and Financial	Student Financial	State-Specific
Receives MTB	X	X			X	X*		X
Single	X	X	X		X	X	X	X
Married filing Jointly	X	X	X		X	X	X	X
Married with Assets	X	X	X		X	X	X	X
Independent Students	X		X		X		X	X

* Families who indicate that they have taken part in qualifying federal benefit programs that require proof of need are allowed to skip the student financial sections of the form, but in the experimental condition they can also skip the parent financial questions.

3.2 Errors

An error was recorded every time the user purposely submitted erroneous data to the system regardless of whether the system produced a validation error. Failure to use system error messages to correct an error was also considered an error. Ambiguous errors and errors that could have been the result of malfunction of the system or inconsistent/incomplete instructions were not counted as errors. Cases in which the participants were exploring the functionality of the form by clicking around to learn more about it were not counted as errors, unless they led to ultimately making the wrong choice and going to the next question. Errors were recorded by one researcher and reviewed for consistency and accuracy by another.

3.3 Satisfaction

The six-question Overall Satisfaction scale from the Questionnaire for User Interaction Satisfaction (QUIS) was used to measure FAFSA satisfaction. Individual scales were combined without weighting to produce an overall score in a 9-point scale. This section of the QUIS has been validated and shown to be a reliable predictor of usability (Harper, Slaughter & Norman, 1997). The questionnaire was presented on the computer screen when the FAFSA was complete. Participants filled out the form before final debriefing which might bias their responses.

4. Power Analysis

To ensure that we collected enough performance data to detect the hypothesized advantages of the Streamlined FAFSA using inferential statistics we performed an A priori sample size estimate for a 1-tailed comparison. Using mean completion times from the Dept. of Education and data-driven estimates of population variability we calculated the number of participants required per test group to detect a 25% advantage in user performance at least 80% of the time (type II error), with a no more than 5% chance of reporting an advantage when there is none (type I error).

Mean (Dependent, Original, and Full Form across all filing quarters) = 32.5 Minutes

Standard Deviation (Treatment); estimated from Nielsen (2006) = 12.7 Minutes

Effect Size (Cohen's d) = 0.54

Desired Power = 0.8

Alpha = 0.05

Minimum sample size per condition = 43 participant

Because standard deviation information is not available from the Dept. of Education for their estimation of completion time, we must adopt an estimated standard deviation drawn from a survey of performance measurements obtained from over 100 usability evaluations. This estimate proved to be somewhat conservative and when combined with a larger than anticipated effect size, it is indicated that a sufficient sample was reached.

III. Results

1. Task Timing

A 2X5 ANOVA supports the speed of performance hypothesis that the Streamlined FAFSA was completed more quickly (M=623, SD=389) than the Current FAFSA (M=1027, SD=389). There was a 39 percent reduction in time to fill out the Streamlined FAFSA compared to the Current FAFSA. The η^2 for this model is 0.52. The Grouping factor was also statistically significant, but not surprising with user groups that are required to report more data on the form taking longer to do so. The interaction between condition and group failed to reach significance, thus supporting the equivalence hypothesis.

Figure 1. Task time for Control and Experimental Conditions across user groups.

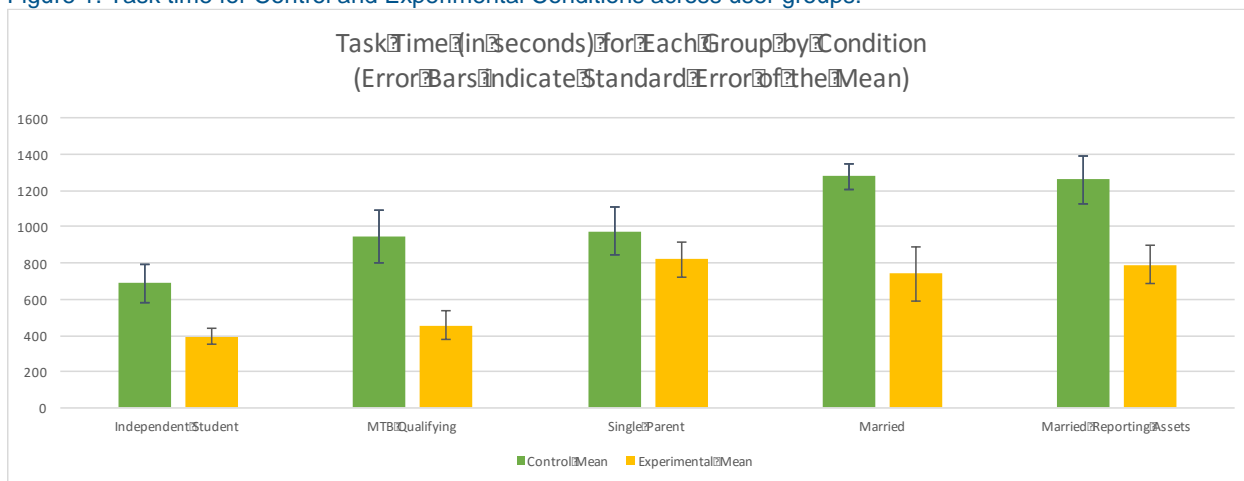


Table 3. Means, Standard Error of the Mean, and Standard Deviation by condition and group.

	Control			Experimental		
	Mean	SEM	SD	Mean	SEM	SD
Independent Student	687	102.14	323	395	41.74	132
Receives MTB	947	144.52	354	456	79.55	225
Single Parent	977	132.00	396	820	94.40	267
Married	1278	69.65	231	741	149.01	365
Married Reporting Assets	1260	130.40	345	792	104.65	296

Table 4. 2X5 ANOVA for task time showing significant main effects for condition and group, with no interaction between the two.

Source	DF	SS	Mean Square	F Value	Pr > F
Condition	1	3408233.460	3408233.460	38.89	<.0001
Group	4	3242365.312	810591.328	9.25	<.0001
Condition*Group	4	420396.222	105099.055	1.20	0.3184

2. Errors

The ANOVA for user errors supports the errors hypothesis that the user errors didn't increase for the experimental condition where task times decreased. In fact, the experimental condition (M=1.23 errors, SD=1.07) produced 56 percent fewer errors than the control condition (M=2.82 errors, SD=1.84). The group factor also reached statistical significance, with groups that have to report less information (Independent Students) generally producing fewer errors. The Interaction effect failed to reach statistical significance, supporting the equivalence hypothesis. The η^2 for this model is 0.37.

Figure 2. User Errors for Control and Experimental Conditions across user groups.

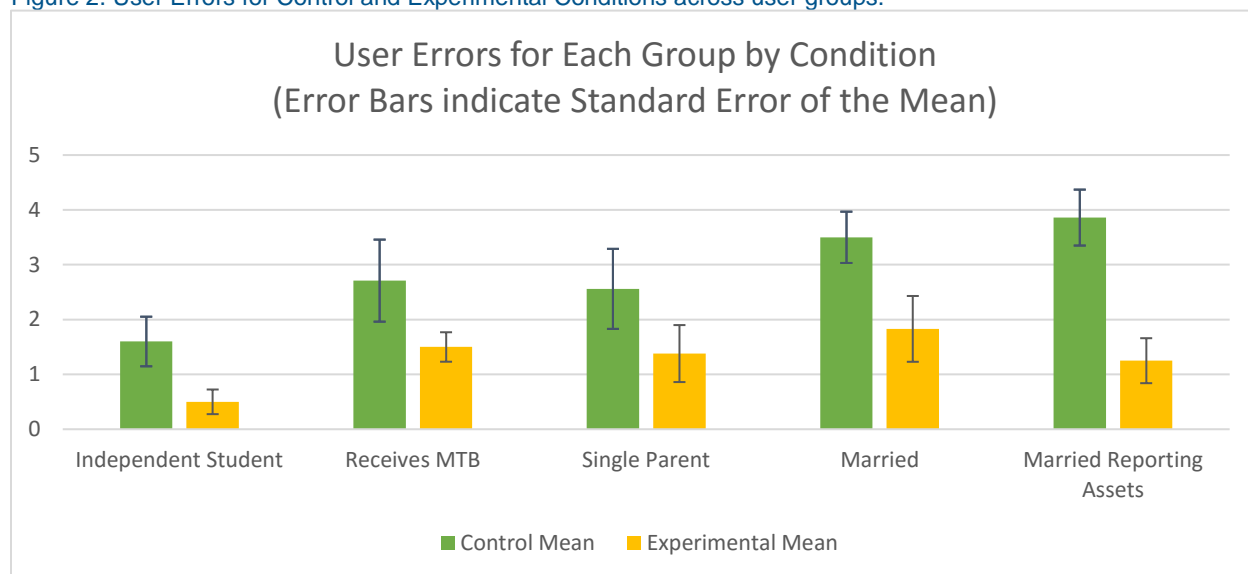


Table 5. Means, Standard Error of the Mean, and Standard Deviation by condition and group.

	Control			Experimental		
	Mean	SEM	SD	Mean	SEM	SD
Independent Student	1.60	0.45	1.43	0.50	0.22	0.71
Receives MTB	2.71	0.75	1.98	1.50	0.27	0.76
Single Parent	2.56	0.73	2.19	1.38	0.52	1.47
Married	3.50	0.47	1.62	1.83	0.60	1.47
Married Reporting Assets	3.86	0.51	1.35	1.25	0.41	1.16

Table 6. 2X5 ANOVA for user errors showing significant main effects for condition and group, with no interaction between the two.

Source	DF	SS	Mean Square	F Value	Pr > F
Condition	1	54.02412500	54.02412500	25.87	<.0001
Group	4	30.71912986	7.67978246	3.68	0.0087
Condition*Group	4	6.21723220	1.55430805	0.74	0.5648

3. Satisfaction

The main effect of condition reached significance, supporting the satisfaction hypothesis. Participants were more satisfied with the Streamlined FAFSA (M=7.45, SD=1.18) than the Current 2016 - 2017 FAFSA (M=6.63, SD=1.39), regardless of their group. Neither the group factor nor the interaction reached significance supporting the equivalence hypothesis. The η^2 for this model is 0.18.

Figure 3. User Satisfaction for Control and Experimental Conditions across user groups

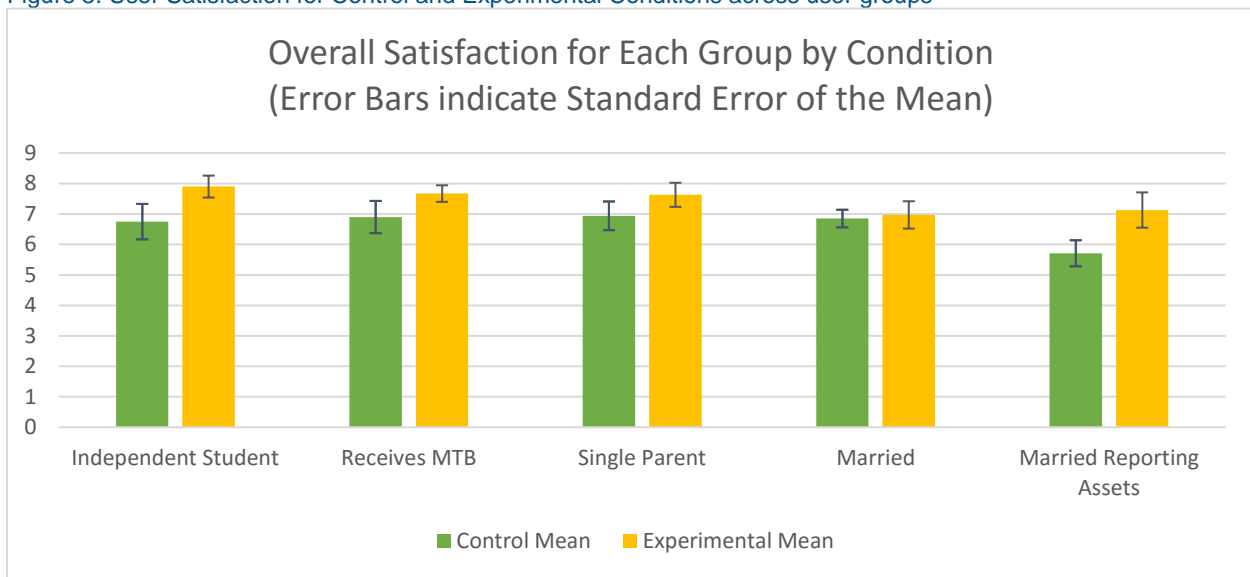


Table 7. Means, Standard Error of the Mean, and Standard Deviation by condition and group.

	Control			Experimental		
	Mean	SEM	SD	Mean	SEM	SD
Independent Student	6.75	0.58	1.84	7.90	0.36	1.14
Receives MTB	6.90	0.53	1.40	7.67	0.27	0.77
Single Parent	6.94	0.47	1.41	7.63	0.40	1.12
Married	6.85	0.29	1.00	6.97	0.45	1.10
Married Reporting Assets	5.71	0.43	1.13	7.13	0.58	1.64

Table 8. 2X5 ANOVA for user satisfaction showing significant main effects for condition, but not group, with no interaction between the two.

Source	DF	SS	Mean Square	F Value	Pr > F
Model	9	27.1001984	3.0111332	1.79	0.0847
Condition	1	14.46180556	14.46180556	8.58	0.0045
Group	4	8.76758220	2.19189555	1.30	0.2774
Condition*Group	4	3.87081066	0.96770266	0.57	0.6820
Error	75	126.3553571	1.6847381		
Corrected Total	84	153.4555556			

IV. Discussion

In all groups and across all measures the Streamlined FAFSA developed by NCAN produced measurably better outcomes than the current 2016-2017 FAFSA on the Web. Much of this

advantage should be attributed to the enhanced use of IRS DRT in the design that reduces the time and effort for applicants by asking them fewer questions and leveraging information that has already been reported to the government. There was a concern that the additional questions required to initiate the DRT and reviewing the imported information would negate some or all of the advantage of the enhanced DRT, but these results put that concern to rest.

It's more difficult to attribute the reduction in errors and improvement in user satisfaction primarily to the enhanced use of the DRT, and other sources should be considered. The usability findings from this study, documented in a companion usability report, can give some indication of the sources of error and satisfaction by users of the Streamlined FAFSA. Some consistent and serious problems with indicating whether a question is asking about students or parents were not solved in either version. Other errors related to Institution Search were partially addressed in the streamlined version, but too often replaced by new errors attributed to the new design. Some other problems were present in the Current FAFSA, but were improved in some way in the Streamlined FAFSA.

- Highlighting the wrong field when the system detects errors on the form leading to confusion about where the error occurred.
- Form entry fields for dates and phone numbers that introduced rather than reduced errors in data entry.
- The DRT eliminated several of the financial questions that led to confusion and poor data quality meaning that users only needed to verify the populated answers for in the streamlined version.
- The Save Key feature was not used in the streamlined version eliminating a consistent source of errors and delay.
- The number of jargon-laden questions about tax forms were reduced significantly due to the use of the DRT.

These significant sources of user error were not solely driven by the reduction in the number of questions, and impacted error rates similarly to the effect of the IRS DRT. It is also likely that the task time difference in combination with design changes that reduced error, and a more modern-looking user experience drove the difference in user satisfaction.

There were early concerns that changes aimed at reducing the burden for more disadvantaged students might increase the burden on other applicant groups by impacting their path through the system. While there are some significant differences between independent students, families who receive MTB, and other groups, every group benefited from the streamlined form.

There remain some significant variances between performance in this study and in the real world that need to be addressed. First, in this study, we are unable to include one of the most challenging tasks facing applicants and their parents: the collection of required information and documents. For a number of practical reasons, we were unable to ask participants to bring in their own tax documents, IDs, and business accounting records. All participants in both conditions were provided with all the materials they need and key personal information to use in the test. We however have no reason to expect that this is a relevant differentiator between the conditions tested here. In both cases these records are required to either enter information from them into the form, or in the streamlined version, to verify the information imported from the IRS DRT.

There is also likely a difference in the perceived importance of filling out the application for this test and in real life for the participants' own future education. When filling out the form at home or with the help of a college counselor, parents and students are motivated to be very careful and accurate because the cost of errors can be so high. That level of risk and therefore care were not replicated in the study, and participants performed to a level that was "good enough" to pass validation and satisfy the researcher. For this reason, the task timing and error rates that are reported here are not an accurate reflection of the performance observed when real applications are submitted. Relative differences between conditions and groups, on the other hand will likely remain.

The FAFSA Usability Report makes a number of recommendations to address usability concerns in both the Current FAFSA and the streamlined version that can improve both forms. In addition to these opportunities for improvement, this study suggests other improvements.

Making the IRS DRT available to the widest number of parents and working students will have the most significant effect on reducing the burden of the FAFSA on applicants. It has the potential to allow applicants who have misplaced records to complete the applications without them.

A related consideration is the prospective perceived burden that parents and some students bring to the FAFSA process. When asked after completing the form how the process compared to their expectations, a few participants said that it met their expectations, but the vast majority reported that the FAFSA, regardless of experimental condition was much easier than they expected, and expressed a lot of amazement that they were done so quickly. These parents often thought back to either their own experience with a paper FAFSA form, or the stories told about that era by families and friends. An effort to convince parents and independent students that the FAFSA burden has been drastically reduced, in combination with the further utilization of the IRS DRT, may bring reluctant parents to the form earlier in the application window because of the additional options open to them.

The simplification of questions and instructions did not appear to lead to inaccurate data being entered more frequently into the form. This would have been detected as user errors, and we would see more of these errors in the experimental condition. A future focus for design and study is the reduction of parent / student confusion errors where applicants enter student information into parent questions and vice versa. The Current FAFSA has tried to make the difference between parent-focused and student-focused questions clear with color changes and labeling with limited success. The streamlined version uses labeling, but the questions are worded to speak consistently to students even if the subject of the question is the parent.

Both of these approaches led to confusion and errors. We found that there are many opportunities to insert a subject, either parent or student, into the wording of questions while consistently addressing for instance the students; for instance, saying "Your parent's marital status" instead of "Marital status" without the context. We recommend implementing this approach consistently across the FAFSA questions and testing specifically for these types of errors as well as tax terminology confusion in a separate small study. Other designs can also be considered in that evaluation.

Other specific recommendations are covered in the FAFSA Usability Report that will have additional impact on burden, data accuracy, user errors and user satisfaction.

V. Methodological Limitations

The methodology used in this study may pose limits in the ability to generalize based on the results.

- Due to the experimental nature of the Streamlined FAFSA some of its functionalities, such as complete online help and extensive documentation, were not available. Some functions like Institution Search were functional but not optimized for performance or usability.
- Although the 2016-2017 FAFSA on the Web remained the same, some of the test data changed during testing due to the early rollout of the 2017-2018 Demo site leading to some materials updating at approximately the midpoint of the study.
- Because the test data was static, students and parents had to assume the identity offered by the system, which may force them to play the role of someone from a different gender when answering some questions, particularly about Selective Service. Participants were asked to use these roles when the system provided their information, and for names, dates of birth, and Social Security number. In all other cases they should use their own information. Participants occasionally used these identities for answering questions instead of providing their own.
- Because of participant availability and the large number of sessions performed, we were unable to control for potential time of observation bias.

VI. References

Harper, B., Slaughter, L., & Norman, K. (1997). Questionnaire administration via the WWW: A validation and reliability study for a user satisfaction questionnaire. Paper presented at WebNet 97, Association for the Advancement of Computing in Education, Toronto, Canada.