Automated formative evaluations for reading comprehension in an English as a foreign language course: benefits on performance, user satisfaction, and monitoring of higher education students in Chile

Santos Lazzeri[1], Ximena Cabezas[2], Luis Ojeda[3], and Francisca Leiva[4]

Abstract. We assess the effect of automated formative evaluations on reading comprehension skills in a course of English for Specific Purposes (ESP) in the area of kinesiology at the Universidad Austral de Chile – Valdivia (UACh). The evaluations were implemented using Questionmark's Perception (QMP) (Questionmark-Corporation, 2015). We investigate: (1) Do formative reading comprehension assessments enhance students' reading comprehension skills? (2) How do students perceive QMP? The experimental design used for this study was pre-test/post-test with control group. The participants were 57 freshmen, kinesiology students from UACh, randomly divided into two groups: G1-experimental, G2-control. After the pre-test, G1 worked on 11 online reading comprehension modules, which included formative evaluations with automated immediate feedback, while G2 did the same work with printed materials. At the end, both groups took the same post-test. The results show that there were no statistically significant differences between the mean grade differences (post-test grade - pre-test grade) of G1 and G2. G1's surveys showed positive attitudes towards the use of automated formative evaluations. Our conclusions are that for our population, the use of computer technology was at least as effective as instruction without technology. Furthermore, QMP was satisfactorily

1. Computer Systems Consultant - Universidad Austral de Chile, Chile; sglazzeri@yahoo.com
2. Facultad de Medicina - Instituto de Medicina - Universidad Austral de Chile, Chile; xcabezas@uach.cl
3. Facultad de Ciencias Económico-Administrativas Instituto de Estadística - Universidad Austral de Chile, Chile
4. Facultad de Medicina - Escuela de Enfermería - Universidad Austral de Chile, Chile

How to cite this article: Lazzeri, S., Cabezas, X., Ojeda, L., & Leiva, F. (2015). Automated formative evaluations for reading comprehension in an English as a foreign language course: benefits on performance, user satisfaction, and monitoring of higher education students in Chile. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 355-361). Dublin: Research-publishing.net. http://dx.doi.org/10.14705/rpnet.2015.000358

evaluated by the students, and it allowed the professor to monitor and timely detect students with performance problems thanks to the different reports it provides.

Keywords: formative evaluations, English as a foreign language, online reading comprehension, English for specific purposes.

1. Introduction

Formative assessment has long been believed to be effective, as documented in Black and Wiliam (1998). However, recent studies, like Kingston and Nash (2011), challenge that belief. This controversy, along with the introduction of multiple technological tools that implement evaluations, motivated us to test one such tool: QMP (Questionmark-Corporation, 2015), applied to reading comprehension in ESP. From a curricular viewpoint, our institution currently applies a competencies-based model (Jabif, 2007; UACh, 2007), which adopts a holistic, integrating vision, with methodologies that are based more on student learning than on the professor's teaching. Therefore, the use of computer based formative assessments is consistent with UACh's policies.

The process we followed to select QMP and descriptive statistics of the study are presented in Lazzeri et al. (2015). Here we present a more comprehensive analysis of our population composition and a performance comparison considering the entrance skill level of the students. We also study the students' preferences according to their survey answers, and consider other advantages that a Computer Based Assessment (CBA) software, such as QMP, can offer.

Our research questions are: (1) Do formative reading comprehension assessments enhance students' reading comprehension skills? (2) How do students perceive QMP?

2. Method

2.1. Experimental design

Study Type: Pre-test/post-test with control group.

Population: 57 freshmen kinesiology students from UACh; Age: average=19, *SD*=2.5; Gender: 51% male, 49% female; High-school type: 38% public, 56% subsidized private, 6% private. Only 19% had a CEFR certification at the ALTE

A2 or B1 levels, which are the goal levels specified by the Chilean government for elementary and high-school graduates, respectively.

Our population was randomly divided into two groups: G1-experimental, G2-control. After the pre-test, G1 worked on seven lessons that were implemented as 11 online reading comprehension modules, which included formative evaluations with automated immediate feedback, while G2 did the same work with printed materials. The only difference was the presentation mode of the material and the automated feedback. Data was collected from a pre-test, post-test, and survey application.

2.2. Instruments

The pre-test and post-test were developed and graded by the course's instructor. The students' satisfaction survey was developed by data analysis specialists. The lessons, used as learning measurement instruments, were designed by the course instructor and implemented by the software administrator in QMP automated formative assessment modules with immediate feedback. Each lesson contained a paper/reading in the field of kinesiology and several exercises related to that reading. Table 1 shows the composition of each lesson in terms of length of paper in words and types of exercises used.

Table 1. Lesson composition

Lesson	**Paper Word Count**	**Number of exercises for each type of question**								**Total**
		Essay	Multiple Choice	Column Match	Cloze	Brief Text	Survey Matrix	T/F	Essay with Extra Text	
1	1167		2	17	10			10		39
2	3047		20	10	10	45				85
3	5363	17		20	10	15		10		72
4	4111			10	10	10		10	10	50
5	4480			10		30		10	10	60
6	5116	10	10	10	10	30	10	10	20	110
7	5504	3		10	20	50		20		103
Total	28788	30	32	87	70	180	10	70	40	519

To determine improvement in reading comprehension skills, we used the dependent variable "Academic Performance on Reading Comprehension" as measured by the grade obtained in the exams, given in a 1-7 scale, where 7 is best, which is the

standard in the Chilean educational system. More precisely, we used the difference in academic performance between the post-test and the pre-test results. G1 and G2 were compared in terms of this variable using Student's t-test for independent samples, since the preconditions to use this kind of test were satisfied. The independent variable was "use of QMP" (Yes/No). The statistical analysis was carried out with SPSS 11.5.

The students' satisfaction survey contained 8 Likert-style questions directly related to the use of QMP that allowed us to get the students' perceptions about the platform. The internal consistency of this survey was determined by computing the Alpha Cronbach reliability coefficient as .747, which is deemed acceptable.

3. Discussion

Table 2 summarizes the results of both groups in the pre-test and post-test.

Table 2. Pre-test and post-test results

Group	Post-test		Pre-test	
	n	Mean	Sdev	Mean	Sdev
G1	28	5.60	0.97	2.96	0.92
G2	29	5.51	0.76	3.24	0.69

It is important to notice that G1 had on average lower scores in the pre-test than G2, but after completing the lessons using the QMP modules, they got a higher average than G2 (using printed materials) in the post-test. Nevertheless, there were no statistically significant differences between the mean grade differences with 95% confidence (t=1.41, p=0.16>0.05) as shown in Table 3.

Table 3. Statistical comparison

	Group	N	Media	Standard Deviation	Levene'sTest for Variance Equality		t-test for media equality	
					F	Sig.	t	Sig.
Difference (Post-test – Pre-test)	G1	28	2.64	0.95				
	G2	29	2.27	1.00	0.90	0.79	1.41	0.16

The answers to the questions related to the students' perceptions about QMP for G1 are summarized in Table 4.

Table 4. Survey results

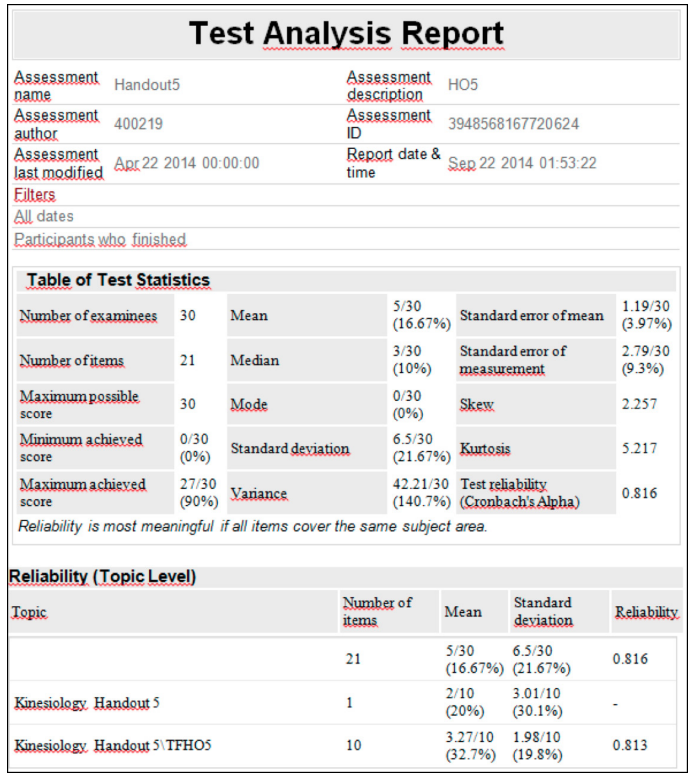
Students' Perception about the use of QMP	Strongly Disagree		Disagree		Neutral		Agree		Strongly Agree	
	n	%	n	%	n	%	n	%	n	%
I liked learning using the methodology based on QMP	1	3.6	1	3.6	8	28.6	13	46	5	17.9
Seeing my classmates' progress motivates me to work	0	0	3	10.7	9	32.1	14	50	2	7.1
At the end of each session I feel that I have learned	0	0	3	10.7	14	50	7	25	4	14.3
The automatic feedback from the platform helps my learning	1	3.6	0	0	8	28.6	7	25	12	42.9
Using QMP made me feel more confident about my knowledge	1	3.6	6	21.4	10	35.7	7	25	4	14.3
I like to have the control over my learning process	0	0	0	0	4	14.3	10	36	14	50
I used the trial and error method as a source of learning.	1	3.6	0	0	4	14.3	19	68	4	14.3
The platform QMP met my expectations.	0	0	5	17.9	10	35.7	11	39	2	7.1

We can highlight that 64% liked the QMP-based methodology, 86% enjoyed controlling their learning process, and 82% used trial and error as a learning strategy. These percentages are obtained by adding the "Agree", and "Strongly Agree" answers for each question. Furthermore, in a separate question, 89.3% recommended their peers to volunteer for the QMP evaluation process. Another positive aspect of QMP is its functionalities to generate useful reports, such as the Test Analysis Report, partially shown in Figure 1, which shows diverse test statistics and a reliability analysis.

4. Conclusions

Despite not finding statistically significant performance improvements, we can conclude that, for our population, the use of computer technology was at least as effective as instruction without technology, which coincides with some of the findings in Grgurović, Chapelle, and Shelley (2013). Furthermore, QMP was satisfactorily evaluated by the students. QMP also allowed the professor to monitor and timely detect students with performance problems thanks to the different reports it provides, which offered relevant information such as students' performance for each exercise and formative evaluation, and items that proved to be easiest or most difficult, among others.

Figure 1. QMP's test analysis report[5]



Test Analysis Report

Assessment name	Handout5	Assessment description	HO5
Assessment author	400219	Assessment ID	3948568167720624
Assessment last modified	Apr 22 2014 00:00:00	Report date & time	Sep 22 2014 01:53:22

Filters
All dates
Participants who finished

Table of Test Statistics

Number of examinees	30	Mean	5/30 (16.67%)	Standard error of mean	1.19/30 (3.97%)
Number of items	21	Median	3/30 (10%)	Standard error of measurement	2.79/30 (9.3%)
Maximum possible score	30	Mode	0/30 (0%)	Skew	2.257
Minimum achieved score	0/30 (0%)	Standard deviation	6.5/30 (21.67%)	Kurtosis	5.217
Maximum achieved score	27/30 (90%)	Variance	42.21/30 (140.7%)	Test reliability (Cronbach's Alpha)	0.816

Reliability is most meaningful if all items cover the same subject area.

Reliability (Topic Level)

Topic	Number of items	Mean	Standard deviation	Reliability
	21	5/30 (16.67%)	6.5/30 (21.67%)	0.816
Kinesiology Handout 5	1	2/10 (20%)	3.01/10 (30.1%)	-
Kinesiology Handout 5\TFHO5	10	3.27/10 (32.7%)	1.98/10 (19.8%)	0.813

Since this is the first time that technology in the area of computer based formative evaluation has been introduced in the Faculty of Medicine at UACh in a course of ESP, we present a preliminary evaluation in this area showing that the use of technology contributes to learning in a different way, more compatible with today's demands from the digital world. However, we cannot generalize the results at this point.

5. Acknowledgements

We would like to thank the Dirección de Investigación y Desarrollo (DID) of the Universidad Austral de Chile – Valdivia for sponsoring this research in the context of project DID S- 2013-41. We would also like to thank the Centro de Idiomas at Universidad Austral de Chile – Valdivia for their assistance in the development of

5. Screenshot reproduced with kind permission from Questionmark™

this work. Finally, We would like to thank Questionmark Corporation for granting permission to include in this paper images generated with their software.

References

Black, P., & Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Grgurović, M., Chapelle, C., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL, 25*(2), 165-198. doi:10.1017/S0958344013000013

Jabif, L. (2007). *La docencia universitaria bajo un enfoque de competencias: orientaciones prácticas para docentes*. Universidad Austral de Chile.

Kingston, N., & Nash, B. (2011). Formative assessment: a meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28-37. doi:10.1111/j.1745-3992.2011.00220.x

Lazzeri, S., Cabezas, X., Ojeda, L., & Leiva, F. (2015). Assessing the impact of computer based formative evaluations in a course of english as a foreign language for undergraduate kinesiology students in chile. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds.), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 348-354). Dublin Ireland: Research-publishing.net. doi:10.14705/rpnet.2015.000357

Questionmark-Corporation. (2015). *Questionmark*. Retrieved from https://www.questionmark.com/us/Pages/default.aspx

UACh. (2007). *Comisión Curricular, Modelo Educacional y Enfoque Curricular de la Universidad Austral de Chile*. Editorial Universidad Austral de Chile.

Published by Research-publishing.net, not-for-profit association
Dublin, Ireland; info@research-publishing.net

Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy
Edited by Francesca Helm, Linda Bradley, Marta Guarda, and Sylvie Thouësny

Typeset by Research-publishing.net
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-28-5 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-29-2 (Ebook, PDF, colour)
ISBN13: 978-1-908416-30-8 (Ebook, EPUB, colour)

Legal deposit, Ireland: The National Library of Ireland, The Library of Trinity College, The Library of the University of Limerick, The Library of Dublin City University, The Library of NUI Cork, The Library of NUI Maynooth, The Library of University College Dublin, The Library of NUI Galway.

Legal deposit, United Kingdom: The British Library.
British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2015.