Estudio de caso

Duolingo Efectiveness Study (2012)

ROUMEN VESSELINOV (CUNY)
JOHN GREGO (USC)

EXECUTIVE SUMMARY

The research study of Duolingo effectiveness was independently conducted in September-November of 2012. The study lasted for approximately eight weeks. A random representative sample was selected from Duolingo users who studied Spanish. The participants were at least 18 years of age, native speakers of English, not from Hispanic origin and not advanced users of Spanish, and all of the participants resided in the United States

The participants took one college placement Spanish language test in the beginning of the study and one test at the end of the study. The test results were measured in points (the higher the better). The improvement of language abilities was measured as the difference between the final and the initial language test results. The effectiveness of Duolingo was measured as language improvement per one hour of study.

MAIN RESULTS

- Overall the average improvement in language abilities was 91.4 points and the improvement was statistically significant.
- The effectiveness measure showed that on average participants gained 8.1 points per one hour of study with Duolingo.
- The 95% Confidence Interval for the effectiveness is from 5.6 points to 10.7 points gained per one hour of study.

Table 1. Spanish WebCAPE Test Cut-off Points

Points	College Semester Placement
Below 270	Semester 1
270-345	Semester 2
346-428	Semester 3
Above 428	Semester 4+

The measure of Effectiveness for this study was defined as follows:

$$Effectiveness = \frac{\text{Effect}}{\text{Efforts}} = \frac{\text{Improvement of language skills}}{\text{Study time}} = \frac{\text{Final-Initial test score}}{\text{Hours of study}}$$

Figure A1. Sample Selection Tree

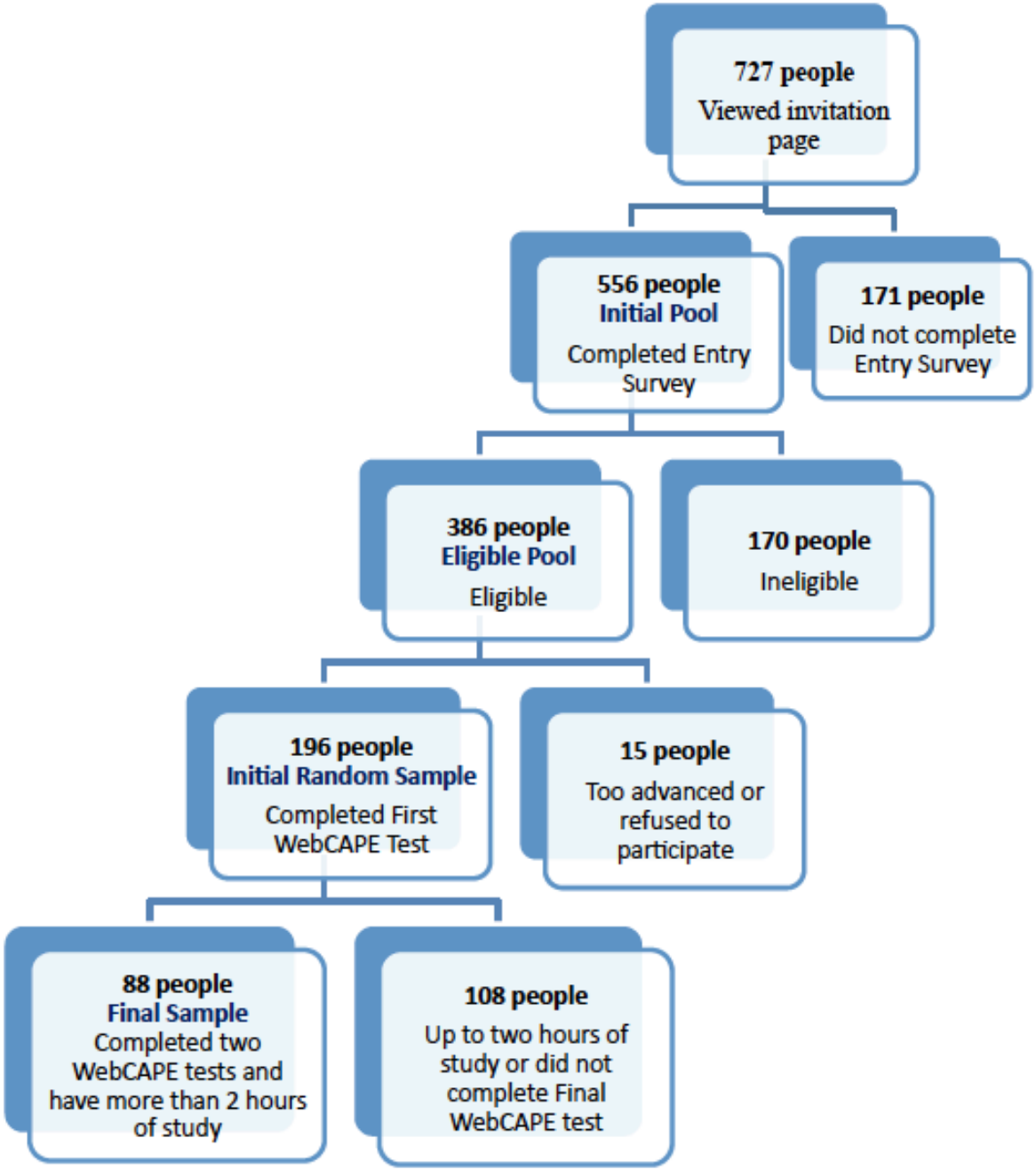


Table 5. Final Study Sample: Age and Gender Decomposition

Age	Female (N)	Male (N)	Total (N)	Percent
18 to 20 years old	0	5	5	5.7
21-30 years old	23	13	36	40.9
31-40 years old	7	14	21	23.9
Over 40 years old	14	12	26	29.5
Total	**44**	**44**	**88**	**100.0**

Tale 7. Study Time

Statistics	Hours of Study
Mean (std)	22.0 (20.4)
Median	16.9
Min	2
Max	133
N	88

Figure 1. Study Time Distribution

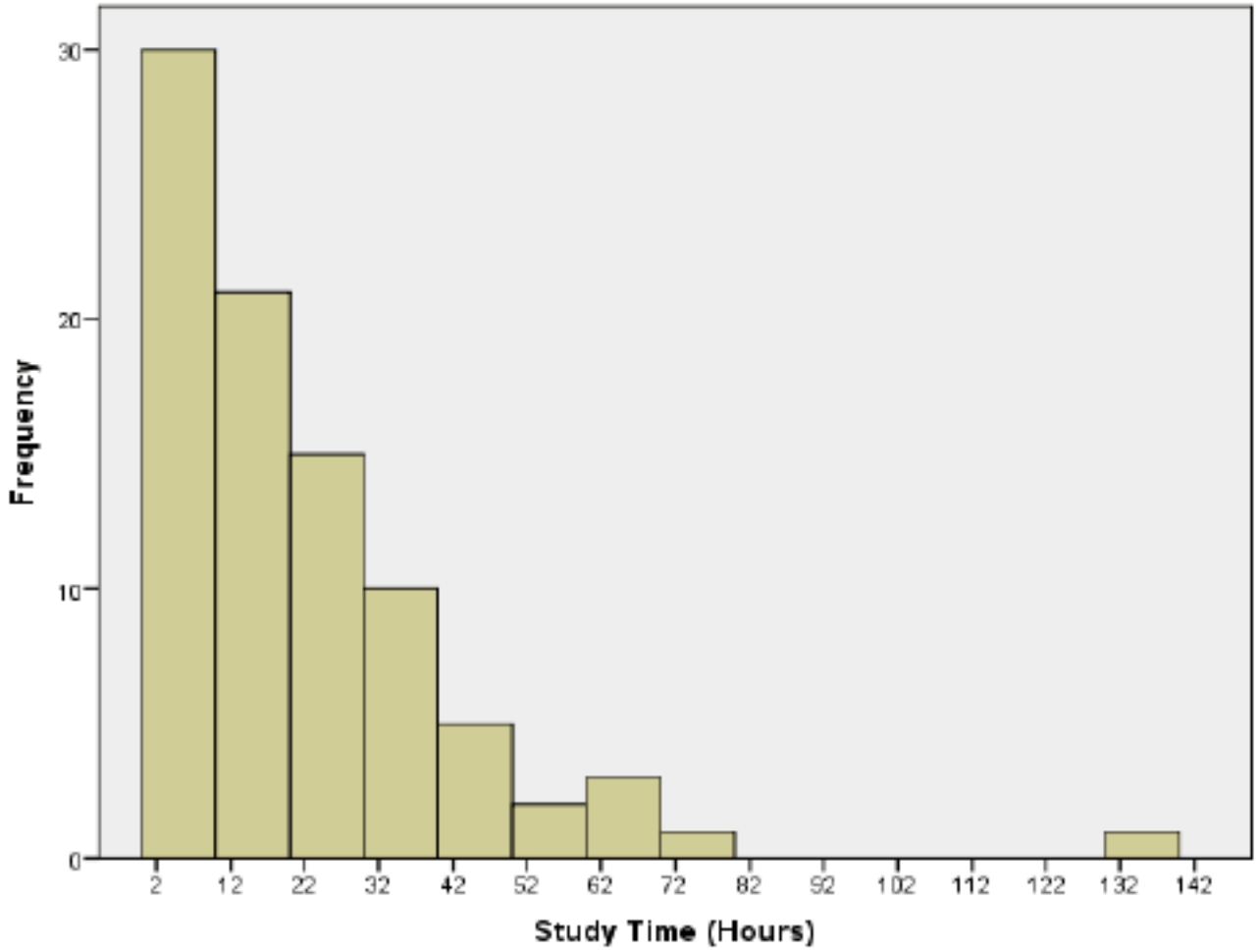


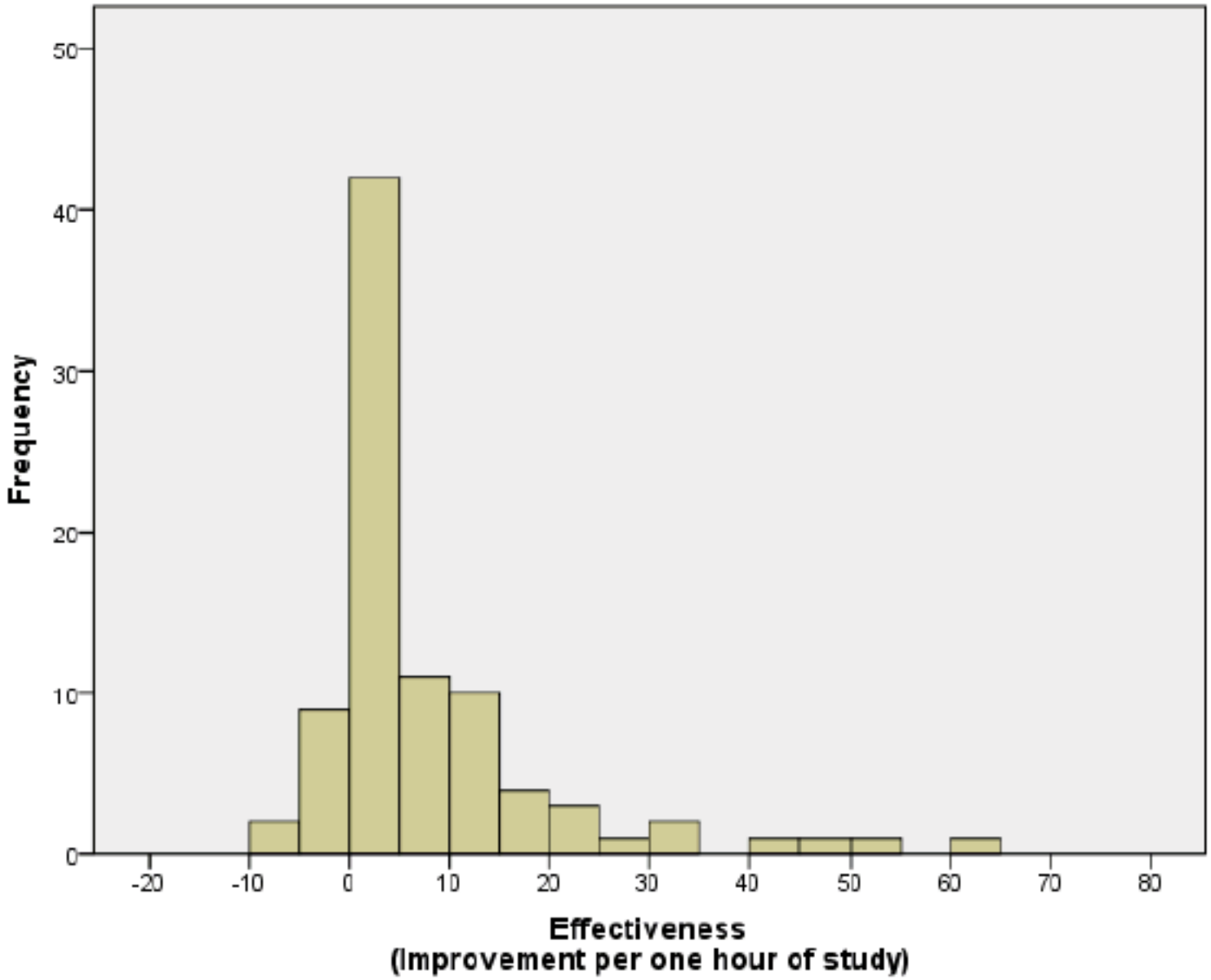
Table 8. Language Improvement

Statistics	Initial WebCAPE	Final WebCAPE	Improvement (Final-Initial)
Mean (std)	162.5 (116.5)	253.9 (110.5)	91.4 (88.0)
Median	161	262.5	81.5
Min	0	0	-57
Max	405	539	341
N	88	88	88

Table 9. WebCAPE Semester Placement

College Semester	Initial Test	Final Test
	Percent (Number)	Percent (Number)
First	77.3 (68)	52.3 (46)
Second	14.8 (13)	29.5 (26)
Third	8.0 (7)	14.8 (13)
Fourth+		3.4 (3)
Total	**100 (88)**	**100 (88)**

Figure 2. Effectiveness Distribution



¿Esto es todo?

Volvamos para atrás…

Figure A1. Sample Selection Tree

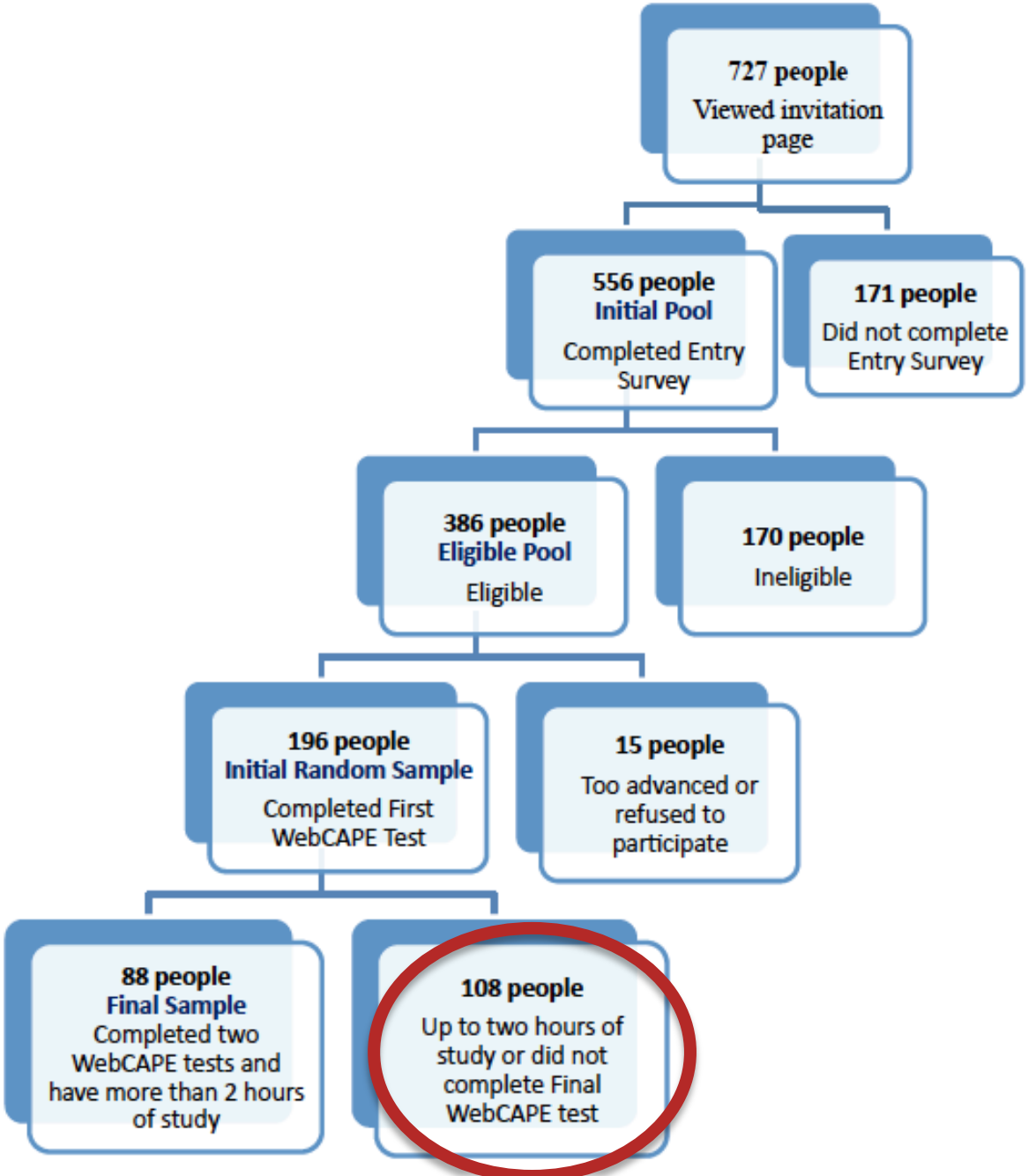


Table 5. Final Study Sample: Age and Gender Decomposition

Age	Female (N)	Male (N)	Total (N)	Percent
18 to 20 years old	0	5	5	5.7
21-30 years old	23	13	36	40.9
31-40 years old	7	14	21	23.9
Over 40 years old	14	12	26	29.5
Total	**44**	**44**	**88**	**100.0**

Table 8. Language Improvement

Statistics	Initial WebCAPE	Final WebCAPE	Improvement (Final-Initial)
Mean (std)	162.5 (116.5)	253.9 (110.5)	91.4 (88.0)
Median	161	262.5	81.5
Min	0	0	-57
Max	405	539	341
N	88	88	88

¿Y qué sabemos del WebCAPE?

WebCAPE Test Description

WebCAPE

The test is multiple-choice and covers grammar, reading, and vocabulary. It is an adaptive test, which means that the question difficulty will vary depending on how you respond. Do not get discouraged if you are unfamiliar with some of the material. Make your best guess and go on to the next question.

¿Y qué pasa con la generalizabilidad?

- Efectividad del tratamiento (Duolingo) medido dos veces sobre los mismos items. ¿Posibles efectos de repetición? ¿Práctica del examen?

¿Y la estadística?

- Autores aplican t-test de muestras pareadas.
- Pero… omiten algunos detalles. ¿Cumple la muestra con la asunción de normalidad de la distribución?
- Analizan relación entre horas de estudio y efectividad con… ¡una t-test! Lo ideal, una correlación de Spearman (dos variables de tipo razón).

Conclusiones

Está la estadística….