e-ISSN 2444-8729

https://doi.org/10.14201/eks2016174111128

One Picture or a Thousand Words? Influence of Question Length and Illustration Support on the Success and Skip Rates on Online Tests

¿Una imagen o mil palabras? Influencia de la extensión de las preguntas y el soporte gráfico en las tasas de éxito y abandono de las pruebas en línea

Ernest Redondo [1], Joaquim Regot [1], David Fonseca [2], Francesc Valls [1], Lluís Giménez [1]

[1] Escuela de Arquitectura de Barcelona (Universidad Politécnica de Cataluña). Spain. ernesto.{redondo, joquim.regot, francesc.valls, lluis.gimenez}@upc.edu

[2] Escuela de Arquitectura, La Salle Universitat Ramon Llull. Spain. fonsi@salleurl.edu

Abstract

The growing popularity of automatically graded online tests, either as an evaluation or self-assessment tool in online or blended education, demands a review of how these questions are designed and delivered to their intended audience. This paper analyzes the results of over 20,000 pre-university mock online quizzes designed to train the students for the Spanish university admission test (known as "Pruebas de Acceso a la Universidad" or "Selectividad") in the technical drawing subject, corresponding to the June and September intakes of 2009 and 2015. The influence of two key aspects on the questions success and skip rates is assessed: (a) the presence or absence of illustration support and (b) the length of the question as a proxy of reading comprehension difficulty. The results support that the presence of an accompanying illustration in the questions result in fewer skipped questions and mode successful answers, while the length of the question has the opposite effect. The performance difference in the 6-year span is also discussed, showing a slight decline over time in the pass rates while the skip rates remain stable. When comparing both two intakes, corresponding to different academic profiles of students that passed the June exam and those who did not, the success ratio is unsurprisingly lower for the students in the second intake. These findings should help improving the design of online quizzes, including more visual content and/or rephrasing the questions to be more concise, to fit the requirements of students educated in a more visual environment of multimedia technologies.

Resumen

La creciente popularidad de las pruebas en línea evaluadas automáticamente, tanto como herramientas de evaluación como de auto-aprendizaje en la educación a distancia o semipresencial, requiere una revisión de cómo estas preguntas se diseñan y presentan a sus receptores. Este artículo analiza los resultados de más de 20.000 exámenes simulados en línea diseñados para preparar el estudiantado para las pruebas de acceso a la universidad española (selectividad) en la materia de dibujo técnico, correspondientes a las convocatorias de junio y setiembre de 2009 y 2015. Se evalúa la influencia de dos aspectos clave en el número de preguntas respondidas correctamente o abandonadas: (a) la presencia o ausencia de una ilustración y (b) la extensión de la pregunta como indicador de la comprensión lectora. Los resultados apoyan que la presencia de una ilustración se traduce en menor número de preguntas abandonadas y mayor número de preguntas respondidas correctamente, mientras que la extensión de la pregunta tiene el efecto contrario. También se examinan los diferentes resultados durante el período de 6 años, que muestran un ligero empeoramiento de los resultados, mientras que la tasa de abandono se mantiene estable. Comparando ambas convocatorias, correspondientes a los perfiles académicos de los estudiantes que aprobaron en junio y los que no lo hicieron, la tasa de éxito es comprensiblemente menor para los estudiantes la segunda. Estos hallazgos deberían permitir mejorar el diseño de las pruebas en línea, incluyendo más contenido visual y/o haciendo las preguntas más concisas, para ajustarse a los requerimientos de unos estudiantes educados en el entorno visual de las tecnologías multimedia.

Keywords

Pre-University Education; Learning Analytics; Online Test; Reading Comprehension; Image Support; University Admission Test.

Palabras Clave

Educación Pre-Universitaria; Analíticas de Aprendizaje; Pruebas En-Línea; Comprensión Lectora; Soporte Gráfico; Prueba de Acceso a la Universidad.

Recepción: 15-12-2016 Revisión: 18-12-2016 Aceptación: 18-12-2016 Publicación: 31-12-2016

© Ediciones Universidad de Salamanca / CC BY NC-NB

1. Introduction

Tests are one of the most often used tools to grade students, and consist on one or more questions whose combined points define the grade the student will receive. Quizzes are standardized tests in which the questions offer a restricted set of possible answers, which allow easier –and often automated– grading.

In the European Higher Education Area, the Bologna process emphasizes the importance of self-learning; online quizzes can be a valuable educational tool, allowing students to self-assess their learning process anytime, anywhere and any number of times, receiving immediate feedback on their performance.

This paper analyzes over 20,000 responses of the preparatory online quiz for the Spanish University Admission Test (*Selectividad*) of the technical drawing subject (of which some or the authors are coordinators), which requires advanced spatial reasoning, and where the students must solve complex problems using abstraction and decomposition, adopting a computational thinking mindset (Wing, 2006). The quiz (Generalitat de Catalunya, 2009) is open to all pre-university students who want to take it online using a web browser in an Internet-connected device to self-assess their skills in different subjects.

The *Selectividad* is closely related to other admission tests to colleges and universities, such as SAT in the United States, A Level in the UK, *Abitur* in Germany or *Baccalauréat* in France. It includes some subjects common for all students as well as some specific subjects depending on their specific *Bachillerato* (high school), divided into 5 blocks: (a) arts, (b) humanities, (c) social sciences, (d) science and engineering, and (e) nature sciences. The subject of technical drawing is included into three of these blocks as an elective: arts, science and engineering, and nature sciences.

In this paper, the variables that increase the number of successful attempts and/or lower the number of questions skipped are explored, with the objective of obtaining insight on how to improve future iterations of the test. After the introductory section, the paper is structured into 4 other main sections:

- The source data used in the study is described in section 2 (experimental data).
- Data transformation processes are detailed in section 3 (data analysis).
- The obtained results are explained in section 4 (results).
- Finally, the results are interpreted and future research is discussed in section 5 (conclusions).

© Ediciones Universidad de Salamanca / CC BY NC-NB

2. Experimental Data

The analyzed experimental data consisted on the responses to an online multiple choice mock test developed to train the students for the technical drawing exam of the *Selectividad*. The data was collected in the years 2009 and 2015, and included the responses for the students preparing for the June and September intakes of the exam (Table 1).

	June 2009	**Sept. 2009**	**June 2015**	**Sept. 2015**
Sample Start	March 18th	June 12th	January 12th	June 12th
Sample End	June 11th	Sept. 9th	June 11th	Sept. 4th
Responses	12369	1295	6268	655

Table 1. Collected responses for each year and intake for the technical drawing online mock test

The total number of responses collected was 20,578, the majority of which were from year 2009 (about twice as many responses than on 2015). The responses corresponding to the students preparing for the June intake of the exam were about 90% of the total responses in both years, because the majority of students who take the September exam are the ones who failed the June exam and the majority of the students pass the exam in June (about 90%).

The data was provided in excel tables that summarized the number of students that passed, failed or ignored each of the 56 questions in the test, which were the same on both years. This made possible to compare the results of the same question in an interval 6 years apart, albeit with different participants who therefore did not constitute a cohort.

With summarized data, the participants could not be individualized, and it was not possible to perform any analysis comparing the performance on different questions at the subject level, because the link between the responses of each student was lost. The data was structured into files that contained tables inside of sheets:

- Each year of collected data (2009 or 2015) was stored in a different file.
- Each intake (June or September) was in a different sheet of those files.
- The sheets contained tables where each line had a question code along with the number of passed, failed and disregarded responses.

In addition, three variables were derived from the content of the questions, to assess their impact on the responses:

- Whether the question was only textual or included an illustration.
- Knowledge subdomain inside the subject of technical drawing.

3. Data Analysis

The data manipulation and analysis was performed in the Open Source Statistics Language R 3.3.1 (R Core Team, 2016) using the following packages: ggplot2 2.1.0 (Wickham, 2009), dplyr 0.5.0 (Wickham & Francois, 2016), stringr 1.1.0 (Wickham, 2016a), tidyr 0.6.0 (Wickham, 2016b), reshape2 1.4.1 (Wickham, 2007), and plyr 1.8.4 (Wickham, 2011). The combined length of the scripts was over 800 lines of code, and data was always manipulated from the sources through code following the principles of reproducible research (Peng, 2011; Wilson, et al., 2014).

3.1. Cleaning Data

Following the relational model defined by Codd (Codd, 1990), each observational unit was planned to be stored in a separate table, breaking the data into the following tables:

- Question responses (the original data).
- Question properties (data about the content of the questions).

3.1.1. Question responses table

The source data was stored in 4 excel sheets, corresponding to the two intakes (June and September) of both years (2009 and 2015). Each sheet stored tabular data in 56 lines (one per question), structured in the following columns for each question:

- Question code (which was not the same on both years, despite the content of the questions being the same).
- Number of successful attempts.
- Number of failed attempts.
- Number of times the question was disregarded.
- Total number of attempts (sum of successful, failed and disregarded attempts).

© Ediciones Universidad de Salamanca / CC BY NC-NB

The four tables were fused into a single one, appending the rows and programmatically adding and filling new fields named "year" and "intake" (which were implicit in the original tables), and a new common question code was introduced joining this table with a helper table relating the question codes of both years with a new synthetic code, with the following variables:

- The new (common) question code.
- The corresponding code of the question in the 2009 data.
- The code of the same question in the 2015 data.

Finally, the original question codes were dropped, as well as the field with the total number of responses since it could be computed from the sum of the responses of each different outcome.

Following the principles for "tidy data" (Wickham, 2014) the table was reshaped to have only one observation per row. In this context, an "observation" was defined as the number of responses to a question, with a specific outcome ("pass", "fail", "skip") in a given year and intake. Therefore, the structure of the table had to be changed from "short" (wide) to "long" (narrow) format, resulting in the following structure with 672 records (56 questions × 3 outcomes × 2 years × 2 intakes):

- Question code (56 questions).
- Year (2009, 2015).
- Intake (June, September).
- Outcome (pass, fail, skip).
- Number of responses (the actual data).

3.1.2. Question properties table

This table, which would later be joined with the question responses table for analysis, stored three variables (two categorical and one quantitative) for each question. After analyzing each question, the following categorical variables were encoded as factors:

- Whether the question included an illustration (about one third of the questions did).
- Knowledge subdomain of the question, with 6 factors (Projection Systems, Shape Geometry, Geometry Theory, Planar Geometry, Drawing Concepts, and Shadows).

From the question textual content, a length metric consisting on the sum of the question length and the length of all 4 different multiple choice responses, measured in number of characters (including spaces), was computed. The answers were included because it was assumed that the participants would read them before choosing their answer. This metric was considered as a preliminary proxy of reading comprehension difficulty for this study.

3.2. Computing Metrics

Once the tables were property formatted for analysis and visualization, the question responses and question properties tables were merged using a join operation.

Since the source data was summarized (count of responses), it was not possible to make conclusions about the participants as individuals, because their information had been aggregated and their unique identifiers had therefore been lost in the summarization process.

Since the analysis was based on aggregated data, some precautions had to be taken into account to avoid making inaccurate conclusions, such as the different participation in both years or the fact that different questions got different number of total responses (Table 2).

The authors considered that even though participation was almost double for year 2009 compared to 2015, the sample size was large enough in both cases to support the analysis, and both sets were comparable.

Year	Total responses	Mean	Std. Dev.
2009	13664	244.0	16.5
2015	6923	123.6	10.6
Combined	20587	367.6	23.2

Table 2. Differences in number of responses, mean number of responses and standard deviation of the number of total responses the questions received

The metric chosen to assess the influence of the different studied factors was the percent rate of each outcome for each question:

- Pass rate: Successful responses / total responses.
- Fail rate: Failed responses / total responses.
- Skip rate: Disregarded responses / total responses.

The aggregated nature of source data only allowed calculating rates individually for each question, globally or conditioned by the variable or variables of interest, to check their influence on the different rates. Since all three rates totaled 100%, this strategy provided a metric that harmonized the different number of responses of each question, allowing their comparison.

Finally, since working with aggregated data could introduce distortions in statistical hypothesis testing, the analysis used the number of responses as a weighting parameter in statistical tests and graphs when necessary to counteract this issue.

4. Results

The results were visualized graphically to explore the different relationships of the studied explanatory variables with the observed results in the responses (success and skip rates) using a systematic approach (Wickham, 2010; Wilkinson, 2005). The explored relationships were:

- Influence of image support.
- Variation across the two studied years.
- Variation between the two different intakes.
- Influence of question length.
- Differences in knowledge subdomain.

4.1. Influence of Image Support

Being technical drawing a subject eminently graphic in its nature, and considering that many of the students in this exam would need exceptional spatial reasoning skills in the degrees they would pursue later in their studies (e.g. Architecture, Engineering), it was important to establish if questions depicting a graphic image had a higher success rate and/or were skipped less.

To explore this relationship, a series of density plots were used, using the Sheather & Jones method to select the bandwidth of a Gaussian kernel density estimator. When plotting the density of the pass, fail and skip rates of the questions, where the rate of each individual question appears at the bottom of each graph as a mark in the rug plot, the differences between text-only questions (left panels) and the questions that were accompanied with a graphic (right panels) could be observed (Figure 1).

The pass ratio was higher when a figure was present (de curve was displaced to the right, where the rate was higher), while the fail rate did the opposite in an almost mirror image (because the number of skipped questions was small, and the sum of all percentages had to be 100%).

© Ediciones Universidad de Salamanca / CC BY NC-NB

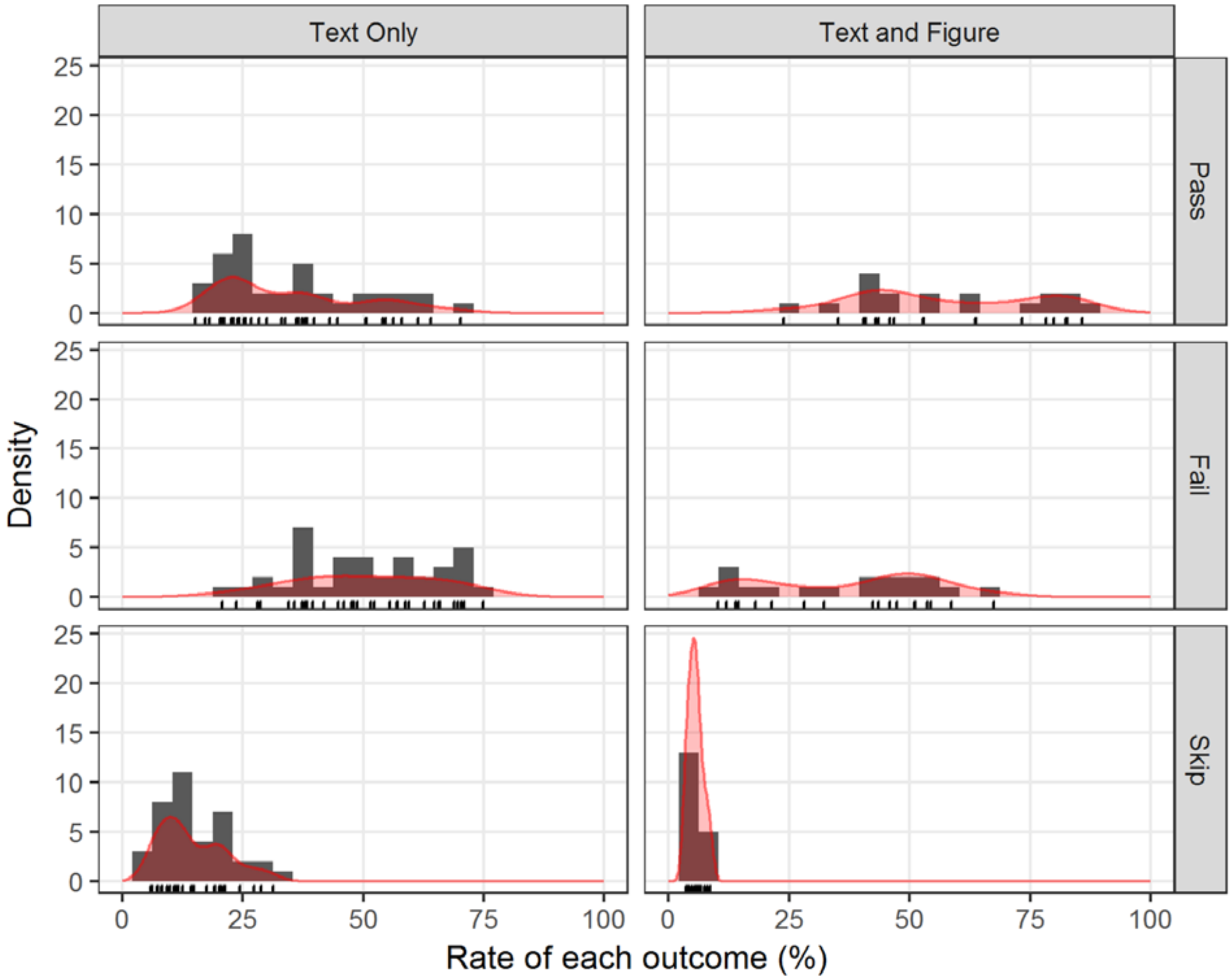

Figure 1. Density graph showing the pass, fail and skip rates when the question is text only (left panels) or has the support of an illustration (right panels)

In addition, the skip rate was much lower when there was a figure present, denoted by a spike around 7%, but when the question was text only the curve was much more spread out and reached over 25% in some cases.

Finally, when overlaying the same graph for both years (Figure 2), the results were very similar regardless of the 6 years elapsed, supporting the conclusion that including images in the question does indeed improve the success ratio and reduces the skip ratio.

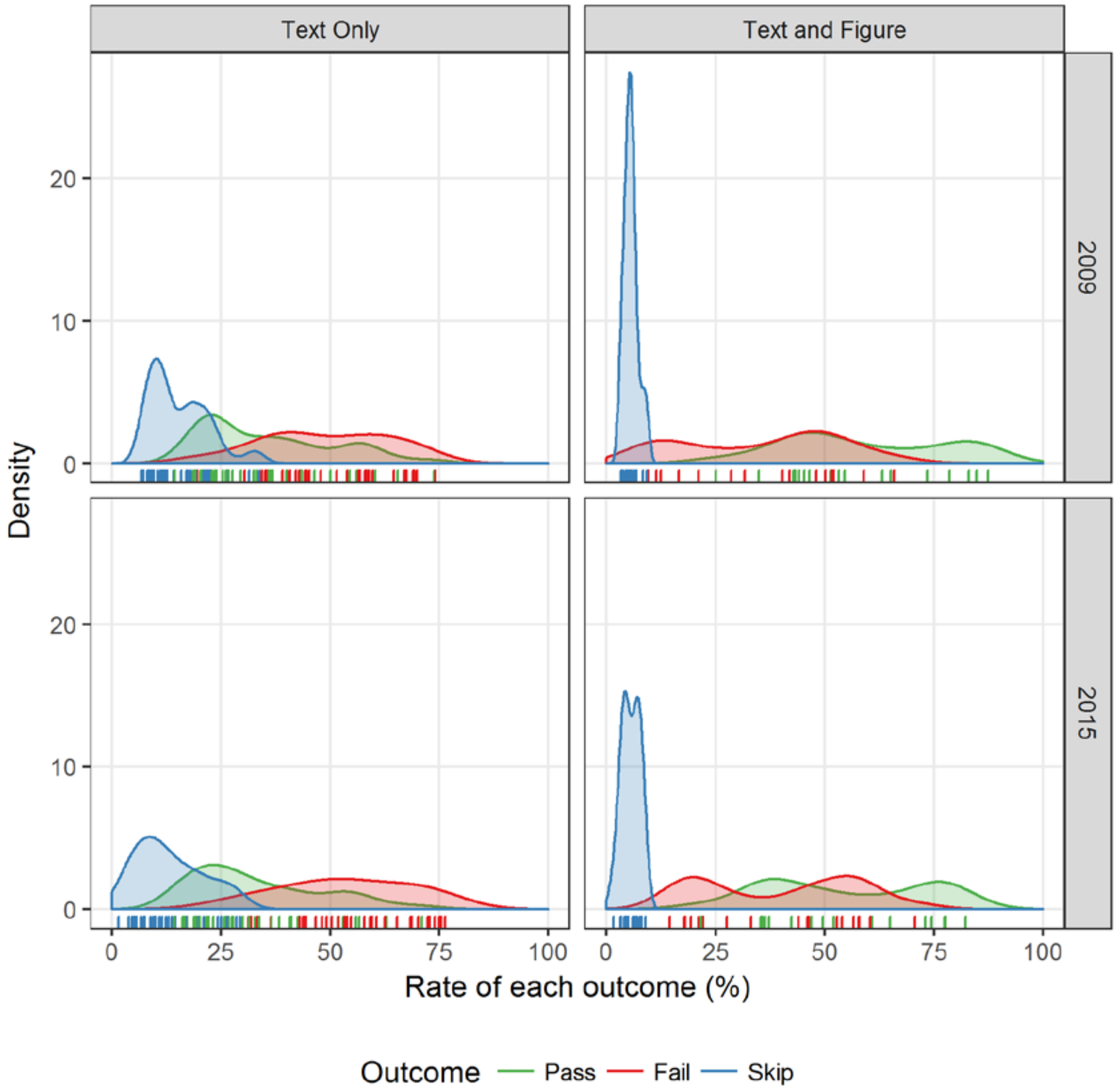

Figure 2. Density graph showing the pass, fail and skip rates when the question is text only (left) or has the support of an illustration (right) in two the different years studied

4.2. Variation Across the Two Studied Years

Since the students were asked the same questions in two different years, separated 6 years apart, it was possible to study the differences and similarities in the success and skip rates, reflecting changes in educational methodology or student profile between those years.

To reveal the changes across time, a graph inspired on the dot plot (Cleveland, 1993) was used, depicting the pass, fail and skip rates of each question in each year (Figure 3):

- The questions were sorted in descending order from left to right, depending on the success rate across both years combined.
- The number of responses was encoded as the size of the dot.

© Ediciones Universidad de Salamanca / CC BY NC-NB

- Each question had two dots, corresponding to each year rate, and the dots were colored according to the corresponding year.
- The smallest dots were drawn on over of the largest to avoid getting obscured by them.
- The graphs were arranged in panels where the leftmost ones contained the questions without pictures and the rightmost ones the questions with illustrations.
- The three outcomes (pass, fail and skip) were placed in three horizontal rows of panels.
- All panels shared the same scale.

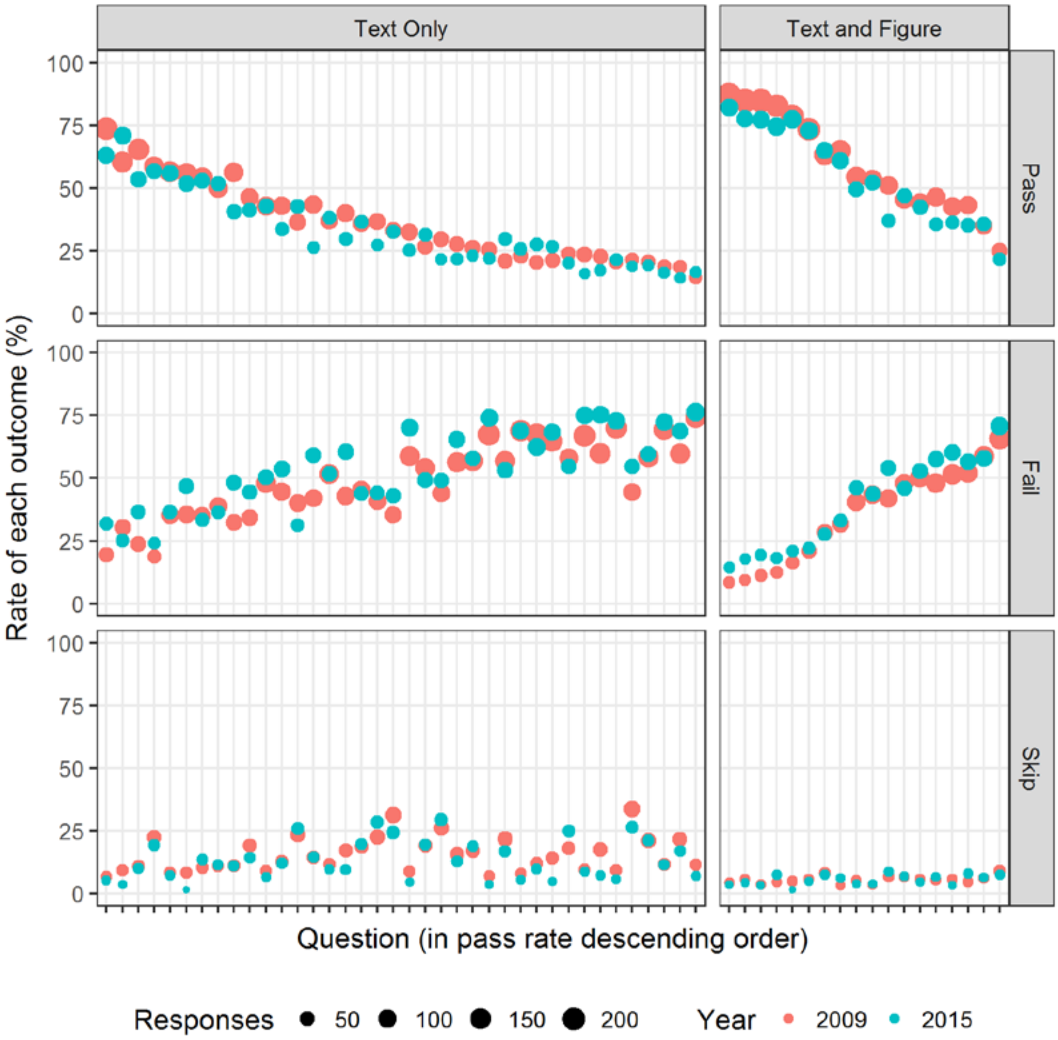

Figure 3. Dot plot of the influence of pass, fail and skip rates in the two years studied, the left panels corresponding to the questions without images and the right panel those with images. Symbol size is proportional to number of responses. Questions are sorted in descending overall rate for both years

In the graph it was easy to spot (from the vertical position of the dot) on which year the rate was higher for every question in each outcome (Table 3), and it could be observed that on 2009 the questions were more successfully answered (57% vs 18%) and less likely to be skipped (42% vs 15%) than 6 years later.

© Ediciones Universidad de Salamanca / CC BY NC-NB

	Higher in 2009	Higher in 2015	Tie (Δ<1%)
Pass	37 (57%)	12 (18%)	7 (11%)
Fail	11 (17%)	41 (63%)	4 (6%)
Skip	27 (42%)	10 (15%)	19 (29%)

Table 3. Number of questions where the rate of the three possible outcomes was higher on 2009, higher on 2015 or their rates difference was under 1%

The graph also showed two aspects that held true on both years for questions with image support compared to questions without: they usually showed higher success rates overall, and their skip ratio was systematically much lower (and in many cases close to zero).

4.3. Variation Between the Two Intakes

The table below (Table 4) summarizes the percentage of students that passed the overall exam in the two intakes of 2009 and 2015, according to official figures (Instituto Nacional de Estadística, 2009; Ministerio de Educación, Cultura y Deporte, 2015). In both years the percentage of students that passed the June intake was around 90%, which matches the roughly 90%-10% split observed in the data corresponding to the technical drawing online mock test on the same years (Table 1).

	June	September
2009	90.67%	66.34%
2015	95.23%	79.26%

Table 4. Percentage of students passing the first (June) and second (September) intakes

Since the students who take the September exam are the ones who did not pass the June exam and their pass rates are significantly lower, it would be expected that participants had different academic profiles. In the following graph (Figure 4) the relationship between the results in both intakes (first June and second September) was examined in detail on both years. The graph combined a violin plot with a box plot and allowed to easily make comparisons between the different factors studied.

Surprisingly, independently of the presence of an image, both years behaved differently: while in the 2009 September intake the pass rate of was lower than on the June intake, on 2015 the results were reversed, although by a very small margin. However, on both years the fail rate was higher in the September intake as expected. The explanation could reside in the skip rate, which on 2015 was lower in the second intake, suggesting that those students were more inclined to answer than on 2009, but were less successful in their attempts.

However, the very different sizes of the samples of both intakes made difficult any definitive conclusions, but confirmed the patterns observed in the analysis of the 6 year variation on subsection 4.2.

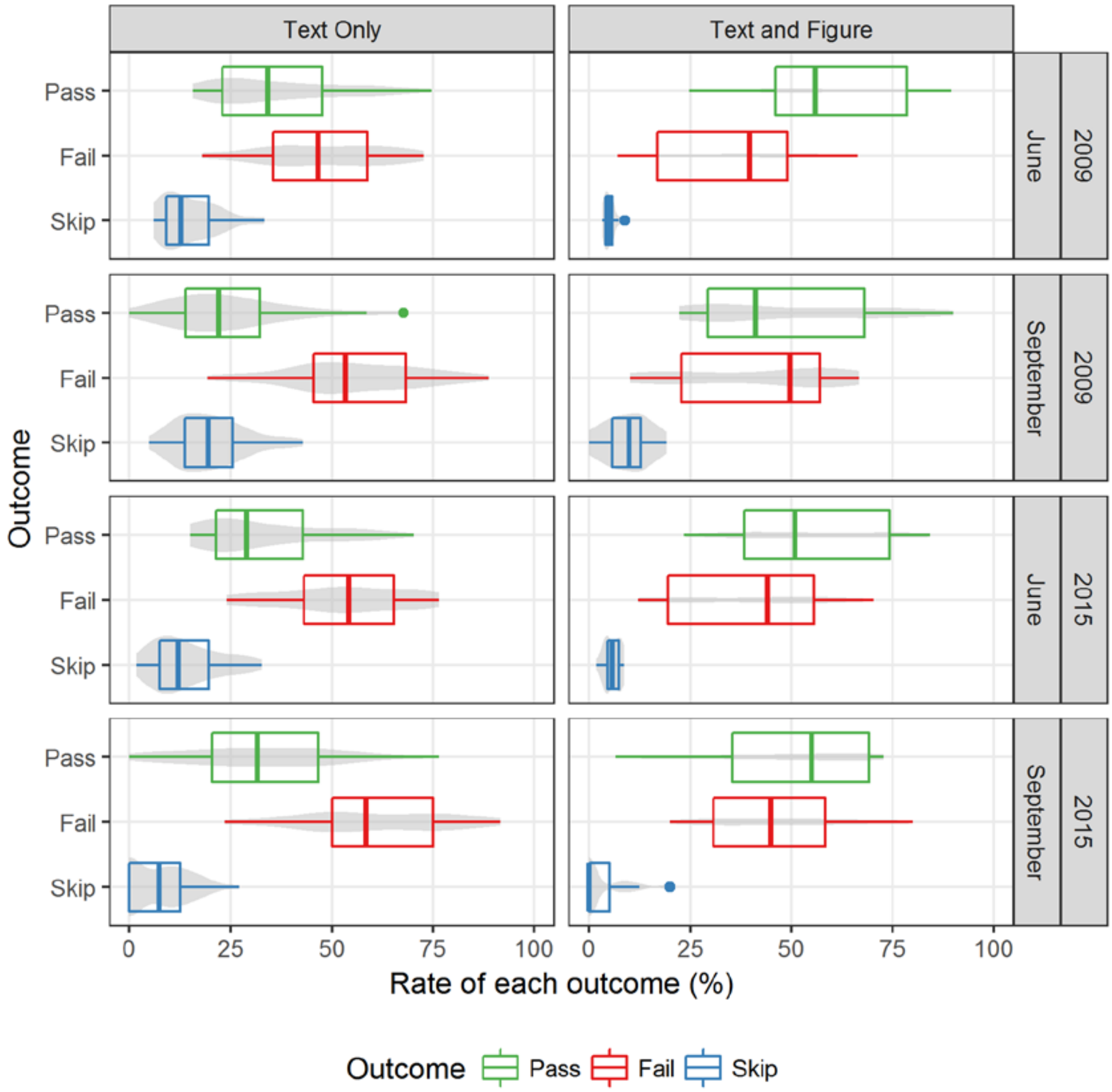

Figure 4. Influence of June or September intakes of both years on pass, fail and skip rates in questions with images or without

4.4. Question Length Influence

Having analyzed the influence of year and intake on success rate, the impact of question length (total characters of each question and four answers) as a measure of reading comprehension was assessed.

The charts used were scatterplots with the length of the question on the horizontal axis and the rate of the outcome (pass, fail and skip) on the vertical axis. There were 6 panels, the three outcomes dividing the space vertically and illustration support horizontally. The size of the symbols was proportional to the number of responses and their color was either their year (Figure 5) or intake (Figure 6). Linear regression lines also were colored accordingly as well as confidence intervals if present, and weighted according to the number of responses.

4.4.1. Influence of question length across years

The scatterplot of success, fail and skip rates versus question length (Figure 5) along the 6 year timespan revealed that performance as measured by the success ratio slightly decreased from 2009 to 2015 (as described in detail in subsection 4.2, which uses the same colors for the years) and greatly increased when a picture was included (as described in subsection 4.1). This ratio slowly decreased as the length of the question increased, with similar slopes in both years but with a lower intercept in 2015.

The skip ratio was very similar on both years, and the participants tended to skip the longer questions more often, and rarely skipped the questions without illustrations.

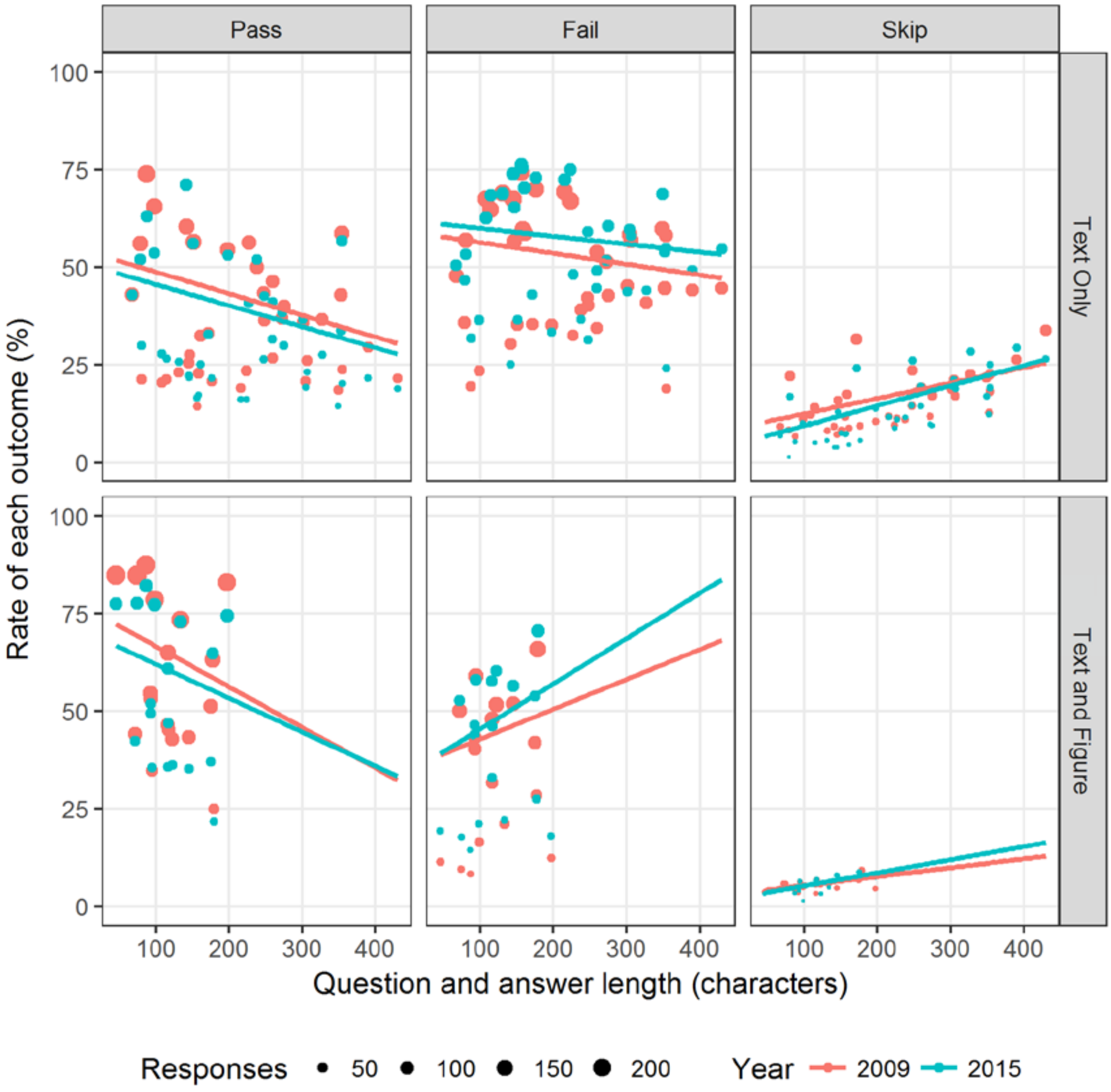

Figure 5. Relationship between question length and pass, fail and skip rates in the two studied years for questions with or without illustration support

4.4.2. Influence of question length between intakes

The scatterplot of the three studied rates versus question length (Figure 6) between the two intakes showed that the success ratio decreased for the students in the second intake (as described in detail in subsection 4.3, where the figure uses the same color scheme) and greatly increased when a picture was included (as described in subsection 4.1). This ratio decreased slowly as the length of the question increased, with similar slopes in both intakes but with a higher intercept in the June intake.

The skip ratio was very similar on both intakes when no picture was included, but higher on the second intake when there was illustration support. Longer questions were skipped more often than shorter ones, and questions with illustrations were less skipped.

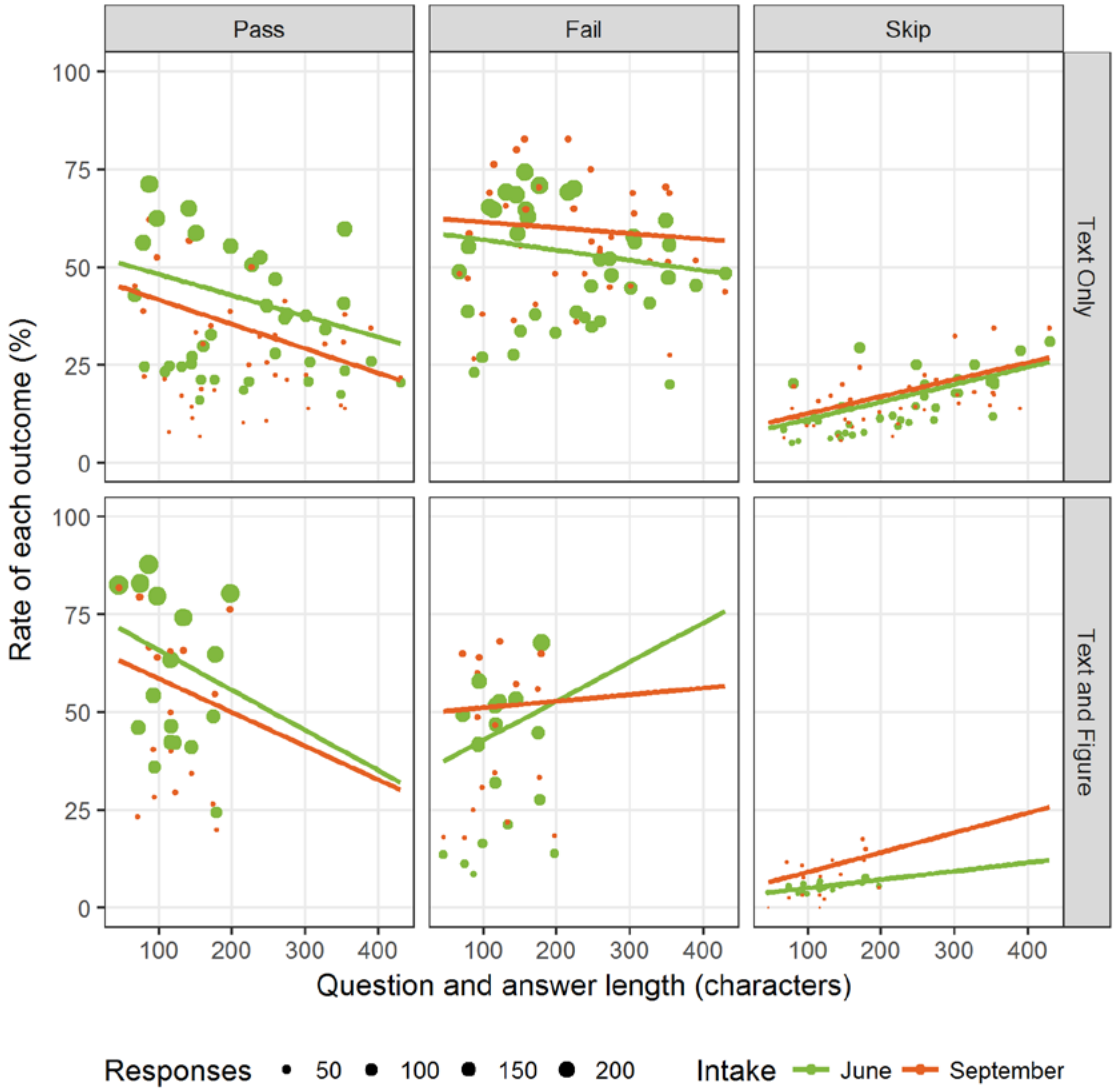

Figure 6. Relationship between question length and pass, fail and skip rates between intakes for questions with or without illustration support

4.5. Differences in Knowledge Subdomain

There were 6 knowledge subdomains defined within the subject of technical drawing (Table 5), which were sorted according to their number of questions in descending order.

Subdomain	Text (% group)	Image (% group)	Total (% all)
Projection Systems	12 (63%)	7 (37%)	19 (34%)

Shape Geometry	7 (47%)	8 (53%)	15 (27%)
Geometry Theory	12 (100%)	-	12 (21%)
Planar Geometry	3 (75%)	1 (25%)	4 (7%)
Drawing Concepts	4 (100%)	-	4 (7%)
Shadows	-	2 (100%)	2 (4%)

Table 5. Number of questions inside each of the 6 knowledge subdomains

When plotted separately (Figure 7), they did not show a distinct density signature, and considering the highly uneven distribution of illustration support inside the groups, and the varying number of questions in each group, it was not possible to make any definite conclusions about the differences in knowledge subdomain.

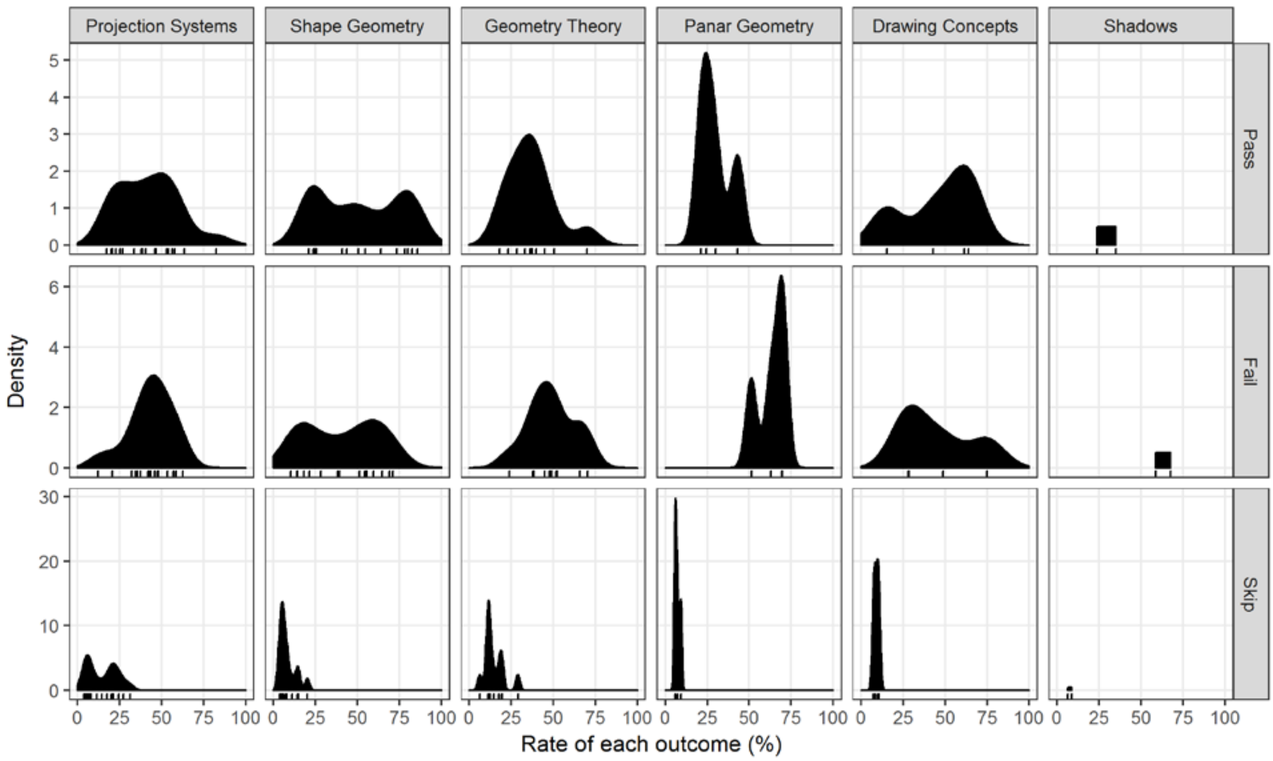

Figure 7. Density graph of success, fail and skip rates in the 6 knowledge subdomains (note that the "shadows" category does not have enough number of samples for the kernel density estimation)

5. Conclusions

The conducted analyses revealed that questions with image support received more successful attempts (i.e. students selected the correct answer) while at the same time significantly reduced the percentage of questions skipped, compared to the questions without images.

These results suggested that the students were more confident about their knowledge when dealing

© Ediciones Universidad de Salamanca / CC BY NC-NB

questions with image support (since they were skipped less often), and these questions were either (a) less difficult, (b) easier to understand or (c) more capable of motivating the students (or a combination of those).

The length of the whole question text (question and possible answers) was also analyzed as a proxy of reading comprehension difficulty, and the results showed that as the length of the question increased, the success rate decreased while the skip rate increased.

These findings suggest that students struggled to fully understand the longer questions or failed to read all the answers of the multiple choice test. The fact that longer questions were skipped more often supports this hypothesis, because students might abandon the question if they perceived that it was too long either because they felt they would leave less time to answer the shorter questions, or because they had trouble understanding what they were asked about.

Identifying the causes of these results should allow the authors to either include more questions with image support or rephrase the non-graphic questions to be shorter or more easily comprehended.

Although the questions in 2009 and 2015 were the same, the analysis of the variation across the 6 years between studies showed a reduction in the success ratio while the skip ratio stayed relatively the same. The authors suggest that while the students have quickly adopted new mobile and communication technologies, the curriculum they are taught and the way they are tested has not adapted yet, and this disconnect manifests itself in lower success rates.

Finally, the variation across intakes likely showed two user profiles: the students whose grades were higher (June) and lower (September). However, the fact that while in the second intake the success rate was lower but the skip rate stayed roughly the same, suggests that the motivation of the September group was comparable to the June group, although they were less knowledgeable overall.

5.1. Future Work

Since the images were not accounted in the question length, this introduced a potential confound because images with pictures were inherently shorter, as their answers were often just a single character (A, B, C or D) and therefore the analysis was biased. In future studies it is expected to correct this either (a) measuring just the length of the question, or (b) increasing the length of questions with images by a fixed number of characters. To measure question complexity more accurately, the semantic analysis of the content of the question will also be explored in future studies.

While current results have provided insight and further research questions using just statistical graphical methods, future studies will use statistical hypothesis testing to provide more robust conclusions while at the same time providing a measure of the influence of each factor using multiple regression analysis.

6. Acknowledgments

This research was supported by the programs BIA2016-77464-C2-1-R of the National Plan for Scientific Research, Development and Technological Innovation 2013-2016, Government of Spain, titled “Gamificación para la enseñanza del diseño urbano y la integración en ella de la participación ciudadana (ArchGAME4CITY)” and BIA2016-77464-C2-2-R of the National Plan for Scientific Research, Development and Technological Innovation 2013-2016, Government of Spain, titled “Diseño Gamificado de visualización 3D con sistemas de realidad virtual para el estudio de la mejora de competencias motivacionales, sociales y espaciales del usuario (EduGAME4CITY)”.

7. References

Cleveland, W. S. (1993). *Visualizing Data* (1 edition). Murray Hill, N.J. : Summit, N.J: Hobart Press.

Codd, E. F. (1990). *The relational model for database management: version 2*. Reading, Mass: Addison-Wesley.

Generalitat de Catalunya. (2009). Posa’t a prova. Retrieved June 14, 2016, from aplicacions.universitats.gencat.cat/posat/

Instituto Nacional de Estadística. (2009, December 17). Pruebas de Acceso a la Universidad. Año 2009. Instituto Nacional de Estadística. Retrieved from http://www.ine.es/prensa/np581.pdf

Ministerio de Educación, Cultura y Deporte. (2015). Datos generales de las Pruebas de Acceso a la Universidad (PAU), año 2015. Retrieved from http://www.mecd.gob.es/educacion-mecd/dms/mecd/educacion-mecd/areas-educacion/universidades/estadisticas-informes/estadisticas/estadistica-de-las-pruebas-de-acceso-a-la-universidad/anno-2015/datos-generales-de-las-pau_2015.xls

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, *334*(6060), 1226. doi:https://doi.org/10.1126/science.1213847

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from http://www.R-project.org/

Wickham, H. (2007). Reshaping Data with the {reshape} Package. *Journal of Statistical Software*, *21*(12), 1-20. doi:https://doi.org/10.18637/jss.v021.i12

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. doi:https://doi.org/10.1007/978-0-387-98141-3

Wickham, H. (2010). A Layered Grammar of Graphics. *Journal of Computational and Graphical Statistics*, *19*(1), 3-28. https://doi.org/10.1198/jcgs.2009.07098

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, *40*(1), 1-29. doi:https://doi.org/10.18637/jss.v040.i01

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10), 1-23. doi:https://doi.org/10.18637/jss.v059.i10

Wickham, H. (2016a). *stringr: Simple, Consistent Wrappers for Common String Operations*. Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2016b). *tidyr: Easily Tidy Data with `spread()` and `gather()` Functions*. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*. Retrieved from https://CRAN.R-project.org/package=dplyr

Wilkinson, L. (2005). *The Grammar of Graphics*. New York: Springer. doi:https://doi.org/10.1007/0-387-28695-0

Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., … Wilson, P. (2014). Best Practices for Scientific Computing. *PLOS Biol*, *12*(1), e1001745. doi:https://doi.org/10.1371/journal.pbio.1001745

Wing, J. M. (2006). Computational Thinking. *Communications of the ACM*, *49*(3), 33-35. doi:https://doi.org/10.1145/1118178.1118215