Measuring Skills @ Scale: Work Readiness Assessment
Field Test Analysis Report
September 2019

INTRODUCTION

Improving the quality of education continues to be a priority around the world, and assessments are considered a vital component in this effort. Typically, Ministries of Education and school systems worldwide use summative assessments, such as national exams, as a means to rank schools and their students. While summative assessments serve an important role in the education system, there has been a gap in the use of formative assessments to help teachers and schools improve the quality and efficiency of their education practices. Through the PSIPSE consortium, Education Development Center (EDC) is developing a soft skills assessment designed, tested, and ultimately implemented with the end-user in mind to bridge that gap.

EDC has developed a soft skills formative assessment that can give access to and empower education officials, teachers, and trainers, for both in-school and out-of-school youth, to use formative assessment data to improve learner outcomes around soft skills. The assessment is human-centered, meaning educators and Ministry Officials will be able to administer and analyze the assessment, and use the results to foster discussion with youth and parents, as well as to improve teaching approaches.

This report details the collaborative effort of Education Development Center and local partners VSO Tanzania and Youth Alive Uganda to develop, pilot, and analyze the validity of a formative soft skills work readiness assessment in Senegal, Rwanda, Tanzania, and Uganda.

The Measuring Skills @ Scale Work Readiness Assessment

Assessment, generally, plays a large role in education systems around the world. At its core, the assessment encompasses a variety of ways that teachers, trainers, or other educators can obtain data about their students' learning and about their own teaching.[1] While there are many forms of assessments, they can ultimately be placed within three key categories: (1) diagnostic assessments, (2) summative assessments, or (3) formative assessments.

Diagnostic assessments serve as a baseline study – an assessment of students' level of education before teaching takes place. Summative assessments are typically administered after learning has been completed, providing information on the success or lack thereof of students to grasp the material taught. Usually, summative assessments are relatively high-stakes, such as national exams or entrance exams. However, between the baseline diagnostic assessment and the endline summative assessment lies room for the formative assessment.[2] Formative assessments provide information during the teaching and learning process, to monitor student learning and provide feedback to educators on the efficiency of their teaching

---

[1] https://www.niu.edu/facdev/_pdf/guide/assessment/formative%20and_summative_assessment.pdf
[2] Ibid.

practices.[3] Formative assessments serve two main purposes – to identify students' strengths and weaknesses, and to inform educators on how to help their students' learning improve. The Measuring Skills @ Scale (MS@S) project will work to increase the use of formative assessments in schools and training programs focused on soft skills development in youth.

The MS@S assessment serves as a formative assessment, evaluating student knowledge of four soft skills standards: (1) Communication, (2) Interpersonal Skills, (3) Dependability, and (4) Problem Solving/Critical Thinking. These standards were selected based on a Developing a Curriculum (DACUM) workshop, as well as a review of YouthPower research from USAID. The DACUM workshop was conducted with EDC's youth workforce development subject-matter experts (SMEs), who homed in on the knowledge, skills, and abilities (KSAs) important for youth success in the workforce. Following the workshop, the facilitator condensed the extensive list of KSAs to eighteen and administered a verification survey to assess these skills on their importance, ease of acquisition, frequency of use, and requirement for starting an entry-level job. Eventually, the four standards were determined based on the following criteria:

> **Skills Standards Measured:**
>
> 1. Communication
>
> 2. Interpersonal Skills
>
> 3. Dependability
>
> 4. Problem Solving / Critical Thinking

- They were deemed most important and useful for success in entry-level work, by work readiness experts.
- They were not too difficult for a youth to develop during his or her time as a student.
- They could be obtained before beginning an entry-level position.
- They were in line with existing research on which skills are linked to success in the workplace.

The formative soft skills assessment will be useful for a variety of stakeholders:

- **Teachers/Trainers** can be equipped to reliably measure both the success of their instruction and their students' work readiness skills.
- **Youth** will receive reliable feedback from their teachers on their work readiness skills and areas for improvement.
- **Larger systems** – such as ministries of education and NGOs – will have reliable data to help them evaluate work readiness curricula and instructional approaches.

Project Purpose

The short-term goal of the MS@S initiative is to design, develop, and test a soft skills assessment that can be used by teachers in their classrooms and trainers in their training programs. The long-term goal of the initiative is to promote an environment in which teachers and students view assessment not only as a summative tool used to judge success or failure, but as a formative tool that can be used to communicate progress and allow teachers to target instruction to improve progress. The Work Readiness Assessment borne out of the MS@S initiative can be a fundamental part of this environment.

---

[3] https://www.cmu.edu/teaching/assessment/basics/formative-summative.html

METHODOLOGY

Item Development

Initially, EDC worked with NOCTI – a US-based test developer with more than five decades of experience in developing tools in the field of Career and Technical Education (CTE) in the United States. For later revisions of the survey, EDC revised and removed questions through a collaborative process with teams from the four countries of interest and with SMEs within EDC.

The assessment is structured as three nearly equivalent forms – forms A, B, and C – measuring the four soft skills standards. Separate subscales of items (or questions) for each of the four soft skills standards were included on each form. The decision to design three forms instead of one allowed for a greater number of items measuring soft skills, more teachable moments aligned with the scenario-structured items, and a variety of forms to use in testing instances to prevent cheating. Some questions repeat across forms, and others are unique to each form. The questions are formatted as scenarios that an employee is likely to encounter in the workplace, each focusing on competencies within one of the four standards.

> **1. What action shows you are listening to a person training you in a task at work?**
>
> A. Looking for some information on your phone.
>
> B. Responding to the trainer's questions.
>
> C. Talking about what you are learning to other workers.
>
> D. Reading from your notebook.

Figure 1: Sample Question

The version of the assessment used in the first pilot in 2018 consisted of near 100 items, but results showed low reliability of this initial version. By the time of the field test in 2019, the assessment had been whittled down to near 70 items by removing questions with relatively low reliability. The final forms had between 15 and 20 items per subscale, or soft skills standard. Table 1 shows an overview of the assessment structure.

Assessment Blueprint		
Standard	**Percent of Test**	**Example Competencies (not exhaustive)**
Communication	28%	• Effectively communicating orally • Provide clear and concise instructions • Participate in group or team discussions
Interpersonal Skills	27%	• Work cooperatively in a workplace team • Provide good customer service • Identify steps of resolving conflict
Dependability	33%	• Take personal responsibility for work • Follow directions • Exhibit flexibility & adaptability
Problem Solving / Critical Thinking	12%	• Identify solutions that consider both benefits and risks • Make decisions considering all facts

Table 1: Assessment Blueprint

The final forms were designed with the intention of providing reliable scores suitable for classroom instruction. Based on the desire for easy instructional use, as well as practical expectations made from the analysis of a second round of pilot (re-pilot) data from March 2019, the reliability target for the subscales on each form was determined to be a Cronbach's alpha of at least .70. Analysis of the re-pilot data already confirmed that Communication, Interpersonal Skills, Dependability, and Problem Solving/Critical Thinking subscales were highly correlated, allowing for a total score based on approximately 70 items per form representing an overall assessment of soft skills. Given the larger number of items used to compute the total soft skills score, Cronbach's alpha targets were set to at least .90 allowing for a total score with sufficient reliability to track individual progress over time.

Implementation

The MS@S work readiness assessment was piloted, re-piloted, and implemented with youth in Senegal, Tanzania, Rwanda, and Uganda. In each country, the tool was translated into the local language and then translated back into English to ensure that the content was preserved. This process allowed for any necessary revisions to the tool before implementation. The final versions of the tool contained text in both the local language (KiSwahili, Kinyarwanda, and Lugandan) and English[4] and were administered in a proctored setting in each country.

The initial pilot took place in the late spring to early summer of 2018 with 50 youth in each of the four countries. Field teams identified schools or training centers for testing sites, and youth within these schools were selected based on certain criteria. The selection criteria required a near equal number of boys and girls, as well as a secondary-school literacy level (for both in-school and out-of-school youth) to meet the language level of the assessment. The reliability testing for the pilot assessment produced low reliability

---

[4] In Senegal, the test was only translated to and offered to participants in French.

results, so the team at EDC conducted a process of revising or removing low-performing questions. Due to these changes, the team conducted another pilot using the revised versions of the tools.

The re-pilot was administered in Senegal and Tanzania in March 2019. The same criteria for test-takers were considered, the tools were translated and back translated, and the tests were again administered in a proctored setting. The reliability testing of the re-pilot was higher than the initial pilot, though some questions were again revised or removed with the support of field team staff members. The revised assessments were then translated and back translated in preparation for implementation in Uganda and Rwanda for the field test.

The field test took place in late April to early June 2019 in all four program countries, with identical selection criteria for test-takers. The samples for each country are as follows:

- **Senegal:** 513 secondary level in-school youth took the assessment on tablets using survey software. The sample included 308 females and 205 males.
- **Tanzania:** 546 out-of-school youth took the assessment, most using tablets and around forty using pen and paper due to technical difficulties. The sample included 281 females and 265 males.
- **Rwanda:** 605 out-of-school youth took the assessment on tablets using survey software. The sample included 331 females and 274 males.
- **Uganda:** 557 secondary level in-school youth took the assessment on tablets using survey software. The sample included 328 females and 229 males.

Analysis Methodology[5]

To ensure the construction of a reliable and valid assessment, both classical item analysis and item response theory (the Rasch model) were used, along with an analysis of Differential Item Functioning (DIF) to analyze test and item characteristics.

To examine the performance of each of the three test forms, three analytic samples were created:

1. A combined sample of students from Rwanda, Senegal, Tanzania and Uganda
2. A sample split by country
3. A sample using equated form scores based on item response theory

Six research questions guided the analysis of field test results.

Research Question #1: Does the reliability of each test form at both the overall test level and at the level of each soft skills standard meet the Cronbach's alpha reliability targets of 0.9 and 0.7, respectively?

The reliabilities of both the overall forms and of each soft skills standard's subscale were determined by computing the Cronbach's alpha for each of these scores on each form. Criteria for success of field test construction at the sub-standard level, as already outlined, consisted of reliabilities of at least .70 for the Communication, Interpersonal Skills, Dependability and Problem Solving/Critical Thinking subscales across all three forms. Criteria for success of the entire work readiness assessment consisted of reliabilities of

---

[5] In this section onward, explanations of words or phrases in light blue appear in a Glossary of Analysis-Related Terms in Appendix G. The words or phrases are only highlighted the first time they appear in the report.

greater than .90 for the overall scores of each form. The reliabilities studied in this research question used the combined sample of students from Rwanda, Senegal, Tanzania, and Uganda.

Research Question #2: Are the levels of difficulty for each soft skills standard's subscale (Communication, Interpersonal Skills, Dependability, and Problem Solving/Critical Thinking) comparable across forms?

Classical item analysis was performed to examine item difficulty (percent of students answering the item correctly) and item discrimination (corrected item-total correlations) for each soft skills standard's subscale independently. Link items – items which appear across multiple forms – for each subscale aided in determining the comparability of groups (each form's test-takers) when examining the classical item results. Further, using these same link items, the Rasch model was applied to equate performance on subscale items across forms. This equating allowed for raw scores for individuals taking forms B and C to be re-expressed as their raw score equivalent on form A. These scores were used to compute adjusted group means for each form, which aided in the comparison of each subscale across forms.

A difference between group means based on raw scores and group means based on adjusted scores allows for a better comparison of a given standard score across forms. In cases where subscales on one form appear particularly easier or difficult than on the other forms, the use of equated form scores was preferable when comparing difficulty across forms. A raw score equating table is presented in Appendix E. This equating table was produced both to assist the current analysis and to provide a convenient method that teachers and trainers can use in the field to compare performance across forms. The levels of difficulty studied in this research question also used the combined sample of students from Rwanda, Senegal, Tanzania, and Uganda.

Research Question #3: Should items be removed or revised based on the results of the item analysis?

Classical item statistics, particularly item discrimination, was used to determine if item scores were consistent with the subscale score to which items belonged. Positive item discriminations along with a positive contribution to subscale reliability were indicative of items performing as intended. Items with low or negative item discrimination values were candidates for further review by research staff to determine if items should be removed, modified, or remain in the final assessment instruments. The analysis conducted for this research question used the combined sample of students from Rwanda, Senegal, Tanzania, and Uganda.

Research Question #4: Is gender bias present, and does the presence of biased items threaten the validity of inferences made from soft skills standard-level and total work readiness scores?

To determine if test items unequally favor boys or girls, all items were examined for Differential Item Functioning (DIF) using logistic regression. DIF analyses were conducted where item scores (correct or incorrect) were modeled using equated total scores, a group membership variable, and an interaction term representing a group by total interaction. This provides a means of testing for uniform DIF, where the item bias is consistent across groups regardless of ability level, and non-uniform DIF where item characteristics for each group differ depending on overall level of achievement. Both the number of items per standard exhibiting DIF favoring either boys or girls and a measure of effect size were considered when assessing the impact of DIF on assessment results. This study of gender bias used the sample of equated form scores based on item response theory.

Research Question #5: (Exploratory) Are there achievement differences with respect to gender within the countries and samples tested?

Since DIF is an item characteristic and not a measure of group ability, exploratory analyses conducted using equated scores examined gender differences within countries for each of the four soft skills standard's subscales. The achievement differences studied in this research question used the sample split by country.

Research Question #6: (Exploratory) Are achievement differences evident between country samples?

An exploratory analysis into mean achievement scores was used to examine mean achievement differences between countries. The achievement differences studied in this research question also used the sample split by country.

Research Question #7: Are the soft skills standards' subscales equally reliable across countries?

To examine the extent to which country-level differences with respect to subscales need to be addressed, Cronbach's alpha coefficients and item-level statistics were considered for each country independently by form and subscale using the non-equated scores. The reliabilities studied in this research question used the sample split by country.

FINDINGS

Reliability for the Combined Sample

Research Question #1: Does the reliability of each test form at both the overall test level and at the level of each soft skills standard meet the Cronbach's alpha reliability targets of 0.9 and 0.7, respectively?

Results from the classical item analysis of field test data using the combined sample of students from Rwanda, Senegal, Tanzania, and Uganda were consistent with study goals and expectations set after the re-pilot in March 2019.

On each of the three forms administered in the field test, Cronbach's alpha coefficients for Communication, Interpersonal Skills, Dependability and Problem Solving/Critical Thinking subscales reached or exceeded the .70 benchmark. When items across subscales were combined to produce a total work readiness score, the obtained Cronbach's alpha coefficient exceeded .90 on each of the three forms. Consistent with the re-pilot, correlations of the four subscale scores within each form were statistically significant and high (>.60), and scree plots examined as part of form-based factor analyses suggested a dominant underlying factor.

Reliabilities by Form and Standard Across Countries
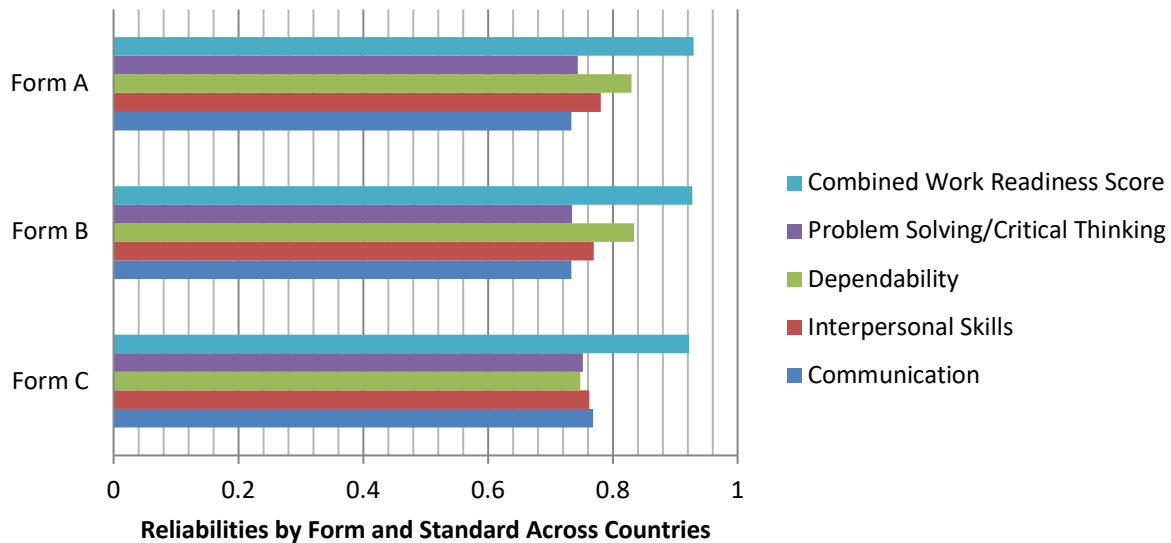
Figure 2: Reliabilities by form and soft skills standard using combined country data

The number of students in the combined sample taking each form were 783, 737 and 720 for forms A, B and C, respectively.

Figure 2 shows the overall reliablity of all items, representing a combined work readiness score for each form as well as the reliability of each soft skills standard's subscale. Scores conceptually range from zero to one, with low scores indicating items hang together poorly, and higher scores indicating items hang together strongly. Since the calculation of Cronbach's alpha is influenced both by the internal consistency of items and the number of items examined, the higher alphas of the combined scores relative to the soft skills standards' subscale scores were expected. As shown in Figure 2, the alphas for the combined work readiness score were comparable across forms, ranging from .92 on form C to .93 on form A. When considering subscale alphas, the reliabilities of the scores for the Communications subscale ranged from .73 on forms A and B to .77 on form C. The reliabilities for the Interpersonal Skills subscale scores ranged from .76 on form C to .78 on form A. Dependability items also proved reliable, ranging from .75 on form C to .83 on form B. The shorter Problem Solving/Critical Thinking subscales ranged from .74 to .75.

Item-level Results for the Combined Sample

Research Question #2: Are the levels of difficulty for each soft skills standard's subscale (Communication, Interpersonal Skills, Dependability, and Problem Solving/Critical Thinking) comparable across forms?

Research Question #3: Should items be removed or revised based on the results of the item analysis?

Overall, the classical item analysis across forms, combined with an examination of equated scores, shows that there are differences in the difficulty of the three forms. This difference favors the use of the score equivalence tables, presented in Appendix E, when comparing raw scores across forms. While forms meet the reliability targets set after the re-pilot in March 2019, item analysis results suggest that overall reliability

could be slightly improved on several soft skills standards' subscales by the removal of a handful of misfitting items.

Overviews of item and subscale difficulty across forms are detailed in the tables below. Each table shows the number of items that were particularly difficult (less than 50% of students answered correctly), items that were difficult (51-65% answered correctly), more moderate items (66 to 84% answered correctly), or easy items (85% or more of students answered correctly). Raw score averages are also presented in the tables. Since each soft skills standards' raw scores are based on different numbers of items across forms, and groups taking each form were not randomly assigned, equated means are also provided. Detailed item statistics for individual forms and items can be found in Appendix A.

Communication (combined sample)

Communication Number of Items in Item Difficulty Categories by Form								
Form	N	Raw Mean	Equated Mean	50% or less	51-65%	66-79%	80-84%	85%+
Form A (18 items)	783	11.57	11.57	5	3	8	1	1
Form B (17 items)	737	11.03	10.83	5	1	9	1	1
Form C (19 items)	720	13.56	11.05	2	4	6	6	1

Table 2: Number of Items by Difficulty Level, by Form for Communication (combined sample)

Communication items on each form spread across the four broad categories of difficulty with most items being answered correctly by between 66-84% of students, as seen in Table 2. Raw score averages appear similar when adjusting for the differing number of Communication items appearing on each form. When considering the equated means, there is less than 1 point difference between group averages. Despite its greater number of items, form C proved somewhat easier based on fewer items in the most difficult category (50% or less), as well as the higher raw and equated means compared to forms A and B. The item level results in Appendix A bear out this picture. With roughly equivalent groups taking each form, form C has seven items that more than 80% of students answered correctly, while forms A and B only have two such items.

Examining the individual item-level results for form A reveals one item that could, based solely on statistical critera, be removed to improve the Communication subscale on that form. Item A7 proved to be difficult for most students with only 52% answering it correctly. The item had a low but positive item-total correlation (.11) indicating that the item did a poor job of seperating high-achieving students from low-achieving students on the Communication subscale. A team review of the item did not make clear why the item statistics were poor compared to other items on the subscale and its content was not determined to provide a sufficiently teachable example to warrant keeping it in the tool, when not contributing to the overall subscale. Given that all other item-total correlations were greater than 0.2 on form A, the removal of this item would increase the reliability of the subscale from .73 to .74.

On form B, a different difficult item (A14B15) was equally appealing to high and low performers alike with an item discrimination of .09. As a link item, this item performed well on form A, but remained a difficult

item on both forms. A review of both additional item statistics and the item itself indicated that more than one answer option could plausibly be correct. Based on field test results, this item could be kept if revised.

Interpersonal Skills (combined sample)

Interpersonal Skills Number of Items in Item Difficulty Categories by Form								
Form	N	Raw Mean	Equated Mean	50% or less	51-65%	66-79%	80-84%	85%+
Form A (17 items)	783	11.50	11.50	1	5	9	1	1
Form B (17 items)	737	12.38	11.23	1	3	8	2	3
Form C (19 items)	720	13.17	11.33	1	4	12	2	0

Table 3: Number of Items by Difficulty Level, by Form for Interpersonal Skills (combined sample)

The distribution of difficulties for items in the Interpersonal Skills subscale, while similar across forms, shows more easy items and slightly fewer moderately difficult items on form B. Indeed, looking at more detailed item level results in Appendix A, form B has five items with difficulties greater than 80%, while only two such items appear on either forms A or C. After adjusting for form difficulty, equated scores for the three groups of test-takers show that average performance was comparable across the three Interpersonal Skills subscales.

The reliability of the Interpersonal Skills items was high across all forms, approaching .80, and all item-total correlations were positive – revealing that scores on the individual items were consistent with the students' total scores on the Interpersonal Skills section. On form A, the two items with the lowest item-total correlations (item A20C36 and item A26) were also among the most difficult items. Item A20C36 also proved to be the most difficult item on Form C with the same item difficulty (percent correct) of 33%. Each of these items had an item-total correlation of approximately .20. While the reliability statistic could be improved for this particular sample of students by removing these items, the solid item discriminations combined with the item dificulty suggest that these items should be retained to keep the reliability stable for more high achieving groups. Both items were reviewed and found to be difficult because more than one answer option could plausibly be correct. Nonetheless, reviewers agreed that the response keyed by test developers was the best of the plausible answers.

On form B, all item-total correlations were above .20, with one exception. This item – item B34C21 – had an item-total correlation near zero, indicating that high and low achievers were equally likely to select the correct response. This same item serves as a link item with form C, where it also performs poorly relative to the other Interpersonal Skills items. Were this a particularly easy item, the low item-total would not be indicative of a problem. Given that the item is among the most difficult on both forms B and C with only 56% of youth correctly answering the item, this item required review to determine if it should be discarded or revised. Seemingly basic, this item asks students to identify, from provided options, the most important factor in helping workers work cooperatively. Both high and low achieving students were attracted to the option "Courage" which was intended as a distractor answer option. While inspiring in a sense, it was ultimately decided to drop the item as it did not directly tap into the work readiness standard being examined.

Dependability (combined sample)

Dependability Number of Items in Item Difficulty Categories by Form								
Form	N	Raw Mean	Equated Mean	50% or less	51-65%	66-79%	80-84%	85%+
Form A (20 items)	783	13.35	13.35	1	7	11	1	0
Form B (21 items)	737	14.50	10.44	3	4	7	4	3
Form C (18 items)	720	11.21	11.53	3	4	11	0	0

Table 4: Number of Items by Difficulty Level, by Form for Dependability (combined sample)

Looking at the Dependability subscale items administered on the field test forms, those on form B appear easiest overall. Along with three items which over 85% of students answered correctly, the form also has 7 items with item difficulties over 80% as compared with only one such item on form A, and no items on form C. Two of the link items across all three forms in particular (A50B55C55 and A42B44C44) appear as the easiest items on form C, among the easier items on form B and in the middle of the difficulty distribution on form A. When both the difficulty levels of items and the lengths of the forms are accounted for, the equated means for the group of test-takers who completed form B drops below the means for the groups who took forms A and C. This implies that if form B test-takers completed form A, their score would be lower.

Looking across forms, item A37B36C53 proved to be a difficult item with item difficulties ranging from .45 to .55. On two of the assessment forms, its removal would increase the alpha of the Dependancy subscale for the groups tested. After a review of the item, and given that the item positively discriminates between high and low achievers in all groups, these results do not warrant the revision or removal of the item.

Forms B and C each contain one item (B49 and C56) with negative item discrimination, meaning low achievers on the Dependability subscales had a greater chance of answering the item correctly than high achievers did. In each case these were the most difficult items on the subscale, with only 13% and 14% responding correctly. Item 49 on form B was reviewed by research staff and determined to be as conceptually a bad fit to the subscale as its statisics suggest. Item 56 on form C was determined to have been miskeyed by test developers, but nonetheless considered for removal when re-analyzed.

Problem Solving/Critical Thinking (combined sample)

Problem Solving/Critical Thinking Number of Item in Item Difficulty Categories by Form								
Form	N	Raw Mean	Equated Mean	50% or less	51-65%	66-79%	80-84%	85%+
Form A (15 items)	783	9.37	9.37	2	7	5	0	1
Form B (15 items)	737	9.59	9.59	2	5	7	0	1
Form C (15 items)	720	9.22	8.86	3	6	5	0	1

Table 5: Number of Items by Difficulty Level, by Form for Problem Solving/Critical Thinking (combined sample)

Students across each of the three forms scored similarly on average in the Problem Solving/Critical Thinking section, as evidenced in the Raw Mean column of Table 5. Comparing the raw means across groups works for two reasons – (1) the Problem Solving/Critical Thinking subscales have the same number of items on each form, and (2) more than half of the items in this subscale are link items across the three forms. Even the distribution of item difficulties was similar with form B having slightly more easier items with item difficulties in the 66-84% range. When equated using the Rasch model with the other Problem Solving/Critical Thinking subscales based on the link item, the equated mean on form C drops slightly.

While all item-total correlations were positive, one item on form B proved both difficult and a poor distinguisher of high- and low- performing students with respect to the Problem Solving/Critical Thinking subscales. Item 68 on form B was only answered correctly by 38% of students and had an item-total correlation of .06. This item was reviewed and had already been revised based on analysis of prior pilot results. As it was not considered an optimal item for the Problem Solving/Critical Thinking subscale to begin with, its poor item characteristics on the field test suggest it should be removed from the tool. Its removal would increase the overall reliablity of the subscale from .74 to .75.

Gender Differential Item Functioning (DIF) Analyses

Research Question #4: Is gender bias present, and does the presence of biased items threaten the validity of inferences made from soft skills standard-level and total work readiness scores?

Overall, the results of the field test indicate that DIF with respect to gender is not a major problem either at the form level or the standard level. While several items exhibited statistically significant DIF, effect size measures were small and almost as many of the items displaying DIF favored girls over boys as the ones that favored boys over girls.

A summary of the number of items on each form that display significant gender based DIF based on Chi-square tests of both uniform DIF and a simultaneous test of uniform and non-uniform DIF for each soft skills standard across forms can be seen in Figure 3. The relevant statistics for individual items are available in Appendix B.

Statistically Significant Gender Chi-Squre DIF Test by Form
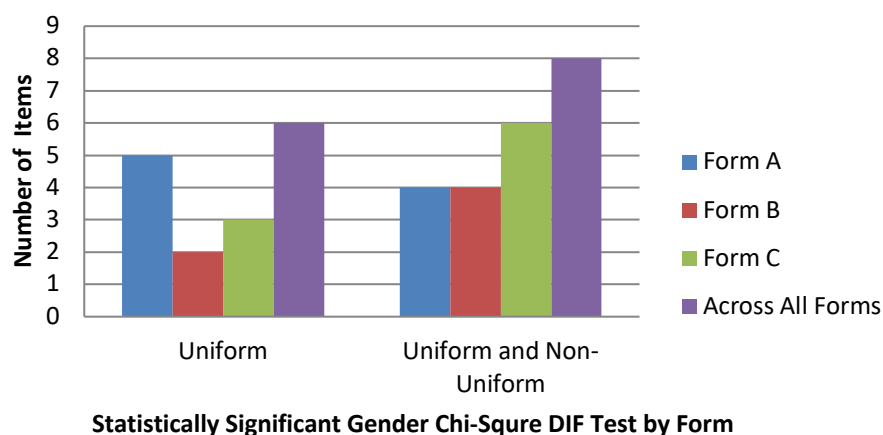
Figure 3: Statistically significant gender Chi-square DIF test, by form

The results show that 14 out of the 132 items that appear across all three forms (10.60%) displayed statistically significant DIF with respect to gender. Of these, 8 items favored boys and 6 items favored girls. While the total number of items displaying DIF in favor of boy and girls is comparable, 5 of the 14 items displaying statistically significant DIF appeared across all three forms. Each of these five items statistically favored boys.

Despite the statistically significant DIF with respect to gender mentioned before, taking the effect size into consideration suggests that gender DIF is not a problem across forms or soft skills standard scores. An effect size measure was computed that consists of the difference between the Nagelkerke R-square value with gender added to the model, and the value when only the total score is used to predict item responses. Using the commonly applied criteria that an item should be considered as displaying DIF if both the Chi-square is statistically significant and the effect size is at least .13, none of the items displaying statistically significant DIF in favor of either males or females reached practical significance. The largest effect size was for an item on Form C (.02), favoring girls. Item-level results can be found in Appendix B.

Although the effect size results suggest that the statistically significant differences with respect to gender were not very large, it would still be wise to adjust the forms based on the gender biases found to ensure that forms do not favor one gender over another when applied to different populations.

Gender Differences in Achievement within Countries

Research Question #5: (Exploratory) Are there achievement differences with respect to gender within the countries and samples tested?

Patterns showed that there were differences in performance based on gender. These results are consistent with the DIF analyses suggesting that gender differences are the result of group differences (differences between test-takers) rather than form bias.

Equated Score Means		Rwanda			Senegal			Tanzania			Uganda		
		N	Mean	Se	N	Mean	Se	N	Mean	Se	N	Mean	Se
Communication	Female	331	12.73	0.15	309	9.85	0.2	281	8.58	0.27	328	11.82	0.16
	Male	274	**13.22***	0.14	205	**11.21***	0.23	265	**10.04***	0.24	227	11.73	0.17
Interpersonal Skills	Female	331	12.96	0.13	309	10.64	0.2	281	8.03	0.23	328	12.35	0.15
	Male	274	**13.35***	0.14	205	**11.61***	0.2	265	**9.57***	0.23	227	12.16	0.18
Dependability	Female	331	13.46	0.19	309	9.92	0.26	281	8.59	0.28	328	13.8	0.19
	Male	274	13.8	0.22	205	**11.81***	0.29	265	9.37	0.28	227	13.52	0.25
Problem Solving/Critical Thinking	Female	331	11.15	0.15	309	8.09	0.19	281	6.76	0.21	328	10.15	0.15
	Male	274	11.3	0.17	205	**8.97***	0.21	265	**7.74***	0.2	227	9.76	0.18

Table 6: Equated means by gender across soft skills standards, within countries

Gender differences within each country based on the equated scores suggest that males, on average, outperformed females in Rwanda, Senegal and Tanzania with respect to the Communication and Interpersonal Skills standards. In Senegal, males also statistically outperformed females with respect to the Dependability and Problem Solving/Critical Thinking standards. In Tanzania, there was also a significant mean difference in group performance between males and females with respect to Problem Solving/Critical Thinking. These observed differences in average performance were small, however, with most being less than a point difference. Notably, in Uganda, there were no statistically significant differences in performance between males and females with respect to any of the four standards measured. While the differences were not statistically significant, the observed means were higher for females than males across the groups tested in Uganda.

Looking at country by form data (Appendix E), gender differences with respect to form appear to be caused by differences in group achievement. In Senegal, for instance, significant gender differences (p<.05) were observed for all soft skills standards on forms A and B, but no significant gender differences were found on any soft skills standards scores on form C. In Tanzania, on the other hand, there were significant gender differences on all standards on form C but not on forms A and B. Similarly, significant differences between males and females on the Interpersonal Skills standard were found on forms A and B in Senegal and form C in Tanzania. In each of these cases, the groups taking each form were not identical, implying that these findings are due to differences in the test-taker rather than problems with the forms.

Differences in Achievement between Countries

Research Question #6: (Exploratory) Are achievement differences evident between country samples?

Scores on the three forms – A, B, and C – were equated using Item-response theory (the Rasch model) to account for potential difficulty differences between forms when comparing results. The equated scores express raw scores on forms B and C on the same scale as form A to allow a single standard score to be computed within each country. The means for each soft skills standard across countries indicates that youth in Rwanda and Uganda performed highest overall. The sample of students from Rwanda had the highest scores for the Communication, Interpersonal Skills and Problem Solving/Critical Thinking standards, and youth in Uganda had the highest average performance on the Dependability standard.[6]

---

[6] All country mean differences within soft skills standards were statistically significant (*p*<0.05), with the exception of the Ugandan and Rwandan means for Dependability.
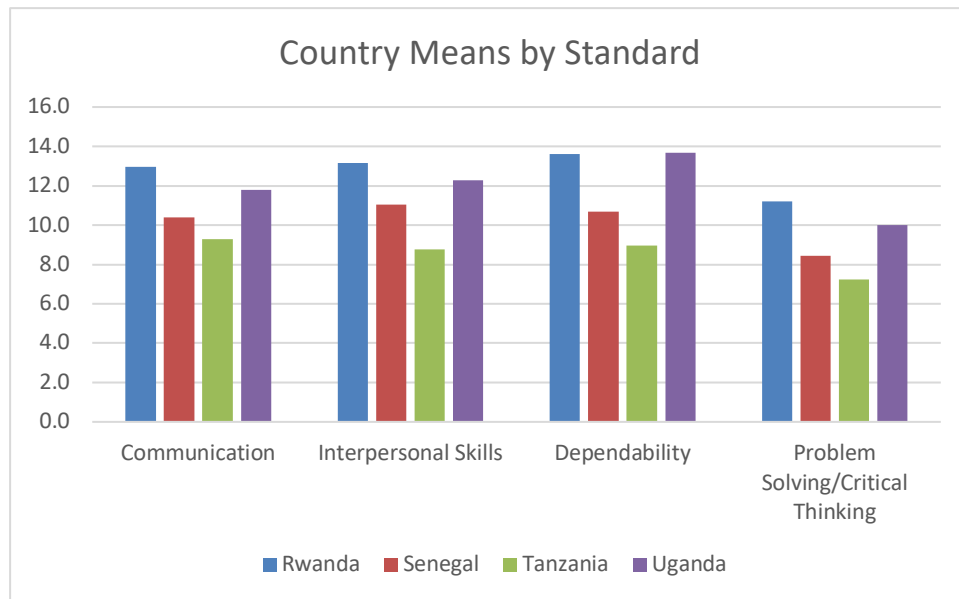
Figure 4: Country means by soft skills standard

While the samples from each country were not completely representative of the student populations, a goal of the study was to build an assessment tool that would be appropriate for in-school and out-of-school youth. To this end, both in-school and out-of-school youth were represented in the analysis sample in both the field test and the earlier second pilot round. Samples of in-school youth were represented by Senegal and Uganda; out-of-school youth were represented by Rwanda and Tanzania.

The resulting average performance of each group of students did not favor a particular group. Out-of-school youth in Rwanda and Tanzania represented both the highest and lowest overall achievers across soft skills standard subscales.

Reliabilities at the Country Level

Research Question #7: Are the soft skills standards' subscales equally reliable across countries?

The reliability statistics that use data combined from all four countries participating in the field test suggest that all subscales met or exceeded the target reliability of .70. Individual country-level data, however, shows that reliability varies substantially across countries, soft skills standard subscales, and forms.

Communication

Communications Form	Items	Senegal	Tanzania	Rwanda	Uganda
A	18	0.702	0.784	0.634	0.631
B	17	0.694	0.818	0.478	0.494
C	19	0.694	0.856	0.573	0.608

Table 7: Reliability of Communication subscale, by country

The Cronbach's alpha reliability statistics for Communication ranged from approximately .70 and .78 or higher in Tanzania to lows of .48 and .49 in Rwanda and Uganda, respectively. This country-to-country difference in reliability is due, in part, to the selection process of items for the field test forms. Items were chosen for the field test forms based on a second round of pilot data collection in Senegal and Tanzania, after revisions borne out of analysis of the first pilot round of data.

An examination of equated means by country for the Communication standard showed that youth in Rwanda and Uganda performed better on average than youth in Tanzania and Senegal (13 and 12, as compared with 10 and 9). While small, these differences are statistically significant.

In a more revealing result, Rwanda and Uganda had a higher percentage of items that 80-90% of youth answered correctly, as indicated by the country-level item statistics in Appendix B. In short, mastery of the Communication items was higher in Rwanda and Uganda, but the items were less effective at separating the highest and lowest achievers. It is important to underscore that in general, a more reliable scale is better for measurement purposes, but the lower reliabilities observed in Rwanda and Uganda as compared with Tanzania and Senegal are not necessarily a concern for this tool. While the lower reliabilities tell us that the tool is doing a poorer job of separating high and low achieving students in these countries, because there are a large number of high achievers in Rwanda and Senegal, this indicates a more nuanced situation. While low reliabilities could indicate that the tool is not functioning to measure mastery level work-readiness skills, here it indicates that youth in the samples tested in these countries have largely attained mastery at the level measured. This is not surprising when you consider that the sample groups in both Rwanda and Senegal had already undergone work readiness training when they participated in the field test. What remains to be determined is if test results have predictive validity; in other words, does obtaining mastery on the test translate to the real world outcomes we are hoping to influence.

Interpersonal Skills

Interpersonal Form	Items	Senegal	Tanzania	Rwanda	Uganda
A	17	0.717	0.713	0.662	0.687
B	17	0.746	0.807	0.544	0.493
C	19	0.732	0.838	0.547	0.656

Table 8: Reliability of Interpersonal Skills subscale, by country

The reliability statistics for Interpersonal Skills were similar across countries. The reliability of the Interpersonal Skills items remained consistently strong across forms in Senegal and Tanzania, but was below .60 in Rwanda and Uganda. Again, groups tested in Rwanda and Uganda had higher average scores once forms were equated, as well as a higher percentage of easier items.

Dependability

Dependability Form	Items	Senegal	Tanzania	Rwanda	Uganda
A	20	.794	.837	.738	.746
B	21	.805	.835	.776	.667
C	18	.711	.758	.650	.631

Table 9: Reliability of Dependability subscale, by country

While average equated form scores for Dependability remained higher for youth in Rwanda and Uganda (almost 14 points) than in Senegal (11) and Tanzania (9), this soft skill standards' subscale items did a better job of separating high and low achievers across countries than the Communication and Interpersonal Skills items did.

Problem Solving/Critical Thinking

Problem Solving Form	Items	Senegal	Tanzania	Rwanda	Uganda
A	15	0.718	0.614	0.667	0.666
B	15	0.641	0.672	0.692	0.547
C	15	0.664	0.741	0.687	0.563

Table 10: Reliability of Problem Solving/Critical Thinking subscale, by country

At the overall level, the Cronbach's alpha reliability statistics indicated that the assessment met the goal of achieving a reliability of at least .70 for the Problem Solving/Critical Thinking soft skills standard subscale, as previously shown in Figure 2. However, when considering countries independently, only Senegal on form A and Tanzania on form C had reliability statistics above .70. The ability of the items in the Problem Solving/Critical Thinking subscale to separate high and low achievers could be increased by including additional items in this standard.

CONCLUSIONS

Taken together, the analyses presented in this report indicate that the three forms, each measuring Communication, Interpersonal Skills, Dependability, and Problem Solving/Critical Thinking, reliably measure foundational skills for work readiness for the sample of students in Rwanda, Senegal, Tanzania and Uganda that were tested. The number of problem items were few, but item analysis did indicate that results could be further improved with minor revision by either removing or reworking several items. Importantly the results did not show that the tools systematically favored either boys or girls when matched on ability. While some individual items did favor either boys or girls, the small effect sizes indicate that gender bias was not a substantial problem. As some items did exhibit statistically significant DIF, however, more work needs to be done to understand why individual item scenarios favored either boys or girls. When such differences arise, items can reflect real gender or cultural differences that are worth exploring. Likewise, these kinds of differences can indicate that an item is measuring something unintended, which can lead to improper generalizations from responses.

While strict form equivalence may not be necessary given the tool's formative focus, score equivalence tables such as the one contained in the appendix can be used in the field to facilitate subscale comparisons across forms based upon existing tool. If Items are further revised, or implementation of items in the field is customized, the existing score equivalence form will no longer be applicable.

When looking at the reliability of subscales across countries, the varied patterns across forms suggest that more difficult items would be required to more finely measure work-readiness levels for a higher achieving group. The lower reliability of some subscale scores observed in Uganda and Rwanda, for example, reflect the high achievement of the test-takers in these countries in relation to the difficulty of items. What remains to be explored is the extent to which mastery of the work readiness tools, and the foundational skills represented on them, has predictive validity with respect to practical work readiness and job attainment. Along with work to optimize the tools' ability to assess gaps in foundational skills, establishing the tools ultimate predictive validity is the next logical step for determining its overall validity. Other evident avenues of interest would include how best to integrate the use of the tool in formative instruction to maximize desired outcomes.

APPENDICES

Appendix A1

Form A Combined Sample Item Statistics: Communication

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
A	1	A1	Communication	73%	.285	.724
A	2	A2_B3_C1	Communication	83%	.293	.724
A	3	A3_B5	Communication	38%	.242	.729
A	4	A4	Communication	67%	.355	.718
A	5	A5_B16_C17	Communication	75%	.419	.713
A	6	A6_B7_C9	Communication	46%	.231	.730
A	7	A7	Communication	52%	.114	.741
A	8	A8_C12	Communication	63%	.365	.717
A	9	A9_B10	Communication	56%	.248	.728
A	10	A10_B11	Communication	78%	.468	.709
A	11	A11	Communication	73%	.384	.716
A	12	A12_B13_C18	Communication	79%	.329	.721
A	13	A13_B12_C11	Communication	75%	.361	.718
A	14	A14_B15	Communication	44%	.245	.729
A	15	A15_B14_C16	Communication	86%	.341	.721
A	16	A16_B1	Communication	50%	.378	.716
A	17	A17_B2_C19	Communication	44%	.333	.720
A	18	A18_B17_C5	Communication	73%	.280	.725

Appendix A2

Form A Combined Sample Item Statistics: Interpersonal Skills

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
A	19	A19	Interpersonal Skills	69%	.451	.764
A	20	A20_C36	Interpersonal Skills	33%	.203	.783
A	21	A21	Interpersonal Skills	76%	.252	.778
A	22	A22	Interpersonal Skills	53%	.282	.777
A	23	A23_B30	Interpersonal Skills	79%	.424	.766
A	24	A24_C26	Interpersonal Skills	67%	.450	.764
A	25	A25	Interpersonal Skills	87%	.342	.772
A	26	A26	Interpersonal Skills	58%	.202	.783
A	27	A27	Interpersonal Skills	51%	.413	.766
A	28	A28	Interpersonal Skills	74%	.421	.766
A	29	A29	Interpersonal Skills	68%	.442	.764
A	30	A30_B31	Interpersonal Skills	62%	.404	.767
A	31	A31_B32	Interpersonal Skills	77%	.510	.760
A	32	A32	Interpersonal Skills	65%	.372	.770
A	33	A33_B26	Interpersonal Skills	69%	.291	.776
A	34	A34_B28_C37	Interpersonal Skills	79%	.528	.759
A	35	A35_B33_C35	Interpersonal Skills	83%	.302	.775

Appendix A3

Form A Combined Sample Item Statistics: Dependability

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
A	36	A36	Dependability	83%	.532	.818
A	37	A37_B36_C53	Dependability	55%	.267	.830
A	38	A38	Dependability	75%	.214	.831
A	39	A39	Dependability	67%	.475	.819
A	40	A40_C54	Dependability	53%	.331	.827
A	41	A41	Dependability	43%	.347	.826
A	42	A42_B44_C44	Dependability	78%	.416	.822
A	43	A43	Dependability	60%	.372	.824
A	44	A44	Dependability	60%	.409	.822
A	45	A45	Dependability	75%	.307	.827
A	46	A46	Dependability	58%	.349	.826
A	47	A47	Dependability	73%	.617	.813
A	48	A48	Dependability	67%	.343	.826
A	49	A49	Dependability	56%	.393	.823
A	50	A50_B55_C55	Dependability	74%	.431	.821
A	51	A51	Dependability	79%	.608	.814
A	52	A52	Dependability	73%	.470	.820
A	53	A53	Dependability	71%	.328	.826
A	54	A54_B37	Dependability	78%	.524	.818
A	55	A55_B54_C51	Dependability	56%	.429	.821

Appendix A4

Form A Combined Sample Item Statistics: Problem Solving/Critical Thinking

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
A	56	A56	Problem Solving/Critical Thinking	73%	.371	.728
A	57	A57_B56_C57	Problem Solving/Critical Thinking	86%	.487	.721
A	58	A58	Problem Solving/Critical Thinking	74%	.315	.733
A	59	A59	Problem Solving/Critical Thinking	50%	.213	.744
A	60	A60	Problem Solving/Critical Thinking	73%	.290	.735
A	61	A61	Problem Solving/Critical Thinking	55%	.412	.723
A	62	A62	Problem Solving/Critical Thinking	65%	.528	.711
A	63	A63_B60_C64	Problem Solving/Critical Thinking	51%	.365	.728
A	64	A64_B61_C65	Problem Solving/Critical Thinking	61%	.347	.730
A	65	A65_B62_C66	Problem Solving/Critical Thinking	64%	.256	.739
A	66	A66_B63_C67	Problem Solving/Critical Thinking	59%	.284	.737
A	67	A67_B64_C68	Problem Solving/Critical Thinking	65%	.278	.737
A	68	A68_C69	Problem Solving/Critical Thinking	28%	.256	.739
A	69	A69_B66_C70	Problem Solving/Critical Thinking	66%	.526	.711
A	70	A70_B67_C71	Problem Solving/Critical Thinking	67%	.300	.735

Appendix A5

Form B Combined Sample Item Statistics: Communication

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
B	1	A16_B1	Communication	47%	.336	.720
B	2	A17_B2_C19	Communication	43%	.256	.728
B	3	A2_B3_C1	Communication	78%	.366	.718
B	4	B4	Communication	69%	.383	.716
B	5	A3_B5	Communication	40%	.250	.729
B	6	B6	Communication	78%	.409	.714
B	7	A6_B7_C9	Communication	43%	.216	.732
B	8	B8	Communication	79%	.252	.728
B	9	B9	Communication	67%	.275	.726
B	10	A9_B10	Communication	52%	.259	.728
B	11	A10_B11	Communication	77%	.476	.708
B	12	A13_B12_C11	Communication	74%	.380	.716
B	13	A12_B13_C18	Communication	81%	.357	.719
B	14	A15_B14_C16	Communication	87%	.335	.722
B	15	A14_B15	Communication	41%	.093	.744
B	16	A5_B16_C17	Communication	73%	.483	.706
B	17	A18_B17_C5	Communication	75%	.393	.715

Appendix A6

Form B Combined Sample Item Statistics: Interpersonal Skills

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
B	18	B18	Interpersonal Skills	76%	.317	.762
B	19	B19	Interpersonal Skills	70%	.406	.755
B	20	B20	Interpersonal Skills	44%	.290	.766
B	21	B21	Interpersonal Skills	66%	.297	.764
B	22	B22	Interpersonal Skills	86%	.448	.754
B	23	B23	Interpersonal Skills	74%	.521	.746
B	24	B24	Interpersonal Skills	89%	.508	.752
B	25	B25	Interpersonal Skills	88%	.526	.750
B	26	A33_B26	Interpersonal Skills	69%	.279	.766
B	27	B27	Interpersonal Skills	65%	.266	.767
B	28	A34_B28_C37	Interpersonal Skills	77%	.500	.748
B	29	B29	Interpersonal Skills	69%	.229	.770
B	30	A23_B30	Interpersonal Skills	81%	.393	.757
B	31	A30_B31	Interpersonal Skills	60%	.395	.756
B	32	A31_B32	Interpersonal Skills	84%	.424	.755
B	33	A35_B33_C35	Interpersonal Skills	79%	.421	.754
B	34	B34_C21	Interpersonal Skills	56%	.063	.785

Appendix A7

Form B Combined Sample Item Statistics: Dependability

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
B	35	B35	Dependability	87%	.467	.826
B	36	A37_B36_C53	Dependability	47%	.222	.837
B	37	A54_B37	Dependability	82%	.409	.827
B	38	B38	Dependability	85%	.556	.822
B	39	B39	Dependability	82%	.587	.820
B	40	B40	Dependability	78%	.582	.820
B	41	B41	Dependability	53%	.423	.827
B	42	B42	Dependability	83%	.461	.825
B	43	B43	Dependability	83%	.387	.828
B	44	A42_B44_C44	Dependability	78%	.461	.825
B	45	B45	Dependability	75%	.399	.828
B	46	B46	Dependability	54%	.348	.830
B	47	B47	Dependability	85%	.466	.825
B	48	B48	Dependability	72%	.516	.822
B	49	B49	Dependability	13%	-.167	.847
B	50	B50	Dependability	64%	.359	.830
B	51	B51	Dependability	70%	.512	.822
B	52	B52	Dependability	73%	.337	.830
B	53	B53	Dependability	61%	.325	.831
B	54	A55_B54_C51	Dependability	48%	.321	.832
B	55	A50_B55_C55	Dependability	76%	.564	.820

Appendix A8

Form B Combined Sample Item Statistics: Problem Solving/Critical Thinking

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
B	56	A57_B56_C57	Problem Solving/Critical Thinking	88%	.471	.712
B	57	B57	Problem Solving/Critical Thinking	63%	.388	.715
B	58	B58	Problem Solving/Critical Thinking	72%	.383	.716
B	59	B59	Problem Solving/Critical Thinking	67%	.312	.723
B	60	A63_B60_C64	Problem Solving/Critical Thinking	58%	.382	.716
B	61	A64_B61_C65	Problem Solving/Critical Thinking	70%	.351	.719
B	62	A65_B62_C66	Problem Solving/Critical Thinking	64%	.252	.730
B	63	A66_B63_C67	Problem Solving/Critical Thinking	57%	.330	.722
B	64	A67_B64_C68	Problem Solving/Critical Thinking	63%	.372	.717
B	65	B65	Problem Solving/Critical Thinking	32%	.227	.732
B	66	A69_B66_C70	Problem Solving/Critical Thinking	70%	.517	.702
B	67	A70_B67_C71	Problem Solving/Critical Thinking	67%	.263	.729
B	68	B68	Problem Solving/Critical Thinking	38%	.060	.750
B	69	B69	Problem Solving/Critical Thinking	78%	.488	.706
B	70	B70	Problem Solving/Critical Thinking	74%	.342	.720

Appendix A9

Form C Combined Sample Item Statistics: Communication

Form	Item #	item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
C	1	A2_B3_C1	Communication	79%	.311	.760
C	2	C2	Communication	83%	.339	.758
C	3	C3	Communication	83%	.408	.754
C	4	C4	Communication	58%	.415	.752
C	5	A18_B17_C5	Communication	75%	.325	.759
C	6	C6	Communication	80%	.393	.755
C	7	C7	Communication	73%	.358	.757
C	8	C8	Communication	81%	.513	.747
C	9	A6_B7_C9	Communication	47%	.174	.772
C	10	C10	Communication	71%	.273	.763
C	11	A13_B12_C11	Communication	77%	.347	.758
C	12	A8_C12	Communication	65%	.313	.760
C	13	C13	Communication	83%	.339	.758
C	14	C14	Communication	59%	.358	.757
C	15	C15	Communication	58%	.297	.762
C	16	A15_B14_C16	Communication	85%	.322	.760
C	17	A5_B16_C17	Communication	76%	.438	.751
C	18	A12_B13_C18	Communication	82%	.290	.762
C	19	A17_B2_C19	Communication	40%	.269	.764

Appendix A10

Form C Combined Sample Item Statistics: Interpersonal Skills

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
C	20	C20	Interpersonal Skills	72%	.378	.748
C	21	B34_C21	Interpersonal Skills	56%	.081	.772
C	22	C22	Interpersonal Skills	64%	.312	.754
C	23	C23	Interpersonal Skills	78%	.523	.739
C	24	C24	Interpersonal Skills	76%	.220	.760
C	25	C25	Interpersonal Skills	70%	.232	.759
C	26	A24_C26	Interpersonal Skills	69%	.375	.749
C	27	C27	Interpersonal Skills	58%	.280	.756
C	28	C28	Interpersonal Skills	77%	.435	.745
C	29	C29	Interpersonal Skills	77%	.547	.737
C	30	C30	Interpersonal Skills	67%	.430	.744
C	31	C31	Interpersonal Skills	64%	.327	.752
C	32	C32	Interpersonal Skills	81%	.413	.747
C	33	C33	Interpersonal Skills	74%	.303	.754
C	34	C34	Interpersonal Skills	72%	.397	.747
C	35	A35_B33_C35	Interpersonal Skills	84%	.278	.756
C	36	A20_C36	Interpersonal Skills	33%	.179	.764
C	37	A34_B28_C37	Interpersonal Skills	78%	.466	.742
C	38	C38	Interpersonal Skills	66%	.211	.761

Appendix A11

Form C Combined Sample Item Statistics: Dependability

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
C	39	C39	Dependability	70%	.353	.734
C	40	C40	Dependability	70%	.446	.726
C	41	C41	Dependability	68%	.200	.747
C	42	C42	Dependability	74%	.219	.745
C	43	C43	Dependability	72%	.455	.726
C	44	A42_B44_C44	Dependability	78%	.380	.733
C	45	C45	Dependability	73%	.524	.720
C	46	C46	Dependability	67%	.340	.735
C	47	C47	Dependability	62%	.462	.724
C	48	C48	Dependability	57%	.337	.735
C	49	C49	Dependability	54%	.357	.734
C	50	C50	Dependability	67%	.362	.733
C	51	A55_B54_C51	Dependability	49%	.325	.737
C	52	C52	Dependability	73%	.380	.732
C	53	A37_B36_C53	Dependability	52%	.207	.747
C	54	A40_C54	Dependability	45%	.153	.752
C	55	A50_B55_C55	Dependability	74%	.552	.718
C	56	C56	Dependability	14%	-.255	.772

Appendix A12

Form C Combined Sample Item Statistics: Problem Solving/Critical Thinking

Form	Item #	Item	Subscale	Item Difficulty	Item Discrimination	Alpha if Deleted
C	57	A57_B56_C57	Problem Solving/Critical Thinking	85%	.486	.730
C	58	C58	Problem Solving/Critical Thinking	78%	.555	.722
C	59	C59	Problem Solving/Critical Thinking	34%	.175	.756
C	60	C60	Problem Solving/Critical Thinking	67%	.398	.735
C	61	C61	Problem Solving/Critical Thinking	66%	.397	.735
C	62	C62	Problem Solving/Critical Thinking	64%	.346	.740
C	63	C63	Problem Solving/Critical Thinking	74%	.517	.724
C	64	A63_B60_C64	Problem Solving/Critical Thinking	49%	.372	.737
C	65	A64_B61_C65	Problem Solving/Critical Thinking	62%	.328	.742
C	66	A65_B62_C66	Problem Solving/Critical Thinking	63%	.120	.762
C	67	A66_B63_C67	Problem Solving/Critical Thinking	58%	.305	.744
C	68	A67_B64_C68	Problem Solving/Critical Thinking	63%	.339	.741
C	69	A68_C69	Problem Solving/Critical Thinking	28%	.226	.751
C	70	A69_B66_C70	Problem Solving/Critical Thinking	69%	.530	.722
C	71	A70_B67_C71	Problem Solving/Critical Thinking	62%	.312	.743

Appendix B1

Form A Communication Standard Item Statistics by Country

Communications Form A	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
A1	57%	0.10	0.79	61%	0.26	0.69	80%	0.30	0.61	92%	0.24	0.62
A2_B3_C2	85%	0.37	0.78	80%	0.38	0.68	72%	0.33	0.61	94%	0.08	0.64
A3_B5	28%	0.13	0.79	22%	0.13	0.71	19%	0.14	0.63	76%	0.30	0.61
A4	62%	0.52	0.76	67%	0.39	0.68	69%	0.34	0.60	69%	0.18	0.63
A5_B16_C17	57%	0.53	0.76	69%	0.24	0.70	82%	0.19	0.62	89%	0.36	0.61
A6_B7_C9	42%	0.46	0.77	45%	0.11	0.71	56%	0.17	0.63	43%	0.19	0.63
A7	44%	-0.01	0.80	63%	0.42	0.68	75%	0.23	0.62	34%	0.16	0.63
A8_C12	49%	0.46	0.77	54%	0.14	0.71	60%	0.25	0.62	84%	0.30	0.61
A9_B10	38%	0.26	0.78	73%	0.23	0.70	33%	0.13	0.63	77%	0.22	0.62
A10_B11	62%	0.62	0.76	73%	0.33	0.69	89%	0.24	0.62	86%	0.39	0.60
A11	62%	0.61	0.76	76%	0.33	0.69	83%	0.25	0.62	74%	0.23	0.62
A12_B13_C18	59%	0.25	0.78	83%	0.35	0.69	89%	0.30	0.61	86%	0.20	0.63
A13_B12_C11	60%	0.32	0.78	71%	0.30	0.69	77%	0.23	0.62	90%	0.32	0.61
A14_B15	43%	0.37	0.77	36%	0.21	0.70	30%	0.08	0.64	64%	0.17	0.63
A15_B14_C16	70%	0.33	0.78	86%	0.19	0.70	90%	0.21	0.62	96%	0.35	0.62
A16_B1	37%	0.48	0.77	51%	0.40	0.68	56%	0.29	0.61	56%	0.24	0.62
A17_B2_C19	36%	0.36	0.77	33%	0.34	0.68	56%	0.36	0.60	50%	0.22	0.63
A18_B17_C5	73%	0.42	0.77	80%	0.43	0.68	80%	0.23	0.62	64%	0.31	0.61

Appendix B2

Form B Communication Standard Item Statistics by Country

Communications Form B	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
A16_B1	39.7%	0.39	0.81	43.1%	0.47	0.66	48.2%	0.20	0.47	56.3%	0.19	0.46
A17_B2_C19	36.2%	0.21	0.82	28.7%	0.37	0.67	53.3%	0.23	0.46	51.1%	0.07	0.49
A2_B3_C2	72.4%	0.35	0.81	75.9%	0.40	0.67	69.3%	0.33	0.44	95.8%	0.29	0.45
B4	55.2%	0.57	0.80	69.0%	0.25	0.69	69.8%	0.15	0.48	82.1%	0.26	0.44
A3_B5	40.8%	0.31	0.82	19.0%	0.20	0.69	25.6%	0.01	0.51	75.3%	0.22	0.45
B6	63.8%	0.67	0.79	60.9%	0.05	0.71	92.5%	0.15	0.48	91.6%	0.20	0.46
A6_B7_C9	40.8%	0.42	0.81	40.2%	0.19	0.69	57.3%	0.28	0.45	31.6%	0.14	0.47
B8	64.9%	0.34	0.81	82.8%	0.32	0.68	94.5%	0.19	0.48	71.6%	0.13	0.47
B9	56.3%	0.56	0.80	61.5%	0.15	0.70	86.4%	0.14	0.48	61.6%	0.11	0.48
A9_B10	28.7%	0.41	0.81	69.0%	0.21	0.69	41.2%	0.06	0.51	69.5%	0.15	0.47
A10_B11	61.5%	0.63	0.80	64.9%	0.40	0.67	87.4%	0.13	0.49	89.5%	0.22	0.45
A13_B12_C11	59.8%	0.44	0.81	63.8%	0.37	0.67	75.4%	0.11	0.49	94.2%	0.06	0.48
A12_B13_C18	63.2%	0.27	0.82	80.5%	0.38	0.67	87.4%	0.36	0.45	91.1%	0.11	0.47
A15_B14_C16	66.1%	0.33	0.81	91.4%	0.15	0.69	88.9%	0.22	0.47	98.9%	0.19	0.47
A14_B15	38.5%	-0.10	0.84	30.5%	0.17	0.69	25.1%	0.05	0.50	67.9%	0.08	0.48
A5_B16_C17	51.7%	0.62	0.80	67.2%	0.37	0.67	76.4%	0.18	0.48	93.2%	0.22	0.46
A18_B17_C5	64.9%	0.63	0.80	72.4%	0.43	0.67	86.9%	0.09	0.49	72.6%	0.27	0.43

Appendix B3

Form C Communication Standard Item Statistics by Country

Communications Form C	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
C1	73.9%	0.47	0.85	80.2%	0.20	0.69	68.8%	0.25	0.59	93.7%	0.13	0.57
A2_B3_C2	80.1%	0.55	0.85	86.3%	0.41	0.67	86.7%	0.17	0.60	78.3%	0.28	0.54
C3	76.1%	0.69	0.84	76.9%	0.25	0.68	93.1%	0.22	0.60	86.8%	0.07	0.58
C4	44.3%	0.48	0.85	54.4%	0.33	0.68	61.8%	0.38	0.57	70.4%	0.32	0.53
A18_B17_C5	71.6%	0.47	0.85	75.3%	0.40	0.67	81.5%	0.14	0.61	70.4%	0.33	0.53
C6	66.5%	0.52	0.85	78.0%	0.28	0.68	80.9%	0.15	0.60	93.1%	0.26	0.55
C7	66.5%	0.48	0.85	55.5%	0.13	0.70	72.8%	0.27	0.59	95.8%	0.29	0.55
C8	71.6%	0.72	0.84	69.8%	0.31	0.68	88.4%	0.27	0.59	94.2%	0.38	0.54
A6_B7_C9	43.8%	0.23	0.86	48.4%	0.39	0.67	63.6%	0.18	0.60	31.7%	0.02	0.60
C10	63.6%	0.27	0.86	51.6%	0.13	0.70	85.5%	0.27	0.59	84.7%	0.16	0.56
A13_B12_C11	63.6%	0.34	0.85	70.9%	0.29	0.68	79.8%	0.26	0.59	93.7%	0.12	0.57
A8_C12	56.8%	0.46	0.85	59.9%	0.22	0.69	59.5%	0.09	0.62	82.0%	0.24	0.55
C13	74.4%	0.50	0.85	81.9%	0.25	0.69	79.8%	0.20	0.60	93.7%	0.08	0.57
C14	51.7%	0.53	0.85	44.5%	0.24	0.69	77.5%	0.30	0.58	61.9%	0.17	0.57
C15	55.1%	0.58	0.84	52.7%	0.33	0.68	66.5%	-0.05	0.64	58.7%	0.15	0.57
A15_B14_C16	72.7%	0.30	0.86	87.9%	0.26	0.68	81.5%	0.28	0.59	97.9%	0.34	0.56
A5_B16_C17	58.5%	0.63	0.84	76.9%	0.27	0.68	76.9%	0.06	0.62	92.1%	0.44	0.53
A12_B13_C18	61.9%	0.23	0.86	88.5%	0.12	0.69	84.4%	0.35	0.58	93.7%	0.28	0.55
A17_B2_C19	31.3%	0.19	0.86	33.5%	0.37	0.67	52.0%	0.36	0.57	45.0%	0.07	0.59

Appendix B4

Form A Interpersonal Skills Standard Item Statistics by Country

Interpersonal Skills Form A	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
A19	49%	0.41	0.69	59%	0.35	0.70	81%	0.21	0.68	85%	0.44	0.63
A20_C36	10%	0.12	0.71	38%	0.19	0.72	60%	0.11	0.70	27%	0.07	0.68
A21	78%	0.33	0.70	63%	0.24	0.71	77%	0.32	0.67	85%	0.23	0.65
A22	33%	0.18	0.71	59%	0.30	0.71	49%	0.24	0.68	71%	0.01	0.69
A23_B30	52%	0.07	0.73	82%	0.45	0.69	91%	0.44	0.66	90%	0.40	0.63
A24_C26	51%	0.54	0.67	65%	0.31	0.71	60%	0.28	0.68	88%	0.45	0.63
A25	80%	0.38	0.69	86%	0.31	0.71	91%	0.20	0.68	90%	0.41	0.63
A26	53%	0.20	0.71	44%	0.22	0.71	79%	0.38	0.66	55%	0.06	0.69
A27	27%	0.41	0.69	52%	0.25	0.71	59%	0.41	0.66	63%	0.23	0.66
A28	46%	0.06	0.73	64%	0.42	0.69	85%	0.34	0.67	98%	0.27	0.66
A29	43%	0.48	0.68	76%	0.31	0.71	78%	0.23	0.68	77%	0.38	0.63
A30_B31	47%	0.36	0.69	48%	0.28	0.71	70%	0.34	0.67	81%	0.29	0.64
A31_B32	53%	0.50	0.68	75%	0.41	0.69	83%	0.27	0.68	95%	0.36	0.64
A32	42%	0.08	0.72	66%	0.32	0.70	61%	0.37	0.66	87%	0.36	0.64
A33_B26	52%	0.17	0.71	75%	0.28	0.71	68%	0.15	0.69	81%	0.27	0.65
A34_B28_C37	48%	0.53	0.67	88%	0.31	0.71	86%	0.25	0.68	94%	0.50	0.63
A35_B33_C35	79%	0.33	0.70	78%	0.31	0.71	81%	0.31	0.67	90%	0.26	0.65

Appendix B5

Form B Interpersonal Skills Standard Item Statistics by Country

Interpersonal Skills Form B	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
B18	59.8%	0.24	0.81	74.1%	0.33	0.74	72.4%	0.15	0.48	95.8%	0.09	0.54
B19	60.9%	0.69	0.78	72.4%	0.47	0.72	72.9%	0.12	0.49	73.7%	0.20	0.53
B20	32.2%	0.44	0.80	35.6%	0.16	0.75	57.3%	0.24	0.46	48.4%	0.16	0.54
B21	59.8%	0.40	0.80	53.4%	0.30	0.74	77.9%	0.15	0.48	71.1%	0.23	0.52
B22	69.5%	0.47	0.79	85.6%	0.35	0.73	95.0%	0.31	0.46	92.1%	0.27	0.52
B23	50.6%	0.70	0.78	58.6%	0.39	0.73	84.4%	0.10	0.49	96.8%	0.09	0.54
B24	71.3%	0.67	0.78	94.3%	0.21	0.74	93.5%	0.17	0.48	96.8%	0.39	0.51
B25	70.1%	0.69	0.78	88.5%	0.29	0.74	92.0%	0.29	0.46	98.4%	0.06	0.55
A33_B26	53.4%	0.11	0.82	69.5%	0.41	0.73	71.4%	0.15	0.48	81.6%	0.20	0.53
B27	62.1%	0.45	0.79	64.9%	0.27	0.74	65.3%	0.07	0.50	68.9%	0.28	0.51
A34_B28_C37	53.4%	0.64	0.78	78.7%	0.23	0.74	80.4%	0.20	0.47	95.3%	0.46	0.50
B29	45.4%	-0.06	0.83	77.0%	0.46	0.72	63.3%	0.09	0.50	87.4%	0.08	0.55
A23_B30	54.0%	0.21	0.81	82.2%	0.39	0.73	94.5%	0.20	0.48	92.1%	0.27	0.52
A30_B31	34.5%	0.37	0.80	44.8%	0.23	0.75	73.4%	0.20	0.47	83.7%	0.30	0.51
A31_B32	72.4%	0.40	0.80	79.3%	0.46	0.72	86.9%	0.26	0.46	97.4%	0.18	0.54
A35_B33_C35	70.1%	0.58	0.79	72.4%	0.40	0.73	82.9%	0.21	0.47	88.4%	0.20	0.53
B34_c21	51.7%	-0.05	0.83	76.4%	0.34	0.73	40.7%	0.09	0.50	58.9%	0.06	0.57

Appendix B6

Form C Interpersonal Skills Standard Item Statistics by Country

Interpersonal Skills Form C	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
C20	65.9%	0.59	0.82	72.0%	0.33	0.72	72.3%	0.25	0.64	76.2%	0.25	0.52
B34_C21	61.9%	0.16	0.84	74.2%	0.32	0.72	41.6%	0.15	0.66	47.1%	0.08	0.56
C22	60.8%	0.37	0.83	63.7%	0.36	0.72	75.7%	0.35	0.63	57.1%	0.24	0.52
C23	53.4%	0.63	0.82	76.9%	0.39	0.72	90.8%	0.29	0.64	91.5%	0.23	0.53
C24	68.8%	0.29	0.84	81.9%	0.36	0.72	88.4%	0.30	0.64	67.2%	-0.08	0.59
C25	76.1%	0.25	0.84	47.8%	0.36	0.72	90.8%	0.32	0.64	67.2%	0.18	0.53
A24_C26	60.2%	0.65	0.82	66.5%	0.24	0.73	52.6%	0.23	0.65	94.2%	0.30	0.52
C27	50.6%	0.44	0.83	52.7%	0.22	0.73	85.0%	0.27	0.64	43.9%	0.20	0.53
C28	59.1%	0.53	0.83	83.0%	0.35	0.72	86.1%	0.34	0.64	79.9%	0.26	0.52
C29	57.4%	0.60	0.82	68.7%	0.46	0.71	86.1%	0.37	0.63	96.3%	0.38	0.52
C30	56.8%	0.63	0.82	55.5%	0.26	0.73	63.6%	0.30	0.64	89.9%	0.37	0.51
C31	58.5%	0.68	0.82	62.6%	0.09	0.74	67.6%	0.17	0.66	67.7%	0.24	0.52
C32	63.1%	0.49	0.83	84.1%	0.26	0.73	86.7%	0.19	0.65	89.4%	0.38	0.51
C33	44.3%	0.14	0.84	73.6%	0.31	0.72	87.9%	0.20	0.65	89.9%	0.06	0.55
C34	67.0%	0.59	0.82	53.8%	0.28	0.72	72.3%	0.32	0.64	93.7%	0.21	0.53
A35_B33_C35	79.5%	0.40	0.83	84.6%	0.31	0.72	82.1%	0.18	0.65	87.8%	0.12	0.54
A20_C36	17.6%	0.03	0.85	31.3%	0.26	0.73	55.5%	0.26	0.64	30.2%	-0.05	0.58
A34_B28_C37	57.4%	0.58	0.82	78.6%	0.31	0.72	78.6%	0.24	0.64	95.2%	0.38	0.52
C38	34.7%	-0.03	0.85	80.8%	0.36	0.72	65.9%	0.08	0.67	81.0%	0.21	0.53

Appendix B7

Form A Dependability Standard Item Statistics by Country

Dependability Form A	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
A36	64%	0.49	0.83	76%	0.42	0.78	95%	0.51	0.73	96%	0.42	0.73
A37_B36_C53	42%	0.11	0.84	53%	0.34	0.79	62%	0.29	0.74	61%	0.20	0.74
A38	76%	0.37	0.83	79%	0.36	0.79	84%	0.24	0.74	65%	0.12	0.75
A39	36%	0.44	0.83	73%	0.43	0.78	77%	0.27	0.74	83%	0.27	0.73
A40_C54	45%	0.23	0.84	40%	0.33	0.79	60%	0.40	0.73	66%	0.30	0.73
A41	30%	0.50	0.83	62%	0.53	0.77	57%	0.36	0.73	28%	0.12	0.74
A42_B44_C44	64%	0.25	0.84	73%	0.53	0.78	87%	0.30	0.74	87%	0.39	0.72
A43	37%	0.24	0.84	55%	0.28	0.79	69%	0.36	0.73	75%	0.26	0.73
A44	43%	0.47	0.83	50%	0.26	0.79	89%	0.37	0.73	59%	0.34	0.72
A45	68%	0.30	0.84	72%	0.24	0.79	70%	0.28	0.74	86%	0.43	0.72
A46	51%	0.55	0.82	59%	0.29	0.79	56%	0.26	0.74	65%	0.28	0.73
A47	45%	0.69	0.82	73%	0.47	0.78	83%	0.41	0.73	91%	0.44	0.72
A48	55%	0.50	0.83	72%	0.30	0.79	75%	0.18	0.75	67%	0.23	0.73
A49	19%	0.03	0.84	56%	0.35	0.79	73%	0.26	0.74	73%	0.37	0.72
A50_B55_C55	57%	0.43	0.83	76%	0.43	0.78	79%	0.20	0.74	84%	0.42	0.72
A51	54%	0.73	0.82	81%	0.38	0.78	91%	0.40	0.73	88%	0.46	0.72
A52	57%	0.57	0.82	80%	0.47	0.78	60%	0.34	0.73	93%	0.50	0.72
A53	56%	0.10	0.85	60%	0.33	0.79	85%	0.42	0.73	83%	0.26	0.73
A54_B37	53%	0.71	0.82	86%	0.33	0.79	79%	0.24	0.74	93%	0.34	0.73
A55_B54_C51	39%	0.56	0.82	41%	0.20	0.80	80%	0.32	0.73	63%	0.34	0.72

Appendix B8

Form B Dependability Standard Item Statistics by Country

Dependability Form B	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
B35	74%	0.50	0.82	84%	0.50	0.79	95%	0.15	0.66	93%	0.32	0.77
A37_B36_C53	42%	0.10	0.84	41%	0.32	0.80	51%	0.16	0.67	52%	0.27	0.78
A54_B37	65%	0.66	0.82	87%	0.37	0.80	83%	0.14	0.67	91%	0.12	0.78
B38	68%	0.62	0.82	79%	0.43	0.79	96%	0.31	0.66	93%	0.46	0.76
B39	60%	0.71	0.81	74%	0.43	0.79	92%	0.25	0.66	97%	0.38	0.77
B40	63%	0.67	0.82	68%	0.50	0.79	89%	0.33	0.65	90%	0.45	0.76
B41	30%	0.50	0.82	47%	0.29	0.80	62%	0.32	0.65	71%	0.21	0.78
B42	60%	0.42	0.83	85%	0.42	0.80	91%	0.27	0.65	93%	0.36	0.77
B43	78%	0.34	0.83	78%	0.45	0.79	83%	0.29	0.65	94%	0.49	0.76
A42_B44_C44	61%	0.35	0.83	68%	0.48	0.79	87%	0.28	0.65	92%	0.36	0.77
B45	69%	0.21	0.84	66%	0.54	0.79	80%	0.38	0.64	82%	0.46	0.76
B46	38%	0.19	0.84	39%	0.14	0.81	65%	0.34	0.64	73%	0.42	0.76
B47	73%	0.54	0.82	76%	0.45	0.79	95%	0.10	0.67	91%	0.26	0.77
B48	47%	0.53	0.82	71%	0.39	0.80	84%	0.31	0.65	85%	0.47	0.76
A52_B49	13%	-0.26	0.85	10%	-0.32	0.82	29%	-0.20	0.71	1%	0.04	0.78
B50	48%	0.22	0.84	64%	0.43	0.79	61%	0.35	0.64	83%	0.35	0.77
B51	49%	0.54	0.82	66%	0.52	0.79	85%	0.26	0.65	77%	0.37	0.77
B52	57%	0.23	0.84	59%	0.20	0.81	78%	0.26	0.65	96%	0.32	0.77
B53	61%	0.29	0.83	47%	0.29	0.80	73%	0.42	0.63	63%	0.35	0.77
A55_B54_C51	46%	0.36	0.83	32%	0.26	0.80	69%	0.25	0.65	45%	0.35	0.77
A50_B55_C55	55%	0.67	0.82	68%	0.49	0.79	86%	0.23	0.66	92%	0.38	0.77

Appendix B9

Form C Dependability Standard Item Statistics by Country

Dependability Form C	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
C39	57%	0.55	0.73	76%	0.36	0.69	90%	0.29	0.61	59%	0.15	0.65
C40	59%	0.53	0.73	51%	0.34	0.69	79%	0.28	0.61	93%	0.28	0.64
C41	68%	0.33	0.75	75%	0.32	0.70	60%	0.17	0.63	68%	0.26	0.63
C42	75%	0.34	0.75	74%	0.42	0.69	85%	0.13	0.63	63%	0.04	0.67
C43	65%	0.49	0.73	58%	0.39	0.69	85%	0.36	0.60	82%	0.34	0.63
A42_B44_C44	69%	0.25	0.75	68%	0.45	0.68	85%	0.16	0.63	89%	0.44	0.62
C45	62%	0.61	0.72	48%	0.45	0.68	87%	0.26	0.61	96%	0.42	0.63
C46	41%	0.17	0.76	57%	0.17	0.71	80%	0.22	0.62	89%	0.44	0.62
C47	43%	0.61	0.72	49%	0.18	0.71	72%	0.22	0.62	83%	0.56	0.60
C48	43%	0.37	0.74	53%	0.28	0.70	79%	0.20	0.62	52%	0.28	0.63
C49	41%	0.50	0.73	55%	0.43	0.69	76%	0.23	0.62	44%	0.12	0.66
C50	35%	0.20	0.76	61%	0.25	0.70	80%	0.27	0.61	90%	0.29	0.63
A55_B54_C51	42%	0.52	0.73	40%	0.13	0.72	73%	0.29	0.61	41%	0.21	0.64
C52	64%	0.33	0.75	66%	0.35	0.69	76%	0.38	0.60	86%	0.38	0.62
A37_B36_C53	47%	-0.01	0.77	53%	0.33	0.70	51%	0.35	0.60	56%	0.27	0.63
A40_C54	35%	-0.07	0.78	36%	0.09	0.72	49%	0.26	0.61	60%	0.14	0.65
A50_B55_C55	56%	0.62	0.72	72%	0.49	0.68	87%	0.33	0.61	82%	0.51	0.60
C56	19%	-0.32	0.79	10%	-0.30	0.74	18%	-0.17	0.67	10%	-0.31	0.69

Appendix B10

Form A Problem Solving/Critical Thinking Standard Item Statistics by Country

Problem Solving Form A	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
A56	58%	0.23	0.60	71%	0.32	0.70	75%	0.35	0.64	85%	0.38	0.65
A57_B56_C57	65%	0.43	0.56	88%	0.39	0.70	95%	0.42	0.65	94%	0.44	0.65
A58	74%	0.25	0.60	71%	0.41	0.70	77%	0.34	0.64	74%	0.44	0.64
A59	43%	0.11	0.62	37%	0.18	0.72	50%	0.26	0.65	66%	0.09	0.69
A60	66%	0.25	0.60	71%	0.40	0.70	75%	0.12	0.67	80%	0.31	0.66
A61	31%	0.41	0.57	56%	0.41	0.69	63%	0.23	0.66	68%	0.26	0.67
A62	27%	0.34	0.58	68%	0.41	0.69	82%	0.35	0.64	83%	0.53	0.63
A63_B60_C64	31%	0.36	0.58	55%	0.38	0.70	59%	0.38	0.63	60%	0.14	0.68
A64_B61_C65	50%	0.40	0.57	60%	0.30	0.71	66%	0.37	0.64	66%	0.25	0.67
A65_B62_C66	43%	-0.20	0.67	61%	0.37	0.70	74%	0.33	0.64	75%	0.16	0.68
A66_B63_C67	42%	0.26	0.59	53%	0.22	0.72	48%	0.13	0.67	87%	0.18	0.67
A67_B64_C68	44%	0.13	0.62	64%	0.15	0.72	71%	0.14	0.67	80%	0.28	0.66
A68_C69	22%	0.12	0.61	22%	0.33	0.70	39%	0.35	0.64	29%	0.19	0.68
A69_B66_C70	43%	0.49	0.55	54%	0.39	0.70	84%	0.34	0.64	81%	0.52	0.63
A70_B67_C71	54%	0.11	0.62	49%	0.22	0.72	84%	0.20	0.66	79%	0.35	0.65

Appendix B11

Form B Problem Solving/Critical Thinking Standard Item Statistics by Country

Problem Solving/Critical Thinking Form B	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
A57_B56_C57	70%	0.52	0.68	89%	0.37	0.66	93%	0.33	0.48	97%	0.42	0.70
B57	47%	0.57	0.67	61%	0.34	0.66	53%	-0.07	0.55	91%	0.47	0.69
B58	63%	0.29	0.71	55%	0.32	0.66	79%	0.22	0.48	90%	0.39	0.69
B59	63%	0.24	0.71	60%	0.32	0.66	72%	0.22	0.48	73%	0.44	0.68
A63_B60_C64	39%	0.51	0.68	56%	0.34	0.66	62%	0.33	0.45	74%	0.07	0.73
A64_B61_C65	56%	0.50	0.68	69%	0.35	0.66	75%	0.21	0.48	77%	0.15	0.72
A65_B62_C66	41%	-0.15	0.75	66%	0.42	0.65	72%	0.25	0.47	75%	0.30	0.70
A66_B63_C67	43%	0.36	0.70	51%	0.20	0.68	44%	0.14	0.50	88%	0.30	0.70
A67_B64_C68	43%	0.37	0.70	51%	0.18	0.68	75%	0.22	0.48	79%	0.32	0.70
B65	32%	0.16	0.72	21%	0.27	0.67	39%	0.21	0.48	34%	0.28	0.70
A69_B66_C70	56%	0.66	0.66	49%	0.38	0.65	78%	0.20	0.49	92%	0.42	0.69
A70_B67_C71	57%	-0.07	0.74	45%	0.24	0.67	80%	0.14	0.50	81%	0.46	0.68
B68	46%	-0.05	0.74	30%	-0.09	0.71	51%	0.05	0.52	51%	0.34	0.70
B69	59%	0.55	0.68	69%	0.41	0.65	90%	0.27	0.48	89%	0.35	0.70
B70	66%	0.47	0.69	68%	0.30	0.66	81%	0.12	0.50	78%	0.35	0.69

Appendix B12

Form C Problem Solving/Critical Thinking Standard Item Statistics by Country

Problem Solving Form C	Tanzania			Senegal			Uganda			Rwanda		
	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD	Pct	Disc.	AifD
A57_B56_C57	67%	.550	.755	91%	.305	.680	89%	.391	.572	93%	.524	0.696
C58	61%	.693	.742	73%	.421	.663	84%	.371	.571	92%	.446	0.665
C59	39%	.393	.768	32%	.127	.699	20%	-.069	.633	44%	.191	0.691
C60	57%	.338	.772	54%	.302	.678	68%	.285	.580	88%	.471	0.654
C61	59%	.560	.753	59%	.291	.679	71%	.200	.595	74%	.414	0.655
C62	56%	.436	.764	66%	.406	.664	66%	.131	.607	69%	.351	0.665
C63	64%	.651	.746	60%	.461	.656	83%	.307	.579	87%	.427	0.658
A63_B60_C64	34%	.423	.765	47%	.394	.665	55%	.314	.574	61%	.162	0.694
A64_B61_C65	46%	.437	.764	69%	.284	.680	64%	.303	.576	69%	.162	0.695
A65_B62_C66	36%	-.287	.819	73%	.249	.684	69%	.097	.612	74%	.232	0.681
A66_B63_C67	44%	.320	.774	56%	.178	.694	41%	.128	.609	88%	.337	0.671
A67_B64_C68	49%	.391	.768	63%	.229	.687	65%	.275	.581	74%	.280	0.674
A68_C69	25%	.192	.783	23%	.163	.693	36%	.280	.580	29%	.298	0.669
A69_B66_C70	54%	.721	.738	59%	.361	.670	75%	.371	.566	87%	.369	0.666
A70_B67_C71	57%	.054	.795	38%	.333	.674	72%	.238	.588	79%	.546	0.638

Appendix C1

Logistic Regression DIF Results for Gender: Communication.

Item	Uniform and NonUniform Gender DIF		Uniform Gender DIF		Nagelkerke R-Squared for Significant results	
	X^2	p-value	X^2	p-value	Attributed to Uniform and NonUniform Gender DIF	Attributed to Uniform Gender DIF
A1	0.637	0.727	0.21	0.646	-	-
A2 B3 C1	5.743	0.057	0.752	0.386	-	-
A3 B5	2.265	0.322	0.038	0.845	-	-
A4	2.052	0.358	1.518	0.218	-	-
A5 B16 C17	2.159	0.34	1.803	0.179	-	-
A6 B7 C9	0.417	0.812	0.004	0.952	-	-
A7	2.023	0.364	1.178	0.278	-	-
A8 C12	0.472	0.79	0.078	0.781	-	-
A9 B10	1.457	0.483	0.055	0.815	-	-
A10 B11	1.342	0.511	1.27	0.26	-	-
A11	0.09	0.956	0.075	0.784	-	-
A12 B13 C18	16.089	<.0001	15.76	<.000	.01	.009
A13 B12 C11	1.394	0.498	0.332	0.564	-	-
A14 B15	0.905	0.636	0.284	0.594	-	-
A15 B14 C16	8.3	0.016	7.162	0.007	-	.005
A16 B1	2.898	0.235	1.104	0.293	-	-
A17 B2 C19	2.589	0.274	1.038	0.308	-	-
A18 B17 C5	2.597	0.273	0.002	0.966	-	-
B4	5.538	0.063	3.071	0.08	-	-
B6	0.609	0.737	0.504	0.478	-	-
B8	6.526	0.038	3.995	0.046	-	-
B9	0.663	0.718	0.65	0.42	-	-
C2	3.746	0.154	3.184	0.074	-	-
C3	8.209	0.17	7.639	0.006	-	.015
C4	2.158	0.34	2.154	0.142	-	-
C6	2.067	0.356	0.115	0.734	-	-
C7	0.728	0.695	0.264	0.607	-	-
C8	1.598	0.45	1.594	0.207	-	-
C10	4.333	0.115	0.828	0.363	-	-
C13	1.991	0.369	0.9	0.343	-	-
C14	1.087	0.581	0.288	0.591	-	-

| C15 | 7.392 | 0.025 | 5.25 | 0.022 | - | - |

Appendix C2

Logistic Regression DIF Results for Gender: Interpersonal Skills.

| Item | Uniform and NonUniform Gender DIF | | Uniform and NonUniform White/Non-White DIF | | Nagelkerke R-Squared for Significant results | |
	X^2	p-value	X^2	p-value	Attributed to Uniform and NonUniform Gender DIF	Attributed to Uniform Gender DIF
A19	0.398	0.82	0.222	0.638	-	-
A20_C36	1.992	0.369	1.871	0.171	-	-
A21	2.586	0.274	1.102	0.314	-	-
A22	1.349	0.509	1.162	0.281	-	-
A23_B30	2.698	0.259	2.233	0.135	-	-
A24_C26	1.372	0.504	0.108	0.742	-	-
A25	3.142	0.208	1.804	0.179	-	-
A26	7.132	0.028	3.247	0.072	-	-
A27	0.306	0.858	0.218	0.641	-	-
A28	1.637	0.441	1.637	0.201	-	-
A29	0.203	0.904	0.1	0.752	-	-
A30_B31	2.298	0.317	1.425	0.233	-	-
A31_B32	0.216	0.898	0.166	0.683	-	-
A32	3.926	0.14	0.607	0.436	-	-
A33_B26	1.319	0.517	1.318	0.251	-	-
A34_B28_C37	0.542	0.763	0.319	0.572	-	-
A35_B33_C35	2.72	0.257	0.703	0.402	-	-
B18	2.58	0.275	2.291	0.13	-	-
B19	0.747	0.688	0.742	0.389	-	-
B20	4.05	0.132	0.265	0.607	-	-
B21	1.942	0.379	1.265	0.261	-	-
B22	1.023	0.6	0.234	0.629	-	-
B23	2.773	0.25	1.708	0.191	-	-
B24	0.077	0.962	0.062	0.803	-	-
B25	0.058	0.971	0.01	0.922	-	-
B27	3.093	0.213	3.075	0.08	-	-
B29	2.968	0.227	0.065	0.799	-	-
B34_C21	3.71	0.156	3.593	0.058	-	-
C20	1.888	0.389	0.573	0.449	-	-
C22	0.397	0.82	0.215	0.643	-	-

C23	2.342	0.31	0.327	0.567	-	-
C24	2.124	0.346	0.527	0.468	-	-
C25	10.745	0.005	0.136	0.713	.019	-
C27	5.149	0.076	2.581	0.108	-	-
C28	1.509	0.47	1.284	0.257	-	-
C29	3.184	0.203	2.746	0.097	-	-
C30	7.143	0.028	6.02	0.014	-	-
C31	5.276	0.072	5.204	0.023	-	-
C32	3.543	0.17	3.373	0.066	-	-
C33	0.547	0.75	0.013	0.91	-	-
C34	1.028	0.598	0.51	0.475	-	-
C38	7.716	0.021	1.813	0.178	-	-

Appendix C3

Logistic Regression DIF Results for Gender: Dependability

Item	Uniform and NonUniform Gender DIF		Uniform Gender DIF		Nagelkerke R-Squared for Significant results	
	X^2	p-value	X^2	p-value	Attributed to Uniform and NonUniform Gender DIF	Attributed to Uniform Gender DIF
A36	4.426	0.109	1.368	0.242	-	-
A37_B36_C53	1.756	0.416	0.444	0.505	-	-
A38	0.603	0.74	0.079	0.778	-	-
A39	0.823	0.663	0.474	0.491	-	-
A40_C54	3.996	0.136	0.001	0.975	-	-
A41	7.353	0.025	1.597	0.206	-	-
A42_B44_C44	0.109	0.947	0.019	0.889	-	-
A43	0.143	0.931	0.129	0.719	-	-
A44	7.965	0.019	7.942	0.005	-	0.01
A45	1.177	0.555	0.325	0.568	-	-
A46	0.603	0.74	0.602	0.438	-	-
A47	1.315	0.518	0.103	0.748	-	-
A48	8.766	0.012	8.717	0.003	-	0.12
A49	0.7	0.705	0.282	0.595	-	-
A50_B55_C55	1.558	0.459	0.029	0.866	-	-
A51	8.441	0.015	8.382	0.004	-	0.01
A52_B49	8.445	0.015	6.128	0.013	-	-
A53	4.006	0.135	2.541	0.111	-	-
A54_B37	6.733	0.035	3.647	0.056	-	-
A55_B54_C51	9.366	0.009	9.225	0.002	.005	0.005
B35	23.13	<.000	0	0.992	.045	-
B38	0.206	0.902	0.082	0.775	-	-
B39	5.573	0.062	0.125	0.724	-	-
B40	3.727	0.155	3.692	0.055	-	-
B41	6.103	0.047	1.223	0.269	-	-
B42	0.184	0.912	0.011	0.915	-	-
B43	0.417	0.812	0.221	0.638	-	-
B45	0.801	0.67	0.012	0.914	-	-
B46	1.865	0.394	1.65	0.199	-	-
B47	1.919	0.383	0.002	0.969	-	-
B48	2.562	0.278	1.131	0.31	-	-
B50	0.041	0.98	0.04	0.841	-	-

B51	2.114	0.347	1.996	0.158	-	-
B52	1.247	0.536	1.023	0.312	-	-
B53	5.924	0.052	5.91	0.015	-	-
C39	3.501	0.174	1.851	0.174	-	-
C40	0.438	0.803	0.423	0.516	-	-
C41	2.163	0.339	1.675	0.196	-	-
C42	0.185	0.912	0.076	0.783	-	-
C43	0.907	0.635	0.147	0.702	-	-
C45	14.773	0.001	12.08	0.001	.02	0.017
C46	0.141	0.932	0.14	0.708	-	-
C47	0.217	0.897	0.149	0.7	-	-
C48	3.915	0.141	2.912	0.088	-	-
C49	3.494	0.174	0.087	0.768	-	-
C50	1.171	0.557	0.205	0.651	-	-
C52	3.077	0.215	0.576	0.448	-	-
C56	8.303	0.016	1.364	0.243	-	-

Appendix C4

Logistic Regression DIF Results for Gender: Problem Solving/Critical Thinking

| Item | Uniform and NonUniform Gender DIF | | Uniform Gender DIF | | Nagelkerke R-Squared for Significant results | |
	X^2	p-value	X^2	p-value	Attributed to Uniform and NonUniform Gender DIF	Attributed to Uniform Gender DIF
A56	1.657	0.437	1.545	0.214	-	-
A57_B56_C57	1.188	0.552	0.97	0.325	-	-
A58	0.131	0.937	0.114	0.736	-	-
A59	0.587	0.745	0.107	0.743	-	-
A60	1.023	0.6	0.122	0.726	-	-
A61	10.581	0.005	5.238	0.022	.013	-
A62	2.115	0.347	1.117	0.29	-	-
A63_B60_C64	1.579	0.454	1.499	0.221	-	-
A64_B61_C65	1.116	0.572	0.495	0.482	-	-
A65_B62_C66	15.515	<.000	11.97	0.001	0	-
A66_B63_C67	0.338	0.845	0.061	0.805	-	-
A67_B64_C68	10.129	0.006	8.441	0.004	.005	.005
A68_C69	4.274	0.118	4.274	0.039	-	-
A69_B66_C70	4.04	0.133	2.756	0.097	-	-
A70_B67_C71.	3.222	0.2	2.63	0.105	-	-
B57	4.087	0.13	1.21	0.271	-	-
B58	5.357	0.069	2.759	0.097	-	-
B59	0.884	0.643	0.579	0.447	-	-
B65	5.42	0.067	5.42	0.02	-	-
B68	3.61	0.164	0.005	0.943	-	-
B69	2.249	0.325	1.207	0.272	-	-
B70	1.593	0.451	0.09	0.764	-	-
C58	2.623	0.269	2.619	0.106	-	-
C59	0.118	0.943	0.117	0.733	-	-
C60	2.754	0.252	0.559	0.455	-	-
C61	0.182	0.913	0.181	0.671	-	-
C62	1.655	0.437	0.211	0.646	-	-
C63	13.818	0.001	10.82	0.001	.019	.015

Appendix D1

Communications Form B and C Score Conversion to Form A

Raw Score on Form B	Converted to Form A	Raw Score on form C	Converted to Form A
0	0	0	0
1	0	1	0
2	1	2	1
3	2	3	2
4	3	4	2
5	4	5	3
6	5	6	4
7	6	7	5
8	7	8	6
9	9	9	7
10	10	10	7
11	11	11	8
12	12	12	9
13	13	13	10
14	14	14	11
15	15	15	12
16	16	16	14
17	17	17	15
18	Not applicable	18	16
19	Not applicable	19	17

Appendix D2

Interpersonal Skills Form B and C Score Conversion to Form A

Raw Score on Form B	Converted to Form A	Raw Score on form C	Converted to Form A
0	0	0	0
1	1	1	1
2	1	2	1
3	2	3	2
4	3	4	3
5	4	5	4
6	5	6	5
7	6	7	6
8	6	8	7
9	7	9	8
10	8	10	8
11	10	11	9
12	11	12	10
13	12	13	11
14	13	14	12
15	14	15	13
16	15	16	14
17	16	17	15
18	Not applicable	18	16
19	Not applicable	19	17

Appendix D3

Dependability Form B and C Score Conversion to Form A

Raw Score on Form B	Converted to Form A	Raw Score on form C	Converted to Form A
0	0	0	0
1	0	1	1
2	0	2	2
3	1	3	3
4	1	4	4
5	2	5	5
6	2	6	6
7	3	7	7
8	4	8	8
9	4	9	9
10	5	10	10
11	6	11	11
12	7	12	12
13	8	13	13
14	9	14	15
15	10	15	16
16	12	16	17
17	13	17	18
18	14	18	20

Appendix D4

Problem Solving/Critical Thinking Form B and C Score Conversion to Form A

Raw Score on Form B	Converted to Form A	Raw Score on form C	Converted to Form A
0	0	0	0
1	1	1	0
2	2	2	1
3	3	3	2
4	4	4	3
5	5	5	4
6	6	6	5
7	7	7	6
8	8	8	7
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15

Appendix E1

Gender Comparisons by Country: Rwanda

Rwanda		Females			Males			
		N	Mean		N	Mean		Significance
Form	Standard	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	
A	Communications	129	12.853	0.249	97	13.701	0.252	0.020
	Interpersonal Skills	129	13.411	0.211	97	13.835	0.252	0.196
	Dependability	129	14.729	0.282	97	15.515	0.340	0.074
	Problem Solving/Critical Thinking	129	10.860	0.230	97	11.361	0.271	0.160
B	Communications	103	12.689	0.233	87	13.126	0.232	0.189
	Interpersonal Skills	103	12.990	0.210	87	13.517	0.216	0.083
	Dependability	103	12.350	0.371	87	12.989	0.414	0.250
	Problem Solving/Critical Thinking	103	11.476	0.257	87	11.977	0.267	0.180
C	Communications	99	12.606	0.282	90	12.800	0.246	0.608
	Interpersonal Skills	99	12.343	0.241	90	12.656	0.229	0.351
	Dependability	99	12.970	0.314	90	12.722	0.333	0.589
	Problem Solving/Critical Thinking	99	11.172	0.309	90	10.589	0.321	0.193

Appendix E2

Gender Comparisons by Country: Senegal

Senegal		Females			Males			
		N	Mean		N	Mean		Significance
Form	Standard	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	
A	Communications	99	10.616	0.324	78	11.974	0.368	0.006
	Interpersonal Skills	99	10.697	0.357	78	11.756	0.320	0.033
	Dependability	99	12.283	0.433	78	14.295	0.407	0.001
	Problem Solving/Critical Thinking	99	8.081	0.324	78	9.692	0.326	0.001
B	Communications	94	8.883	0.340	80	11.150	0.379	0.000
	Interpersonal Skills	94	9.777	0.361	80	12.225	0.316	0.000
	Dependability	94	7.181	0.434	80	10.638	0.424	0.000
	Problem Solving/Critical Thinking	94	7.745	0.311	80	9.188	0.318	0.001
C	Communications	116	9.974	0.335	47	10.064	0.412	0.879
	Interpersonal Skills	116	11.302	0.314	47	10.340	0.417	0.088
	Dependability	116	10.121	0.364	47	9.681	0.486	0.500
	Problem Solving/Critical Thinking	116	8.371	0.346	47	7.383	0.394	0.101

Appendix E3

Gender Comparisons by Country: Tanzania

Tanzania		Females			Males			
		N	Mean		N	Mean		Significance
Form	Standard	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	
A	Communications	114	9.746	0.389	83	9.506	0.413	0.678
	Interpersonal Skills	114	8.246	0.305	83	8.651	0.390	0.407
	Dependability	114	10.018	0.450	83	9.771	0.522	0.722
	Problem Solving/Critical Thinking	114	6.982	0.270	83	6.916	0.304	0.871
B	Communications	78	8.000	0.540	96	9.146	0.451	0.102
	Interpersonal Skills	78	7.949	0.448	96	9.094	0.424	0.066
	Dependability	78	7.038	0.515	96	7.708	0.498	0.355
	Problem Solving/Critical Thinking	78	7.821	0.360	96	7.781	0.347	0.938
C	Communications	89	7.607	0.491	86	11.547	0.338	0.000
	Interpersonal Skills	89	7.831	0.458	86	10.977	0.333	0.000
	Dependability	89	8.135	0.441	86	10.826	0.365	0.000
	Problem Solving/Critical Thinking	89	5.539	0.441	86	8.477	0.354	0.000

Appendix E4

Gender Comparisons by Country: Uganda

Uganda		Females			Males			
		N	Mean		N	Mean		Significance
Form	Standard	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	
A	Communications	111	12.135	0.285	72	11.625	0.302	0.237
	Interpersonal Skills	111	12.694	0.285	72	12.389	0.334	0.494
	Dependability	111	15.279	0.289	72	14.861	0.469	0.423
	Problem Solving/Critical Thinking	111	10.459	0.257	72	10.389	0.334	0.666
B	Communications	119	11.555	0.234	80	11.575	0.273	0.955
	Interpersonal Skills	119	11.992	0.222	80	11.825	0.279	0.638
	Dependability	119	12.445	0.335	80	12.425	0.421	0.970
	Problem Solving/Critical Thinking	119	10.454	0.206	80	9.788	0.256	0.043
C	Communications	98	11.786	0.332	75	11.987	0.306	0.666
	Interpersonal Skills	98	12.398	0.283	75	12.293	0.318	0.806
	Dependability	98	13.786	0.315	75	13.387	0.370	0.411
	Problem Solving/Critical Thinking	98	9.429	0.319	75	9.133	0.318	0.521

Appendix F1

Country Level Comparisons: Communication

Communication		Form A			Form B			Form C		
(I) country2	(J) country2	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.
Rwanda	Senegal	2.002	0.325	.000	2.730	0.316	.000	2.444	0.351	.000
	Tanzania	3.572	0.315	.000	3.891	0.316	.000	3.334	0.354	.000
	Uganda	1.282	0.322	.000	1.279	0.305	.000	0.764	0.356	.139
Senegal	Rwanda	-2.002	0.325	.000	-2.730	0.316	.000	-2.444	0.351	.000
	Tanzania	1.570	0.335	.000	1.161	0.322	.002	0.890	0.357	.062
	Uganda	-0.720	0.341	.150	-1.451	0.312	.000	-1.680	0.359	.000
Tanzania	Rwanda	-3.572	0.315	.000	-3.891	0.316	.000	-3.334	0.354	.000
	Senegal	-1.570	0.335	.000	-1.161	0.322	.002	-0.890	0.357	.062
	Uganda	-2.290	0.332	.000	-2.612	0.312	.000	-2.569	0.362	.000
Uganda	Rwanda	-1.282	0.322	.000	-1.279	0.305	.000	-0.764	0.356	.139
	Senegal	0.720	0.341	.150	1.451	0.312	.000	1.680	0.359	.000
	Tanzania	2.290	0.332	.000	2.612	0.312	.000	2.569	0.362	.000

Appendix F2

Country Level Comparisons: Interpersonal Skills

Interpersonal Skills		Form A			Form B			Form C		
(I) country2	(J) country2	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.
Rwanda	Senegal	2.429	0.302	.000	2.188	0.310	.000	1.532	0.356	.000
	Tanzania	5.177	0.293	.000	4.556	0.310	.000	3.523	0.359	.000
	Uganda	1.019	0.299	.004	1.228	0.300	.000	0.160	0.361	.971
Senegal	Rwanda	-2.429	0.302	.000	-2.188	0.310	.000	-1.532	0.356	.000
	Tanzania	2.748	0.311	.000	2.368	0.317	.000	1.991	0.363	.000
	Uganda	-1.410	0.317	.000	-0.960	0.307	.010	-1.372	0.364	.001
Tanzania	Rwanda	-5.177	0.293	.000	-4.556	0.310	.000	-3.523	0.359	.000
	Senegal	-2.748	0.311	.000	-2.368	0.317	.000	-1.991	0.363	.000
	Uganda	-4.158	0.309	.000	-3.328	0.307	.000	-3.363	0.367	.000
Uganda	Rwanda	-1.019	0.299	.004	-1.228	0.300	.000	-0.160	0.361	.971
	Senegal	1.410	0.317	.000	0.960	0.307	.010	1.372	0.364	.001
	Tanzania	4.158	0.309	.000	3.328	0.307	.000	3.363	0.367	.000

Appendix F3

Country Level Comparisons: Dependability

Dependability		Form A			Form B			Form C		
(I) country2	(J) country2	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.
Rwanda	Senegal	1.897	0.394	.000	3.403	0.398	.000	2.433	0.336	.000
	Tanzania	5.153	0.383	.000	4.903	0.398	.000	3.228	0.339	.000
	Uganda	-0.048	0.391	.999	0.113	0.385	.991	-0.661	0.340	.212
Senegal	Rwanda	-1.897	0.394	.000	-3.403	0.398	.000	-2.433	0.336	.000
	Tanzania	3.256	0.407	.000	1.500	0.407	.001	0.795	0.342	.093
	Uganda	-1.945	0.414	.000	-3.291	0.394	.000	-3.094	0.343	.000
Tanzania	Rwanda	-5.153	0.383	.000	-4.903	0.398	.000	-3.228	0.339	.000
	Senegal	-3.256	0.407	.000	-1.500	0.407	.001	-0.795	0.342	.093
	Uganda	-5.201	0.403	.000	-4.791	0.394	.000	-3.888	0.346	.000
Uganda	Rwanda	0.048	0.391	.999	-0.113	0.385	.991	0.661	0.340	.212
	Senegal	1.945	0.414	.000	3.291	0.394	.000	3.094	0.343	.000
	Tanzania	5.201	0.403	.000	4.791	0.394	.000	3.888	0.346	.000

Appendix F4

Country Level Comparisons: Problem Solving/Critical Thinking

Problem Solving/Critical Thinking		Form A			Form B			Form C		
(I) country2	(J) country2	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.	Mean Difference (I-J)	Std. Error	Sig.
Rwanda	Senegal	2.284	0.285	.000	3.297	0.293	.000	2.447	0.314	.000
	Tanzania	4.121	0.277	.000	3.906	0.293	.000	3.608	0.317	.000
	Uganda	0.644	0.283	.104	1.519	0.283	.000	1.490	0.318	.000
Senegal	Rwanda	-2.284	0.285	.000	-3.297	0.293	.000	-2.447	0.314	.000
	Tanzania	1.837	0.294	.000	0.609	0.300	.177	1.160	0.320	.002
	Uganda	-1.641	0.300	.000	-1.778	0.290	.000	-0.958	0.321	.016
Tanzania	Rwanda	-4.121	0.277	.000	-3.906	0.293	.000	-3.608	0.317	.000
	Senegal	-1.837	0.294	.000	-0.609	0.300	.177	-1.160	0.320	.002
	Uganda	-3.477	0.292	.000	-2.387	0.290	.000	-2.118	0.324	.000

Uganda	Rwanda	-0.644	0.283	.104	-1.519	0.283	.000	-1.490	0.318	.000
	Senegal	1.641	0.300	.000	1.778	0.290	.000	0.958	0.321	.016
	Tanzania	3.477	0.292	.000	2.387	0.290	.000	2.118	0.324	.000

Appendix G

Glossary of Analysis-Related Terms

Term	Definition
Adjusted Group Means	Adjusted Group Means refer to group means calculated based upon the equated scores rather than raw scores for each form.
Classical Item Analysis	Classical Item Analysis uses student responses to items to examine how well individual items function for the groups tested. The principle statistics examined are item difficulty, item discrimination, and the contribution of the item to scale reliability.
Differential Item Functioning	Differential Item Functioning (DIF) predicts item responses (correct or incorrect) as a function of group membership (gender), total score, and group by total interaction. This is used in the current analyses to determine if an item exhibits bias in favor of boys or girls when matched on total score.
Effect Size	An Effect Size of .13 is used in the DIF analyses to determine if statistically significant differences between groups for a given item represent meaningful differences that could impact the fairness of the assessments.
Equated Form Scores	Using the Rasch Model predicted scores on form A were computed for scores on forms B and C. This was done in the current analyses to facilitate comparisons across forms.
Form-Based Factor Analysis	Factor Analysis examines the interrelationship of items to determine how many scales or factors are represented by the data.
Item Difficulty	Item difficulty in the context of classical item analysis represents the percent of students answering an item correctly.
Item Discrimination	Item Discrimination in the context of classical Item analysis represent the relationship of responses to an item to the total scores for the scale. A high positive value is desirable, indicating that high achievers on the test are also high achievers on the given item.
Item Response Theory (Rasch Model)	The Rasch Model is an item-response-theory model that simultaneously estimates the ability of the subject and the difficulty of items. In the current analyses, it is used to equate test forms so that scores can be directly compared.
Link Items	Link items are common items used across forms.
NonUniform DIF	NonUniform DIF implies that boys and girls of the same ability have a different probability of answering a given item correctly and that this relationship changes depending on the level of ability level considered.
Test Reliability	Test Reliability is a measure of the internal consistency of items comprising a scale.
Uniform DIF	Uniform DIF implies that boys and girls of the same ability level have a different probability of answering a given item correctly.