Kepler Year Three Impact Evaluation
Endline Report
October 2016

Paul Wang: paul.wang@idinsight.org
Harlan Downs-Tepper: harlan.downs.tepper@idinsight.org
Nicole Hoesterey: nicole.hoesterey@idinsight.org
Table of Contents

Executive Summary ......................................................................................................................... 3
Introduction ........................................................................................................................................ 4
  About Kepler ................................................................................................................................. 4
  IDInsight’s role ............................................................................................................................... 4
Evaluation design .......................................................................................................................... 5
  Evaluation methodology ................................................................................................................. 5
    Background of comparison group selection process ................................................................. 5
    Impact measurement methods .................................................................................................. 5
  Metrics ........................................................................................................................................... 6
  Testing periods .............................................................................................................................. 7
Previous findings ............................................................................................................................ 8
  Year 1 evaluation results .............................................................................................................. 8
  Year 2 evaluation results .............................................................................................................. 8
Year 3 evaluation results ................................................................................................................. 9
  Interpreting effect sizes ................................................................................................................ 9
  Summary of results ....................................................................................................................... 9
Results by test ................................................................................................................................ 11
  Critical thinking .......................................................................................................................... 11
  Cognitive skills ............................................................................................................................. 11
  English language .......................................................................................................................... 11
  Computer literacy ......................................................................................................................... 11
Year 3 survey findings .................................................................................................................... 12
  Employment plans ....................................................................................................................... 12
  Resources used in the employment search ................................................................................. 12
  Sector ........................................................................................................................................... 13
  Terms of employment .................................................................................................................. 13
  Expected stay ............................................................................................................................... 14
Conclusion ..................................................................................................................................... 15
Appendices .................................................................................................................................... 16
Executive Summary

Kepler, an innovative university in Rwanda, seeks to provide high-quality, low-cost education for promising Rwandan youth. Kepler employs a blended learning model that combines low-cost online courses with in-person course facilitation. The academic material is tailored to provide the practical skills most demanded in the Rwandan and international job market: English language skills, computer literacy, professional competencies, and critical thinking.

IDinsight has collaborated with Kepler since the first cohort of students enrolled in 2013 to evaluate the impact of Kepler on student learning in critical thinking, cognitive skills, English language, and computer literacy compared to a matched cohort of students attending traditional Rwandan universities.

The Year 3 evaluation had three objectives:

1. Understand the cumulative three-year impact of Kepler for the 2013 cohort
2. Assess the discrete impact of the third year for the 2013 cohort
3. Learn about post-graduation employment plans for the 2013 cohort

IDinsight used a propensity score matching methodology in 2013 to identify a group of comparison students entering traditional Rwandan universities who were similar to the incoming cohort of Kepler students on key demographics. The same group of students has been tested each of the past three years.

This work builds on the evaluations conducted in previous years. At the end of the first year, Kepler students performed better than comparison students in cognitive skills, English, and computer literacy, but there was no statistically significant impact on critical thinking. The Year 2 evaluation found significant impact on all evaluation areas.

This report presents three main findings from the Year 3 evaluation:

- **Kepler students in the 2013 cohort continued to outperform comparison students** in all areas: critical thinking, cognitive skills, English language, and computer literacy. Kepler had the largest effect on critical thinking and computer literacy.

- **Kepler students maintained higher performance levels during the third year of Kepler, but did not make additional gains relative to comparison students.** Comparison students made gains relative to Kepler students on the English grammar section of the cognitive skills test. Differences between Kepler and comparison students remained unchanged for all other subjects.

- **Kepler students are much more likely to have secured full-time employment by the end of their third year of university than their peers.** The Kepler students who have secured employment are likely to work in business, education, or communications.
Introduction

Kepler, an innovative university in Rwanda, has engaged IDinsight for the past three years to conduct an impact evaluation of its model on student academic outcomes. The purpose of the evaluation was to assess the impact of Kepler relative to traditional Rwandan universities on students’ critical thinking, cognitive skills, and computer literacy. This report describes the context, research objectives, evaluation background, methodology, previous results, and findings from the Year 3 endline evaluation. Findings come from tests and a survey administered to the 2013 cohort of Kepler students and matched comparison students in April–May 2016.

About Kepler

Kepler aims to dramatically improve learning and employment outcomes while lowering the cost of higher education to make it more widely accessible. To achieve these goals, Kepler employs a blended learning model that combines low-cost online courses with in-person course facilitation. The academic material is tailored to provide the practical skills most demanded in the Rwandan and international job market: English language skills, computer literacy, professional competencies, and critical thinking.

During their first two years at Kepler, students work towards their associates degree through a core curriculum and online content. Third-year Kepler students pursue the bachelor’s degree through self-paced learning on online platforms and engage in the workforce through internships and jobs, while their peers at traditional Rwandan universities continue to learn in the classroom.

Kepler started its inaugural cohort of 50 students in 2013, and has grown to accept cohorts of 100–150 students in the following years.

IDinsight’s role

IDinsight has been engaged since the beginning of Kepler to evaluate Kepler’s impact on student learning in critical thinking, cognitive skills, English language, and computer literacy compared to a matched cohort of students attending traditional Rwandan universities.

The Year 3 evaluation had three objectives:

1. Understand the cumulative three-year impact of Kepler for the 2013 cohort
2. Assess the discrete impact of the third year for the 2013 cohort
3. Learn about post-graduation employment plans for the 2013 cohort
Evaluation design

Evaluation methodology

**Background of comparison group selection process**

IDinsight used propensity score matching to create a comparison group of incoming university students that had similar characteristics to the 2013 cohort of Kepler students.\(^1\) We identified students who would have had a high likelihood of enrolling in Kepler if given the opportunity, based on key characteristics such as prior education and socioeconomic status.

At the Year 1 baseline, IDinsight selected 100 comparison students from a pool of over 500 students who would be attending a traditional Rwandan university: University of Rwanda (UR) College of Business and Economics, UR College of Education, UR College of Science and Technology, and Adventist University of Central Africa. Each Kepler student was matched to two non-Kepler students from the comparison group with similar baseline characteristics.

Figure 1: Comparison student selection process (2013 cohort)

IDinsight created a comparison group that was statistically similar to the incoming cohort of Kepler students based on observable characteristics.\(^2,3\)

**Impact measurement methods**

The impact of the Kepler model was estimated as follows:

1. **Three-year impact for the 2013 cohort:** The cumulative three-year impact of Kepler for the 2013 cohort was estimated by calculating the difference in average Year 3 endline test results between Kepler and comparison students.

2. **Third year impact for the 2013 cohort:** The discrete impact of the third year of Kepler was estimated by comparing the change in test scores between the Year 2 endline and the Year 3 endline for Kepler and comparison students.

All estimates include controls for important characteristics likely to affect learning outcomes (“covariates”), such as age, gender, previous academic achievement, family background, and other relevant characteristics. For further details on the covariates, propensity score matching, and a discussion on limitations of the study, please refer to Appendix A and B.

---

\(^1\) Kepler considered randomizing admissions for a portion of the incoming class, but the admissions process did not yield enough qualified students to randomize without decreasing the size or quality of the incoming cohort at the time.

\(^2\) Propensity score matching balanced the treatment and comparison group on a set of observable characteristics that are correlated with the main outcomes. However, propensity score matching does not guarantee balance on unobservable characteristics, which can affect the results if comparison students did not go to Kepler for reasons that are correlated with the outcomes.

\(^3\) At baseline, Kepler students were slightly less likely to be poor than comparison students based on the Progress out of Poverty index (a metric that estimates the probability that the respondent lives below the national poverty line, www.progressoutofpoverty.org) and nearly identical on Ubudehe score (a Rwandan measure of socioeconomic status).
Metrics
Kepler identified critical thinking, cognitive skills (English grammar, math, and logic), English language skills, and computer literacy as the most important indicators of success of the Kepler academic program. A series of exams was administered at the end of the 2015–2016 academic year to assess student learning in these areas.

Table 1: Summary of Year 3 endline exams

<table>
<thead>
<tr>
<th>Category</th>
<th>Summary</th>
</tr>
</thead>
</table>
| Critical thinking  | **Watson-Glaser II:** Paper-based multiple-choice exam to measure inference, deduction, interpretation, recognizing assumptions, and evaluating arguments.  
CLA+ exam: Computer-based exam with one essay response and multiple-choice questions to assess critical thinking, problem solving, scientific and quantitative reasoning, writing, and the ability to make and critique arguments. |
| Cognitive skills   | **IDInsight-developed multiple-choice test:** Paper-based test modeled on popular standardized exams (SLE, SAT, ACT), which includes English, math, and logical reasoning questions. |
| English language   | **Abridged IELTS reading and writing tests:** Paper-based tests with multiple-choice and fill-in-the-blank questions in response to reading passages. The writing test was an essay response to a general prompt. |
| Computer literacy  | **Typing speed and IDInsight-developed tasks:** Typing speed test, web research, website credibility, Microsoft Word, Microsoft Excel, and email. |

Critical thinking
The Watson-Glaser II Critical Thinking Appraisal (“Watson-Glaser”) was used as one measure of students’ critical thinking ability. The test measures (i) inference, (ii) deduction, (iii) interpretation, (iv) recognizing assumptions, and (v) evaluating arguments. The paper-based multiple-choice exam lasted 50 minutes. The test version administered in April–May 2016 was also administered at Year 1 endline in 2014; however, Watson-Glaser’s guidelines permit administration of the same version after as little as one year without adverse effects.

The CLA+ exam was used as a second measure of critical thinking ability. The exam consists of two sections—a “performance task” and “selected-response” questions—used to measure critical thinking, problem solving, scientific and quantitative reasoning, writing, and the ability make and critique arguments. In the performance task, students were presented with a real-world situation and asked to address the issue, suggest a solution, or recommend a course of action based on the supporting documentation in essay form in two hours. In the selected-response section, students responded to ten questions to assess scientific and quantitative reasoning, ten to assess critical reading and evaluation, and five to assess the ability to critique an argument over one hour. The Council for Aid to Education, the company that designed and owns the exam, graded the CLA+ exam.

Cognitive skills
IDInsight developed a multiple-choice cognitive skills test that included English grammar, math, and logical reasoning questions. The test was modeled on common standardized tests, such as the SAT, ACT, and the Scholastic Level Exam (SLE). All test content was original, or derived from public domain resources. Students were given 20 minutes to complete the 30 questions on the test.
English language

IDinsight administered abridged versions of the International English Language Testing System (IELTS) reading and writing tests. Questions were sourced from the General Training and Academic IELTS practice tests. The 2013 cohort Kepler and comparison students had not previously taken versions of the IELTS tests administered this year.

The IELTS reading test consisted of multiple-choice and fill-in-the-blank questions in response to two reading passages. Students had 40 minutes to complete 25 questions.

The IELTS writing test asked students to respond to a general prompt in essay format in 40 minutes. Essays were graded on task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy by a grader experienced in evaluating IELTS essays.

Computer literacy

The computer literacy test was comprised of six sections:

1. **Typing speed test.**
2. **Website credibility:** Assess if given websites are credible enough to be cited in an academic paper.
3. **Web research:** Find answers to a list of questions using a web search.
4. **Microsoft Word:** Edit a Word document according to instructions, such as changing font size, adding hyperlinks, and using spellcheck.
5. **Microsoft Excel:** Edit an Excel spreadsheet according to instructions, such as adding a table, sorting, and drawing a scatterplot.
6. **Email:** Submit Word and Excel tests via email. Graded on ability to log in and send email, use the subject line appropriately, add appropriate email content, and attach documents.

The computer tests were administered sequentially over approximately one hour. The aggregate computer literacy score was constructed by equally weighting each section’s score expressed as a percent of questions answered correctly.  

**Testing periods**

Tests have been administered in various combinations during baseline and endline administrations over the past three years. The Year 3 endline test administration included all tests.

<table>
<thead>
<tr>
<th>Table 2: Test administration timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year 1 Baseline</strong></td>
</tr>
<tr>
<td>CLA+</td>
</tr>
<tr>
<td>Watson-Glaser II</td>
</tr>
<tr>
<td>Cognitive skills: English</td>
</tr>
<tr>
<td>Cognitive skills: Math</td>
</tr>
<tr>
<td>Cognitive skills: Logic</td>
</tr>
<tr>
<td>English (IELTS)</td>
</tr>
<tr>
<td>Computer literacy</td>
</tr>
</tbody>
</table>

---

4 There is no maximum typing score; therefore, 50 words per minute or above (the target typing speed for Kepler) was considered as 100%.
Previous findings

Years 1 and 2 of the impact evaluation found that the Kepler curriculum was associated with positive learning gains relative to comparison universities.

Table 3: Overview of previous test results (percentage point difference between Kepler and comparison)

<table>
<thead>
<tr>
<th></th>
<th>Year 1 Endline</th>
<th>Year 2 Endline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical thinking</td>
<td>Inconclusive</td>
<td>5.4 pp</td>
</tr>
<tr>
<td>Cognitive skills</td>
<td>9.8 pp</td>
<td>16.4 pp</td>
</tr>
<tr>
<td>English language*</td>
<td>20.0 pp</td>
<td>10.4 pp</td>
</tr>
<tr>
<td>Computer literacy</td>
<td>24.7 pp</td>
<td>35.9 pp</td>
</tr>
</tbody>
</table>

* Different tests administered at Y1 and Y2 endline

Year 1 evaluation results

In the first year of the program, Kepler wanted to conduct an impact evaluation to generate indicative evidence about whether the model was promising, or if it needed a significant pivot.\(^5\)

Year 1 results indicated:

- **Kepler students performed better than comparison students in cognitive skills, English, and computer literacy.** Kepler students performed 9.8 percentage points (pp) better than comparison students on an IDinsight-developed cognitive skills test administered at midline. Kepler students performed 20pp better on the English comprehension exam at endline and 24.7pp better on the midline computer literacy test that assessed typing speed, internet research, Microsoft Word, and Microsoft Excel.

- **There was no evidence of an impact of Kepler on critical thinking.** The small difference between Kepler and comparison scores on the Watson-Glaser paper-based critical thinking exam was not statistically significant. Kepler students performed better than the comparison group on the CLA+ exam, but the results were not statistically significant and can likely be attributed to Kepler students’ better computer and language skills.

Year 2 evaluation results

For the 2013 cohort, Year 2 of the evaluation indicated:\(^6\)

- **The effect of Kepler on critical thinking emerged in the second year.** The evaluation found that Kepler students performed 5.4pp higher on the Watson-Glaser test than comparison students. This was in contrast to the Year 1 results, which showed no significant differences in critical thinking between Kepler and comparison students.

- **Kepler students continued to outperform their peers from traditional Rwandan universities in cognitive skills, English language, and computer literacy.** Kepler students did not show additional gains relative to their peers in cognitive skills during the second year. The Year 2 tests introduced the IELTS English reading and writing tests, on which Kepler students scored 10.5pp and 3.7pp higher than comparison students, respectively. Kepler students scored 35.9pp higher than comparison students on the computer literacy tests.

---

\(^5\) Due to the study’s limitations—small sample size, lack of a true baseline before the beginning of the school year, and the possibility that the matched comparison group may differ in unobservable ways from the Kepler students—the results were qualified. Please refer to **Year One Kepler Endline Report** for additional information on Year One methodology and findings.

\(^6\) In Year 2, IDinsight also identified a matched comparison group for the 2014 cohort of Kepler students and tested 2014 cohort students at baseline and endline. Please refer to **Kepler Year Two Evaluation Endline Report** for additional information on Year 2 methodology and findings.
Year 3 evaluation results

Interpreting effect sizes

Results from the Year 3 endline examination of the 2013 cohort are presented below. Test scores were converted from raw points to the percentage of total possible points on the test. Differences in test scores between Kepler students and comparison students on the Year 3 endline tests are presented both in terms of standardized effect sizes, expressed in standard deviations (SD), and percentage points, where relevant. The standardized effect size was calculated by dividing the difference in test scores by the standard deviation of the comparison students’ Year 3 endline results in order to assess the difference in tests scores relative to the spread in scores. Standardization enables comparisons of the magnitudes of effect sizes irrespective of the different grading scales used on different tests.

In education studies, a standardized effect size of less than 0.1 SD is considered small, while a standardized effect size of more than 0.3 SD is considered large, and a standardized effect size of more than 0.5 SD is considered very large. However, these definitions depend strongly on the outcome being tested and the assessment.

All test scores and impact estimates presented in this report control for differences in covariates, including student demographics and previous academic performance. Due to this adjustment, average test scores presented here vary slightly from raw test scores that did not adjust for differences in covariates. Please refer to Appendix C for unadjusted average test scores.

For full regression results, please refer to Appendix D.

Summary of results

Kepler students in the 2013 cohort continued to perform better than comparison students after three years at Kepler. IDinsight measured the three-year effect of Kepler on critical thinking, cognitive skills, English language, and computer literacy (Figure 2). The biggest difference between Kepler and comparison students, in terms of standardized effect sizes, was in critical thinking as measured by the CLA+ exam and computer literacy. The English grammar section of the cognitive skills exam showed the smallest difference between Kepler and comparison students, although Kepler students continued to perform better than comparison students on the reading and writing IELTS English tests.

Figure 2 shows the average difference in Year 3 endline test scores after controlling for differences in student demographics. The height of the bar represents the three-year learning gains in terms of the standardized effect size. The average difference in test scores is presented below the bars for each test expressed in percentage points as well, along with the significance of the difference as shown by the p-value.

---

7 The standard deviations of the comparison students’ Year 3 endline test scores are used as the standardized effect size unit because they represent the spread in test scores that would be expected from students who did not attend Kepler.

8 A p-value denotes the likelihood that the observed difference between Kepler and comparison students is due to chance. A lower p-value increases our confidence that the difference is attributable to Kepler rather than to extraneous factors, such as small baseline differences between Kepler and comparison students.
Figure 2: Three-year impact of Kepler

Kepler students maintained their higher performance levels during their third year, but did not make additional gains relative to comparison students. There were no statistically significant differences between the cumulative two-year and three-year impacts of Kepler, except for the English grammar section of the cognitive skills test, in which comparison students partially closed the gap with Kepler students (Figure 3). Although Kepler students did not show further gains relative to the comparison students during their third year, they maintained their higher performance levels in all skill areas.

Figure 3 illustrates the impact of the third year of Kepler on student academic performance in terms of standardized effect sizes. For each test, the bar on the left represents the cumulative two-year impact of Kepler, while the bar on the right represents the three-year impact. The difference in impact between the third year of Kepler and the second year is presented below the bars.

Figure 3: Third year impact of Kepler
Results by test

The following sections describe the results for each test conducted in the Year 3 endline, including the difference-in-differences (DID) results between Year 2 and Year 3 outcomes in terms of standardized effect size. The DID estimate helps reveal the impact of the third year Kepler curriculum on learning outcomes. A positive DID estimate indicates that Kepler students made additional gains relative to comparison students during the third year of university (i.e. between Year 2 and Year 3 endline exams); a negative DID estimate indicates that comparison students are closing the gap with Kepler students during the third year.

Critical thinking

Kepler students demonstrated higher critical thinking ability than comparison students on two different measures of critical thinking: the Watson-Glaser exam and the CLA+ exam. After the first year of Kepler, Kepler students did not show significantly better critical thinking skills on the Watson-Glaser exam or CLA+ compared to the control group. At the end of the second year, Kepler students performed 0.75 SD better than the comparison group on the Watson-Glaser exam (the CLA+ was not administered in Year 2). After three years, Kepler students continued to show better performance on critical thinking than their peers. However, the differences in test scores between Kepler and comparison students remained the same on the Watson-Glaser exam, indicating no additional gains from the third year curriculum. The standardized effect size of CLA+ exam is large (1.52 SD), which could be partially due to the Kepler students’ comfort using computer-based programs and faster typing speed.⁹

Cognitive skills

Kepler students performed better than comparison students on all three areas of cognitive skills—English, math, and logic—although Kepler students lost some advantage in English grammar. Kepler students continued to outperform their peers in cognitive skills, with differences of 0.39 SD in English, 0.69 SD in math, and 0.75 SD in logic. However, the difference in English grammar skills was 0.57 SD smaller in the third year than it was at the end of the second year. Kepler students performed almost the same on the English section of the cognitive skills exam at the end of Year 2 and Year 3, while comparison students improved their scores, therefore significantly decreasing the gap.

English language

Kepler students performed significantly better on English reading and writing exams than the comparison students. Kepler students performed 1.24 SD and 0.98 SD better on the English reading and writing exams, respectively, at the end of Year 3. The DID estimates for reading and writing were positive, but not significant.

Computer literacy

Kepler students continued to far outperform the comparison students on computer literacy tests, but did not make further gains in the third year. Each year, Kepler students have shown large, positive advantages in computer skills over comparison students. The difference remained well over 2 SD at the end of Year 3. The difference in computer literacy test scores was slightly lower in Year 3 than at the end of Year 2 (2.78 SD in Year 2 compared to 2.64 SD in Year 2), although the DID estimate is not significant.

⁹ The CLA+ exam was not administered in Year 2, therefore, a DID estimate was not possible for this test.
Year 3 survey findings

All students who participated in the Year 3 endline exams completed an online survey at the conclusion of the testing session about post-graduation employment plans, including whether or not they have secured a full-time position, sector, salary, hours, benefits, and resources used to seek and secure employment.

Employment plans

Kepler students are more likely to have secured post-graduation employment than comparison students at the end of the third year of university. 73.8% of 2013 cohort Kepler students have confirmed full-time employment for after university at the time of surveying (April–May 2016). 10.3% of comparison students had secured employment at this time. No Kepler students reported planning to be self-employed after graduation, even though Kepler encourages entrepreneurship.

Table 4: Post-graduation employment plans

<table>
<thead>
<tr>
<th>Employment Status</th>
<th>Kepler</th>
<th>Count</th>
<th>Non-Kepler</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes: employment secured</td>
<td>73.8%</td>
<td>31</td>
<td>10.5%</td>
<td>9</td>
</tr>
<tr>
<td>No: employment not secured</td>
<td>26.2%</td>
<td>11</td>
<td>81.4%</td>
<td>70</td>
</tr>
<tr>
<td>Will be self-employed</td>
<td>0%</td>
<td>0</td>
<td>8.1%</td>
<td>7</td>
</tr>
</tbody>
</table>

N=42 N=86

Resources used in the employment search

Kepler students used career services, previous employers, and active networking most often in the employment search to seek and secure employment (Figure 4). Students who have not yet secured employment have been searching for jobs for eight months on average (as of May 2016).

Figure 4: Resources used by Kepler students in the employment search

\[10\] Most comparison students have a fourth year of university before graduation.
Sector

Kepler students who have secured employment will be employed primarily in business, education, and communication sectors (Figure 5). The comparison students who have secured employment will be working mostly in business (7 students).

Figure 5: Sector of confirmed employment for Kepler students

![Sector of post-graduate employment (Kepler students)](chart.png)

Terms of employment

Kepler students with confirmed employment will work 40-hour weeks. Comparison students expect to work 25 hours per week on average in their confirmed employment.

Kepler students who have secured employment will earn 211,860 RWF per month, on average. Comparison students reported a higher average starting salary (302,500 RWF per month), however, three of the nine reported salaries seemed unusually high (over 500,000 RWF per month). When the potential outliers are removed, comparison students with confirmed employment will earn 203,333 RWF per month. There was a range of satisfaction among Kepler students with their starting salaries (Figure 6).
Most Kepler students with confirmed employment will receive medical and sick leave benefits. Students with secured jobs after graduation are generally satisfied with the benefits they will receive in their employment (Figure 7).

### Table 5: Employment benefits for Kepler students

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical</td>
<td>29</td>
<td>93.5%</td>
</tr>
<tr>
<td>Sick leave</td>
<td>21</td>
<td>67.7%</td>
</tr>
<tr>
<td>Paid vacation</td>
<td>14</td>
<td>45.1%</td>
</tr>
<tr>
<td>Transport</td>
<td>11</td>
<td>35.4%</td>
</tr>
<tr>
<td>Retirement</td>
<td>9</td>
<td>29.0%</td>
</tr>
<tr>
<td>Dental</td>
<td>9</td>
<td>29.0%</td>
</tr>
<tr>
<td>Free meals</td>
<td>5</td>
<td>16.1%</td>
</tr>
<tr>
<td>Unpaid vacation</td>
<td>4</td>
<td>12.9%</td>
</tr>
<tr>
<td>Tuition credit</td>
<td>3</td>
<td>9.7%</td>
</tr>
<tr>
<td>Housing</td>
<td>3</td>
<td>9.7%</td>
</tr>
</tbody>
</table>

N=31

**Expected stay**

Kepler students expected to stay with the organization that hired them for approximately 4.5 years after graduation. Comparison students with confirmed employment expected to stay for slightly longer (5.6 years), although the difference is not significant.
Conclusion

After three years at Kepler, students continued to perform better in critical thinking, cognitive skills, English, and computer literacy than their peers in traditional Rwandan universities. However, the differences in performance between Kepler and comparison students have plateaued during the third year. This could be explained by the fact that most third year students are primarily completing self-paced online learning and participating in the workforce, with limited time in the classroom, while their peers in traditional Rwandan universities continue to attend classes full-time. In addition to academic outcomes, Kepler is preparing students for the workforce, as demonstrated by the fact that nearly three-quarters of the 2013 cohort has secured full-time employment after graduation as of May 2016.
Appendices

Appendix A: Propensity score matching for the 2013 cohort .................................................. 17
Appendix B: Study Limitations .................................................................................................. 19
Appendix C: Raw test scores .................................................................................................... 20
Appendix D: Regression results ................................................................................................. 21
Appendix E: Subgroup analysis ................................................................................................. 23
  Gender .................................................................................................................................... 23
  Domicile ................................................................................................................................. 23
Appendix F: Example test questions .......................................................................................... 25
  Watson-Glaser II .................................................................................................................. 25
  CLA+ .................................................................................................................................... 26
  Cognitive skills ...................................................................................................................... 29
  IELTS Reading ...................................................................................................................... 30
  IELTS Writing ...................................................................................................................... 32
  Computer literacy .................................................................................................................. 32
Appendix A: Propensity score matching for the 2013 cohort

This study used a propensity score matching methodology to identify students in traditional Rwandan universities who would likely have enrolled in Kepler if given the opportunity. A logistic regression with key covariates was used to estimate the probability (“propensity score”) that the comparison student would have enrolled with the 2013 Kepler class.

Covariates included:

- Age
- Gender
- Urban/rural background
- Poverty level
- Parents’ living status
- Previous employment
- Secondary school marks
- Secondary school type
- Years of computer use
- Expected earnings five years after graduation.

Selection and matching

Comparison group selection was a four-step process:

1. Screened 500+ students from traditional universities in Kigali (Adventist University of Central Africa, University of Rwanda (UR) College of Science and Technology, UR College of Business and Economics, and UR College of Education).
2. Selected 200 students that best match the Kepler demographic based on age, gender, secondary school marks, poverty level, and national exam scores.
3. Administered a survey on key background characteristics, including urban/rural background, parents’ living status, previous employment, secondary school type, years of computer use, and expected earnings five years after graduation.
4. Selected 100 comparison students that best matched the Kepler students using propensity score matching.

Almost all comparison students who were invited to participate in the evaluation did ultimately participate.

Comparison students did not attend Kepler because most did not apply in the first place. In its first year, Kepler's recruitment did not reach all possible applicants. As such, matching attempts to identify those students who were likely to have applied and been accepted to Kepler if given the opportunity.

Baseline balance after matching

Kepler and comparison students were balanced on most key observable characteristics at baseline. Differences between Kepler and comparison students’ senior 6 marks and national exam scores were negligible. Kepler students were on average slightly less likely to be poor by one poverty measure (PPI)\(^{11}\) and nearly identical in terms of Ubudehe. While there was an evident difference in students’ urban vs. rural status, this difference was not statistically significant at the 5% significance level. There were no differences between Kepler students and comparison students on demographics, including gender and age. There was also no significant difference in computer use. The lack of

\(^{11}\) Between the Kepler and comparison group, there were significant differences on the following poverty questions: home floor material, home lighting source, and cooking fuel used, all of which point to differences in the urban vs. rural makeup of the two groups.
differences suggests that Kepler and comparison students were similar on average, and differences in outcomes between the two groups can likely be attributed to Kepler.

The table below shows the covariate averages between the Kepler and matched comparison groups with corresponding $p$-values.

Table 1: Baseline balance

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Kepler</th>
<th>Matched Comparison</th>
<th>Difference</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior 6 marks</td>
<td>77.82</td>
<td>76.32</td>
<td>1.50</td>
<td>.101</td>
</tr>
<tr>
<td>National exam percentage</td>
<td>83.41</td>
<td>83.16</td>
<td>.25</td>
<td>.926</td>
</tr>
<tr>
<td>Ubudehe score</td>
<td>2.62</td>
<td>2.68</td>
<td>-.06</td>
<td>.653</td>
</tr>
<tr>
<td>Progress Out of Poverty index</td>
<td>56.92</td>
<td>49.81</td>
<td>7.11</td>
<td>.004***</td>
</tr>
<tr>
<td>% from an urban location</td>
<td>54.00</td>
<td>39.00</td>
<td>15.00</td>
<td>.082*</td>
</tr>
<tr>
<td>Gender (% Female)</td>
<td>48.00</td>
<td>45.00</td>
<td>3.00</td>
<td>.730</td>
</tr>
<tr>
<td>Age</td>
<td>21.40</td>
<td>21.82</td>
<td>-.42</td>
<td>.129</td>
</tr>
<tr>
<td>Years of computer use</td>
<td>4.92</td>
<td>4.81</td>
<td>0.11</td>
<td>.799</td>
</tr>
</tbody>
</table>
Appendix B: Study Limitations

Limitations of matching

The 2013 cohort students were matched based on initial screenings of 500 potential comparison students’ previous academic performance and demographic surveys. For comparison, the 2014 cohort comparison students were identified from an initial pool of 2800+ candidates who went through a more rigorous simulation of the Kepler admission process. The comparison students matched to the 2013 cohort may be less accurate matches than the 2014 cohort due to these different approaches.

Additionally, regardless of the approach to matching, any matching methodology cannot directly capture characteristics that are not fully quantifiable or observable, such as ambition, work ethic, and motivation. Although the treatment and comparison groups appeared balanced on observable characteristics, it is impossible to be certain they were balanced on unobservable characteristics as well.

Delayed baseline

Due to timing challenges, the Year 1 evaluation did not include a true baseline before university classes began. The testing window in October 2013 began 18 days after the school year had started for Kepler students and between 29 and 36 days for comparison students. Kepler students performed significantly better than comparison students on baseline tests (CLA+, SLE, and typing speed), which is unlikely due to differences in aptitude and critical thinking alone since Kepler and comparison students have statistically similar high school grades, national exam scores, and demographics. Instead, these differences could have been due to the fact that Kepler students received significant computer and English training in their first weeks of school before the baseline was administered. Due to these differences, Year 1 baseline tests scores were not used as controls in analysis.

Sample size

The sample size of the evaluation was constrained by the size of the first Kepler cohort. A small sample size decreases the power of the evaluation to detect differences between the treatment and comparison groups. Each of the 50 Kepler students was matched to two comparison students to improve the precision of the impact estimates.

Computer literacy of Kepler students

Kepler students were more proficient in computer skills than comparison students. This may have lent an advantage to Kepler students on the CLA+ exam since they could more easily navigate the test and type more quickly. In order to disentangle the effect of computer proficiency from critical thinking skills, typing speed was added as a control when assessing the difference in CLA+ exam scores between treatment and control groups.

Attrition

If students drop out of the sample over the course of the evaluation (“attrite”), then the problems of small sample and possible imbalance between Kepler and comparison groups may be exacerbated. Attrition was reduced by offering incentives for participation: (1) all students received 7500 RWF as compensation for their time and transportation, and (2) a 5000 RWF bonus was offered to students who scored in the top 10% from each university (aggregate score, all tests weighted equally) and the two students from each school who showed the largest improvement from Year 2 endline tests to Year 3 endline tests.

Attrition was very low, with only a few students dropping out due to reasons unrelated to the evaluation. Three Kepler students were not tested because they were no longer in Kepler or due to illness. Five comparison students refused to participate in testing this year.
Appendix C: Raw test scores

The following table lists the means and standard deviations for tests administered at the Year 3 endline in April–May 2016. These scores have not controlled for differences in covariate values, i.e. they do not account for differences in scores that may be due to student demographics or previous academic performance. Results presented elsewhere in this report controlled for the covariates.

Table 2: Raw test scores (unadjusted for covariates)

<table>
<thead>
<tr>
<th>Test</th>
<th>Kepler N</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Comparison N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical thinking: Watson-Glaser</td>
<td>41</td>
<td>54.8%</td>
<td>12.2 pp</td>
<td>87</td>
<td>48.0%</td>
<td>8.5 pp</td>
</tr>
<tr>
<td>Critical thinking: CLA+*</td>
<td>39</td>
<td>934.8</td>
<td>130.4 points</td>
<td>55</td>
<td>750.5</td>
<td>110.4 points</td>
</tr>
<tr>
<td>Cognitive skills: English</td>
<td>41</td>
<td>50.2%</td>
<td>19.9 pp</td>
<td>87</td>
<td>39.8%</td>
<td>19.0 pp</td>
</tr>
<tr>
<td>Cognitive skills: Logic</td>
<td>41</td>
<td>43.0%</td>
<td>16.7 pp</td>
<td>87</td>
<td>30.9%</td>
<td>14.4 pp</td>
</tr>
<tr>
<td>Cognitive skills: Math</td>
<td>41</td>
<td>58.5%</td>
<td>21.1 pp</td>
<td>87</td>
<td>41.2%</td>
<td>18.8 pp</td>
</tr>
<tr>
<td>IELTS Reading</td>
<td>41</td>
<td>43.8%</td>
<td>13.7 pp</td>
<td>88</td>
<td>28.7%</td>
<td>10.4 pp</td>
</tr>
<tr>
<td>IELTS Writing</td>
<td>41</td>
<td>65.1%</td>
<td>7.0 pp</td>
<td>88</td>
<td>57.3%</td>
<td>7.8 pp</td>
</tr>
<tr>
<td>Computer literacy**</td>
<td>41</td>
<td>64.7%</td>
<td>11.5 pp</td>
<td>88</td>
<td>28.1%</td>
<td>13.2 pp</td>
</tr>
</tbody>
</table>

*CLA+ is reported in points, rather than a percentage score

**Computer literacy test consists of typing speed, web research, website credibility, Microsoft Word, Microsoft Excel, and email tests.
Appendix D: Regression results

Below are tables that list coefficients (with p-values in brackets) from regression results. All test scores are expressed in standard deviations of the 2013 cohort’s comparison students’ test scores for that test. In all tables, one asterisk indicates statistical significance at the 10% significance level, two asterisks indicate statistical significance at the 5% significance level, and three asterisks represent statistical significance at the 1% significance level.

Table 3: Three-year impact for 2013 cohort

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (student enrolled in Kepler)</td>
<td>0.681**</td>
<td>1.519***</td>
<td>0.393*</td>
<td>0.695***</td>
<td>0.749***</td>
<td>0.984***</td>
<td>1.244***</td>
<td>2.644***</td>
</tr>
<tr>
<td></td>
<td>[0.015]</td>
<td>[0.000]</td>
<td>[0.078]</td>
<td>[0.002]</td>
<td>[0.001]</td>
<td>[0.000]</td>
<td>[0.000]</td>
<td>[0.000]</td>
</tr>
<tr>
<td>Age</td>
<td>-0.043</td>
<td>-0.064</td>
<td>-0.062</td>
<td>-0.153**</td>
<td>-0.078</td>
<td>-0.044</td>
<td>-0.116*</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td>[0.515]</td>
<td>[0.245]</td>
<td>[0.229]</td>
<td>[0.010]</td>
<td>[0.110]</td>
<td>[0.408]</td>
<td>[0.063]</td>
<td>[0.262]</td>
</tr>
<tr>
<td>Female (binary)</td>
<td>-0.153*</td>
<td>-0.327*</td>
<td>-0.254</td>
<td>-0.655***</td>
<td>-0.478***</td>
<td>-0.336*</td>
<td>-0.440**</td>
<td>-0.442**</td>
</tr>
<tr>
<td></td>
<td>[0.457]</td>
<td>[0.096]</td>
<td>[0.174]</td>
<td>[0.002]</td>
<td>[0.010]</td>
<td>[0.075]</td>
<td>[0.020]</td>
<td>[0.019]</td>
</tr>
<tr>
<td>Exposure to English at home (binary)</td>
<td>-0.193</td>
<td>-0.196</td>
<td>-0.191</td>
<td>-0.053</td>
<td>-0.011</td>
<td>-0.004</td>
<td>0.021</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>[0.475]</td>
<td>[0.435]</td>
<td>[0.401]</td>
<td>[0.811]</td>
<td>[0.963]</td>
<td>[0.988]</td>
<td>[0.932]</td>
<td>[0.643]</td>
</tr>
<tr>
<td>Urban (binary)</td>
<td>0.073</td>
<td>0.162</td>
<td>0.073</td>
<td>0.131</td>
<td>-0.131</td>
<td>-0.220</td>
<td>0.030</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>[0.754]</td>
<td>[0.425]</td>
<td>[0.726]</td>
<td>[0.563]</td>
<td>[0.484]</td>
<td>[0.309]</td>
<td>[0.883]</td>
<td>[0.741]</td>
</tr>
<tr>
<td>PPI score</td>
<td>-0.004</td>
<td>0.009</td>
<td>0.004</td>
<td>0.009</td>
<td>0.021***</td>
<td>0.005</td>
<td>0.014</td>
<td>0.016*</td>
</tr>
<tr>
<td></td>
<td>[0.646]</td>
<td>[0.322]</td>
<td>[0.569]</td>
<td>[0.300]</td>
<td>[0.004]</td>
<td>[0.606]</td>
<td>[0.121]</td>
<td>[0.059]</td>
</tr>
<tr>
<td>Both parents alive (binary)</td>
<td>-0.551***</td>
<td>0.144</td>
<td>-0.370***</td>
<td>-0.139</td>
<td>-0.107</td>
<td>0.032</td>
<td>-0.178</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>[0.006]</td>
<td>[0.445]</td>
<td>[0.044]</td>
<td>[0.453]</td>
<td>[0.552]</td>
<td>[0.859]</td>
<td>[0.355]</td>
<td>[0.735]</td>
</tr>
<tr>
<td>Senior 5 marks</td>
<td>0.004</td>
<td>0.017</td>
<td>0.022</td>
<td>0.030</td>
<td>0.006</td>
<td>-0.013</td>
<td>-0.010</td>
<td>-0.021</td>
</tr>
<tr>
<td></td>
<td>[0.893]</td>
<td>[0.487]</td>
<td>[0.408]</td>
<td>[0.171]</td>
<td>[0.806]</td>
<td>[0.577]</td>
<td>[0.653]</td>
<td>[0.230]</td>
</tr>
<tr>
<td>Senior 6 marks</td>
<td>-0.000</td>
<td>0.013</td>
<td>0.004</td>
<td>-0.012</td>
<td>-0.000</td>
<td>0.013</td>
<td>0.020</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>[0.986]</td>
<td>[0.487]</td>
<td>[0.865]</td>
<td>[0.534]</td>
<td>[0.985]</td>
<td>[0.502]</td>
<td>[0.325]</td>
<td>[0.376]</td>
</tr>
<tr>
<td>Years of computer use</td>
<td>0.060</td>
<td>0.029</td>
<td>0.068*</td>
<td>0.035</td>
<td>0.038</td>
<td>0.029</td>
<td>0.072*</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>[0.172]</td>
<td>[0.501]</td>
<td>[0.096]</td>
<td>[0.399]</td>
<td>[0.291]</td>
<td>[0.374]</td>
<td>[0.096]</td>
<td>[0.369]</td>
</tr>
<tr>
<td>Household owns computers (binary)</td>
<td>0.602</td>
<td>0.104</td>
<td>-0.046</td>
<td>-0.519*</td>
<td>0.187</td>
<td>0.020</td>
<td>-0.155</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>[0.108]</td>
<td>[0.777]</td>
<td>[0.892]</td>
<td>[0.086]</td>
<td>[0.497]</td>
<td>[0.945]</td>
<td>[0.679]</td>
<td>[0.676]</td>
</tr>
<tr>
<td>Constant</td>
<td>0.871</td>
<td>-1.448</td>
<td>-0.902</td>
<td>1.638</td>
<td>0.277</td>
<td>0.850</td>
<td>0.968</td>
<td>0.788</td>
</tr>
<tr>
<td></td>
<td>[0.752]</td>
<td>[0.506]</td>
<td>[0.680]</td>
<td>[0.435]</td>
<td>[0.879]</td>
<td>[0.658]</td>
<td>[0.638]</td>
<td>[0.676]</td>
</tr>
<tr>
<td>Observations</td>
<td>129</td>
<td>128</td>
<td>128</td>
<td>128</td>
<td>129</td>
<td>129</td>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.200</td>
<td>0.397</td>
<td>0.157</td>
<td>0.273</td>
<td>0.293</td>
<td>0.239</td>
<td>0.371</td>
<td>0.701</td>
</tr>
</tbody>
</table>
### Table 4: Third year impact for the 2013 cohort

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (student enrolled in Kepler)</td>
<td>-0.065</td>
<td>-0.571**</td>
<td>0.020</td>
<td>0.090</td>
<td>0.224</td>
<td>0.297</td>
<td>-0.134</td>
</tr>
<tr>
<td></td>
<td>[0.825]</td>
<td>[0.022]</td>
<td>[0.930]</td>
<td>[0.729]</td>
<td>[0.285]</td>
<td>[0.118]</td>
<td>[0.350]</td>
</tr>
<tr>
<td>Age</td>
<td>0.054</td>
<td>0.028</td>
<td>-0.092*</td>
<td>-0.087</td>
<td>-0.025</td>
<td>0.075</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>[0.491]</td>
<td>[0.608]</td>
<td>[0.092]</td>
<td>[0.155]</td>
<td>[0.647]</td>
<td>[0.145]</td>
<td>[0.402]</td>
</tr>
<tr>
<td>Female (binary)</td>
<td>0.356</td>
<td>0.188</td>
<td>-0.235</td>
<td>-0.061</td>
<td>-0.343*</td>
<td>0.012</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>[0.159]</td>
<td>[0.336]</td>
<td>[0.246]</td>
<td>[0.796]</td>
<td>[0.093]</td>
<td>[0.940]</td>
<td>[0.598]</td>
</tr>
<tr>
<td>Exposure to English at home (binary)</td>
<td>-0.375</td>
<td>-0.497**</td>
<td>-0.235</td>
<td>-0.298</td>
<td>-0.355</td>
<td>-0.102</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>[0.239]</td>
<td>[0.036]</td>
<td>[0.315]</td>
<td>[0.347]</td>
<td>[0.233]</td>
<td>[0.624]</td>
<td>[0.333]</td>
</tr>
<tr>
<td>Urban (binary)</td>
<td>0.110</td>
<td>-0.009</td>
<td>0.301</td>
<td>-0.068</td>
<td>-0.055</td>
<td>-0.243</td>
<td>-0.344**</td>
</tr>
<tr>
<td></td>
<td>[0.686]</td>
<td>[0.969]</td>
<td>[0.175]</td>
<td>[0.772]</td>
<td>[0.787]</td>
<td>[0.189]</td>
<td>[0.019]</td>
</tr>
<tr>
<td>PPI score</td>
<td>-0.012</td>
<td>0.017*</td>
<td>-0.006</td>
<td>0.029***</td>
<td>-0.001</td>
<td>0.008</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>[0.318]</td>
<td>[0.078]</td>
<td>[0.504]</td>
<td>[0.001]</td>
<td>[0.944]</td>
<td>[0.360]</td>
<td>[0.341]</td>
</tr>
<tr>
<td>Both parents alive (binary)</td>
<td>-0.476*</td>
<td>-0.386*</td>
<td>0.089</td>
<td>-0.073</td>
<td>-0.183</td>
<td>-0.408**</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>[0.063]</td>
<td>[0.051]</td>
<td>[0.632]</td>
<td>[0.737]</td>
<td>[0.344]</td>
<td>[0.018]</td>
<td>[0.681]</td>
</tr>
<tr>
<td>Senior 5 marks</td>
<td>0.043</td>
<td>0.042</td>
<td>0.017</td>
<td>0.016</td>
<td>-0.009</td>
<td>0.020</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>[0.145]</td>
<td>[0.111]</td>
<td>[0.408]</td>
<td>[0.609]</td>
<td>[0.725]</td>
<td>[0.325]</td>
<td>[0.524]</td>
</tr>
<tr>
<td>Senior 6 marks</td>
<td>-0.059**</td>
<td>0.010</td>
<td>-0.008</td>
<td>-0.026</td>
<td>0.004</td>
<td>-0.010</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>[0.047]</td>
<td>[0.662]</td>
<td>[0.711]</td>
<td>[0.256]</td>
<td>[0.879]</td>
<td>[0.598]</td>
<td>[0.911]</td>
</tr>
<tr>
<td>Years of computer use</td>
<td>0.108**</td>
<td>0.014</td>
<td>0.011</td>
<td>0.037</td>
<td>-0.003</td>
<td>0.004</td>
<td>-0.046*</td>
</tr>
<tr>
<td></td>
<td>[0.036]</td>
<td>[0.715]</td>
<td>[0.779]</td>
<td>[0.419]</td>
<td>[0.934]</td>
<td>[0.918]</td>
<td>[0.059]</td>
</tr>
<tr>
<td>Household owns computers (binary)</td>
<td>0.816*</td>
<td>-0.215</td>
<td>-0.265</td>
<td>-0.271</td>
<td>0.126</td>
<td>0.168</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>[0.067]</td>
<td>[0.544]</td>
<td>[0.453]</td>
<td>[0.444]</td>
<td>[0.694]</td>
<td>[0.612]</td>
<td>[0.824]</td>
</tr>
<tr>
<td>Constant</td>
<td>0.134</td>
<td>-5.206**</td>
<td>1.556</td>
<td>1.237</td>
<td>1.303</td>
<td>-2.467</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>[0.969]</td>
<td>[0.020]</td>
<td>[0.427]</td>
<td>[0.609]</td>
<td>[0.505]</td>
<td>[0.218]</td>
<td>[0.924]</td>
</tr>
<tr>
<td>Observations</td>
<td>129</td>
<td>128</td>
<td>128</td>
<td>128</td>
<td>129</td>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.132</td>
<td>0.126</td>
<td>0.070</td>
<td>0.132</td>
<td>0.068</td>
<td>0.117</td>
<td>0.107</td>
</tr>
</tbody>
</table>
Appendix E: Subgroup analysis

The subgroup analyses below examine whether the effect of the Kepler curriculum on test scores differed by student gender or urban/rural domicile. Both the three-year cumulative impact and the third year impact for the 2013 cohort were estimated for these subgroups. Due to small sample sizes, these subgroup analyses only have sufficient statistical power\(^\text{12}\) to detect very large differences between subgroups; smaller differences may be statistically indistinguishable from zero.

Gender

In both Kepler and comparison groups, males outperformed females on nearly all exams. Over the course of the evaluation, the gender gap widened for Kepler females compared to females in other universities in critical thinking as measured by the Watson-Glaser critical thinking exam. The female comparison students improved critical thinking scores in the third year, while female Kepler students did not show much improvement. However, other tests do not provide much evidence for Kepler either narrowing or widening the pre-existing gender gap.

Table 5: Subgroup analysis by gender

<table>
<thead>
<tr>
<th>Test</th>
<th>Effect size (SD)</th>
<th>Perc. pt. diff-in-diff</th>
<th>p-value</th>
<th>Third year impact</th>
<th>Effect size (SD)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Watson-Glaser</td>
<td>-0.744</td>
<td>-6.4 pp</td>
<td>0.114</td>
<td></td>
<td>-1.143**</td>
<td>0.024</td>
</tr>
<tr>
<td>CLA+</td>
<td>-0.283</td>
<td>-31.259 points</td>
<td>0.517</td>
<td></td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Cognitive skills – English</td>
<td>-0.752*</td>
<td>-14.3 pp</td>
<td>0.055</td>
<td></td>
<td>-0.553</td>
<td>0.182</td>
</tr>
<tr>
<td>Cognitive skills – math</td>
<td>-0.496</td>
<td>-7.1 pp</td>
<td>0.215</td>
<td></td>
<td>-0.271</td>
<td>0.503</td>
</tr>
<tr>
<td>Cognitive skills – logic</td>
<td>-0.538</td>
<td>-10.1 pp</td>
<td>0.199</td>
<td></td>
<td>-0.504</td>
<td>0.316</td>
</tr>
<tr>
<td>IELTS writing</td>
<td>0.181</td>
<td>1.4 pp</td>
<td>0.625</td>
<td></td>
<td>-0.053</td>
<td>0.896</td>
</tr>
<tr>
<td>IELTS reading</td>
<td>-0.399</td>
<td>-4.2 pp</td>
<td>0.387</td>
<td></td>
<td>-0.202</td>
<td>0.577</td>
</tr>
<tr>
<td>Computer literacy</td>
<td>0.137</td>
<td>1.8 pp</td>
<td>0.689</td>
<td></td>
<td>0.036</td>
<td>0.892</td>
</tr>
</tbody>
</table>

Domicile

Endline scores did not differ systematically across all tests for students from urban and rural backgrounds for either Kepler or comparison students. In the results below, a positive difference indicates that there was a greater impact for rural students while a negative difference indicates a greater impact for urban students. Overall, the results from the subgroup analyses do not provide much evidence for Kepler having a greater impact on either urban or rural students.

\(^{12}\) Statistical power is the probability that a study will detect an effect when there is truly an effect present. Statistical power can be increased by having a larger sample size in each group that is being compared. The subgroup analyses presented here in general have less statistical power than the main analyses because more groups are being compared, and so there are fewer students (data points) per analysis.
Table 6: Subgroup analysis by urban/rural domicile

<table>
<thead>
<tr>
<th>Test</th>
<th>Three-year cumulative impact</th>
<th>Third year impact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effect size (SD)</td>
<td>Perc. pt. diff-in-diff</td>
</tr>
<tr>
<td>Watson-Glaser</td>
<td>0.579</td>
<td>4.9 pp</td>
</tr>
<tr>
<td>CLA+</td>
<td>-0.054</td>
<td>-5.946 points</td>
</tr>
<tr>
<td>Cognitive skills – English</td>
<td>-0.816**</td>
<td>-15.5 pp</td>
</tr>
<tr>
<td>Cognitive skills – math</td>
<td>-0.010</td>
<td>-0.1 pp</td>
</tr>
<tr>
<td>Cognitive skills – logic</td>
<td>0.210</td>
<td>4.0 pp</td>
</tr>
<tr>
<td>IELTS writing</td>
<td>0.071</td>
<td>0.6 pp</td>
</tr>
<tr>
<td>IELTS reading</td>
<td>0.277</td>
<td>2.9 pp</td>
</tr>
<tr>
<td>Computer literacy</td>
<td>0.059</td>
<td>0.8 pp</td>
</tr>
</tbody>
</table>
Appendix F: Example test questions

Watson-Glaser II

Directions: Inferences

An inference is a conclusion a person can draw from certain observed or supposed facts. For example, if the lights are on in a house and voices can be heard coming from the house, a person might infer that someone is at home. But this inference may or may not be correct. Possibly the people in the house did not turn the lights and the television off when they left the house.

In this test, each exercise begins with a statement of facts that you are to regard as true. After each statement of facts you will find several possible inferences—that is, conclusions that some persons might draw from the stated facts. Examine each inference separately, and make a decision as to its degree of truth or falsity.

For each inference you will find spaces on the answer sheet labeled T, PT, ID, PF, and F. For each inference make a mark on the answer sheet under the appropriate heading as follows:

T if you think the inference is definitely TRUE; that it properly follows beyond a reasonable doubt from the statement of facts given.

PT if, in the light of facts given, you think the inference is PROBABLY TRUE; that it is more likely to be true than false.

ID if you decide that there are INSUFFICIENT DATA; that you cannot tell from the facts given whether the inference is likely to be true or false; if the facts provide no basis for judging one way or the other.

PF if, in the light of the facts given, you think the inference is PROBABLY FALSE; that it is more likely to be false than true.

F if you think the inference is definitely FALSE; that it is wrong, either because it misinterprets the facts given, or because it contradicts the facts or necessary inferences from those facts.

Sometimes, in deciding whether an inference is probably true or probably false, you will have to use certain commonly accepted knowledge or information that practically every person has. This will be illustrated in the example that follows.

Example and answers

Two hundred students in their early teens voluntarily attended a recent weekend student conference in a Midwestern city. At this conference, the topics of race relations and means of achieving lasting world peace were discussed, because these were the problems the students selected as being most vital in today’s world.

1. Statement: As a group, the students who attended this conference showed a keener interest in broad social problems than do most other students in their early teens.

   Answer: Inference 1 is probably true (PT) because (as is common knowledge) most people in their early teens do not show so much serious concern with broad social problems. It cannot be considered definitely true from the facts given because these facts do not tell how much concern other young teenagers may have. It is also possible that some of the students volunteered to attend mainly because they wanted a weekend outing.

2. Statement: The majority of the students had not previously discussed the conference topics in their schools.
Answer: Inference 2 is probably false (PF) because the students’ growing awareness of these topics probably stemmed at least in part from discussions with teachers and classmates.

3. **Statement: The students came from all sections of the country.**
   
   **Answer:** There is no evidence for inference 3. Thus there are insufficient data (ID) for making a judgment on the matter.

4. **Statement: The students discussed mainly labor relations problems.**
   
   **Answer:** Inference 4 is definitely false (F) because it is given in the statement of facts that the topics of race relations and means of achieving world peace were the problems chosen for discussion.

5. **Statement: Some teenage students felt it worthwhile to discuss problems of race relations and ways of achieving world peace.**
   
   **Answer:** Inference 5 necessarily follows from the given facts; it therefore is true (T).

**CLA+**

**PERFORMANCE TASK**

This is an example of a brief Performance Task. In the course of this practice performance task, you will prepare a written response to a hypothetical but realistic situation. The Performance Task is made up of an introductory scenario, a question, and some documents that include several information sources. You will use information from the documents in carrying out the task.

While your personal values and experiences are important, you should base your response on the evidence provided in these documents.

**Role**

You are a staff member for an organization that analyzes the accuracy of policy claims made by political candidates. The organization is non-partisan, meaning that it is not influenced by, affiliated with or supportive of any one political party or candidate.

**Scenario**

Leila Jainson is running for reelection as the mayor of Stoneville. Mayor Jainson’s opponent in this contest is Dr. Carl Greer. Dr. Greer is a member of the Stoneville City Council. During a recent TV interview about cell phone use, Dr. Greer claimed that these phones interfered with people’s ability to operate a motorized vehicle and caused vehicle-related accidents in Stoneville. Dr. Greer said that reducing cell phone usage while driving motorized vehicles would lower the city’s vehicle-related accident rate. To support this argument, Dr. Greer presented a chart that compared the percentage of drivers who use cell phones while driving to the number of vehicle-related accidents. Dr. Greer based this chart on cell phone use and community data tables that were provided by the Stoneville Police Department and government population counts.

**Task**

Your job is to evaluate Dr. Greer’s claims. To do so, please answer the question that follows, using the supporting documents provided (labeled A and B). Your answers should include the appropriate or relevant evidence (drawn from documents A and B) necessary to support your position.

**Question**

Dr. Greer claims that “reducing cell phone usage while driving motorized vehicles would lower the city’s vehicle-related accident rate” (Document B exhibits the chart Dr. Greer used to support this statement).
1. What are the strengths and/or limitations of Dr. Greer’s position on this matter? What specific information in Documents A and B led you to this conclusion? What additional information, if any, would you like to have had?

Document A

STONEVILLE POLICE: SERVING STONEVILLE FOR OVER 50 YEARS

The two tables below present data about the city’s five regions. The percentage of registered drivers who use cell phones while operating a motorized vehicle (Table 1) was obtained from a population survey. The middle column of Table 1 shows the number of registered drivers involved in a motorized vehicle-related accident. The number of registered drivers (Table 1) and the percentage who are college graduates (Table 2) are based on 2005 government population counts. The percentage of moving violation offenders in Stoneville (Table 2) is based on 2005 Stoneville Police Department data.

Table 1: Vehicular accidents statistics

<table>
<thead>
<tr>
<th>Region</th>
<th>Percentage of drivers who use cell phones while operating a motorized vehicle</th>
<th>Number of registered drivers involved in a vehicle-related incident</th>
<th>Number of registered drivers</th>
<th>Number of vehicle-related accidents per 1,000 drivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>1</td>
<td>72</td>
<td>8,396</td>
<td>8.58</td>
</tr>
<tr>
<td>South</td>
<td>3</td>
<td>110</td>
<td>13,099</td>
<td>8.40</td>
</tr>
<tr>
<td>North</td>
<td>5</td>
<td>171</td>
<td>18,886</td>
<td>9.05</td>
</tr>
<tr>
<td>West</td>
<td>8</td>
<td>204</td>
<td>23,993</td>
<td>8.50</td>
</tr>
<tr>
<td>City Center</td>
<td>10</td>
<td>222</td>
<td>25,875</td>
<td>8.58</td>
</tr>
</tbody>
</table>

Table 2: Demographic characteristics

<table>
<thead>
<tr>
<th>Region</th>
<th>Percentage of moving violation offenders in Stoneville</th>
<th>Percentage of drivers who are college graduates</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>60</td>
<td>22</td>
</tr>
<tr>
<td>South</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>North</td>
<td>40</td>
<td>11</td>
</tr>
<tr>
<td>West</td>
<td>35</td>
<td>9</td>
</tr>
<tr>
<td>City Center</td>
<td>45</td>
<td>3</td>
</tr>
</tbody>
</table>

Document B

DR. GREER’S CHART
Dr. Greer used the chart below during a TV interview to show the relationship between the number of people who use cell phones while driving a motorized vehicle and vehicular accidents in Stoneville. This chart is based on data that were provided to Dr. Greer by the Stoneville Police Department (Document A).

![Motorized Vehicle Accident and Cell Phone Use in Stoneville](chart)

**SELECTED-RESPONSE QUESTIONS**

**Scenario**

Are grades in college and overall productivity in college negatively affected by student’s use of Facebook? A group of researchers hypothesized that students who use Facebook on a regular basis do get lower grades in college than students who do not regularly use Facebook.

In order to test this hypothesis, researchers collected data from students at a large university. The researchers stood on a popular corner of campus, the corner where popular restaurants and bars are located, and asked 50 students who were walking by to answer a few questions.

Researchers found that 79% of the students said they did not think that spending time on Facebook interferes with their grades. The researchers decided to compare the average amount of time that students spend on Facebook each week with each student’s current GPA. The results are shown in Figure 1.

**Example question**

1. Which of the following, if true, would strengthen the researchers’ claim that spending more time on Facebook leads to a decrease in a college student’s GPA?

   A. Results of a study showing that Frequent Users spend their free time in bars while Infrequent Users spend their free time in libraries.
   
   B. Results of a study showing that there is no relationship between Frequent Users and a dislike of studying.
   
   C. Results of a study showing that Frequent Users play video games when they don’t have access to Facebook.
   
   D. Results of a study showing that people who do poorly in high school, tend to not do well in college, either.
Cognitive skills

English language

1. Choose the word that is grammatically correct:
   They said that _________ on the way to the movie now.
   - their
   - there
   - them
   - they're

2. Which word will not correctly complete this sentence:
   Some children cannot ______ restraint when consuming candy.
   - provide
   - demonstrate
   - show
   - exhibit

3. Constant | intermittent: these two words have:
   - Similar meanings
   - Contradictory meanings
   - Neither the same nor opposite meanings

Math

4. If $x = 0.3$, which of the following represents the largest number?
   - $3x$
   - $0.80x$
   - $x^2$
   - $\frac{4}{x}$
   - $4x$

5. If Matthew is 32 and his sister is three-quarters his age, how old will Matthew’s sister be when Matthew is 44 years old?
   - 20
   - 28
   - 33
   - 36
6. It’s currently 11:25AM and I have a meeting at 1:10PM at my office. If it takes me 50 minutes to travel, which option is the latest I can leave to arrive to the meeting on time?
   - Right now
   - 50 minutes from now
   - 60 minutes from now
   - 105 minutes from now

Logic
7. Assume the first two statements are true.

(1) Tanya is older than Eric
(2) Cliff is older than Tanya
(3) Eric is older than Cliff

If the first two statements are true, the third statement is:
   - True
   - False
   - Uncertain, not enough information

8. Which person would have the fastest average daily speed (assuming no other movement during the day)?
   - Person A rode a bicycle for 3 hours
   - Person B walked for 8 hours
   - Person C rode on a speed train for 2 hours
   - Person D drove a car for 30 minutes

9. The father of my aunt’s brother is my:
   - Cousin
   - Uncle
   - Father
   - Grandfather

IELTS Reading
Do the following statements agree with the information given in Reading Passage 1?
   - TRUE if the statement agrees with the information
   - FALSE if the statement contradicts the information
1. Chronobiology is the study of how living things have evolved over time.
   - True
   - False
   - Not given

2. The rise and fall of sea levels affects how sea creatures behave.
   - True
   - False
   - Not given

Choose the correct answer.

3. What did researchers identify as the ideal time to wake up in the morning?
   - 6.04
   - 7.00
   - 7.22
   - 7.30

The text has eight paragraphs, A–H.
Which paragraph contains the following information?
Write the correct letter, A–H, at the bottom of each question.

4. The origin of the word ‘calisthenics’

Complete the summary below.
Choose NO MORE THAN TWO WORDS from the text for each answer.
Write your answers below.

During the sixties and seventies, attaining huge muscles became more important than

5 …………………… or having an attractive-looking body.
IELTS Writing

Write about the following topic:

Some people believe that teaching children at home is best for a child's development while others think that it is important for children to go to school.

Discuss the advantages of both methods and give your own opinion.

Give reasons for your answer and include any relevant examples from your own knowledge or experience. Write at least 250 words.

Computer literacy

Typing speed

Students were instructed to complete a three-minute typing tests at http://www.typingtest.com.

Adjusted typing speed in words per minute (wpm) was calculated by:

1. Dividing the number of characters typed by five to get number of words typed
2. Subtracting the number of errors made as a penalty
3. Dividing by three to get words per minute

Web research

Use the internet to find answers to the following questions. Write the answers in the spaces below. You will be graded on the accuracy of your answer.

1. Who won the silver medal for Women’s Singles tennis in the 1992 Olympics?

Website credibility

Determine if the sites would be considered a credible website.

Credible means that it is likely to have reliable information that you could cite in an academic paper.

<table>
<thead>
<tr>
<th>#</th>
<th>Website</th>
<th>Credible? Circle answer.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wikipedia</td>
<td>YES</td>
</tr>
<tr>
<td>2</td>
<td>National Institute of Statistics and Research, Rwanda</td>
<td>YES</td>
</tr>
</tbody>
</table>

Microsoft Word

Students were instructed to open a prepared Microsoft Word document and make edits such as the following:

- Change the font (letter style) of the document to Times New Roman.
- Insert a header with the words “Breaking News”
- Change the margins of the document to 0.75 inches on all sides.

Microsoft Excel

Students were instructed to open a prepared Microsoft Excel spreadsheet and make edits such as the following:

- Adjust the column titles so that all words appear in one cell using wrap text.
- Change the top row’s background color to green.
• Make a bar graph (column) with columns **region** and **total number of households**

**Email**

Students were given the following instructions to send an email to a given email address:

• Write a formal email, stating that you are submitting the computer test documents, attaching the Word and Excel documents.

Students without an email account were given a Gmail email account and password to use.