7 Assessing the efficacy of VR for foreign language learning using multimodal learning analytics

Tom Gorham[1], Sam Jubaed[2], Tannishtha Sanyal[3], and Emma L. Starr[4]

Abstract

This chapter describes a small-scale pilot study in which participants in the experimental group learned how to write Japanese kanji characters within an immersive Virtual Reality (VR) graffiti simulator (the Kingspray Graffiti Simulator on the Oculus Rift VR system). In comparing the experimental group to the non-VR control group in the context of embodied cognition, the authors used a multimodal learning analytics approach: the participants' body movements were recorded using a full-body 3D motion-tracker and clustered with a machine learning algorithm. The participants were also compared on the basis of a written posttest and a follow-up survey.

Keywords: computer assisted language learning, virtual reality, machine learning, multimodal learning analytics, Japanese writing.

1. Harvard Graduate School of Education, Cambridge, United States; tsg181@mail.harvard.edu

2. Harvard Graduate School of Education, Cambridge, United States; saj229@mail.harvard.edu

3. Harvard Graduate School of Education, Cambridge, United States; tas651@mail.harvard.edu

4. Harvard Graduate School of Education, Cambridge, United States; ems519@mail.harvard.edu

How to cite this chapter: Gorham, T., Jubaed, S., Sanyal, T., & Starr, E. L. (2019). Assessing the efficacy of VR for foreign language learning using multimodal learning analytics. In C. N. Giannikas, E. Kakoulli Constantinou & S. Papadima-Sophocleous (Eds), *Professional development in CALL: a selection of papers* (pp. 101-116). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.28.873

1. Introduction

1.1. Language learning, VR, and embodied cognition

Around the world there is encouragement for students to learn a foreign language (Devlin, 2015; Jackson, 2013). There is good reason for this; in addition to the potential economic and social benefits of being bilingual, there is evidence that it can also improve executive functions in children – a suite of cognitive skills that are strong predictors of future success, which include inhibition, working memory, and cognitive flexibility (Bialystok, 2015). Unfortunately, learning a foreign language can be an arduous and occasionally frustrating experience for many students (Lightbown & Spada, 2013). One of the underlying reasons for the current study is to examine one potential avenue that could make this process more effective, efficient, and enjoyable.

The present study examines one facet of foreign language learning: writing in a foreign script, particularly in one that is significantly different to a student's mother language. More specifically, it investigates the possible affordances that a fully immersive VR environment may offer for facilitating this process. For this study, a VR graffiti simulator was chosen as a comparison to traditional pen and paper approaches to foreign language writing practice. The novel VR experience might increase the participants' interest and enjoyment and hence improve their motivation and attitude (Lightbown & Spada, 2013). There is evidence that students who take handwritten notes of class lectures have better recollection and understanding of the material compared to students who type their notes using their laptops (Mueller & Oppenheimer, 2014). This may be due to the relatively slower speed of handwriting or it may result from the different patterns of brain activation that are caused by the fine motor manipulation of the pen (Kiefer et al., 2015). At any rate, to date, there has been little to no research done in making a comparison between foreign language writing practice done with pen and paper and similar practice done in an immersive VR environment.

It is worthwhile to research the educational affordances of VR because of the phenomenon of embodied cognition. It is the strong reciprocal relationship between the body and the mind, wherein the "perceptual and motor systems influence the way we construct concepts, make inferences, and use language" (Repetto, Serino, Macedonia, & Riva, 2016, para. 3). VR has the potential to leverage this embodied cognition to help improve the language learning outcomes of students (Macedonia, Müller, & Friederici, 2011; Repetto, Cipresso, & Riva, 2015). Because of this, an immersive virtual graffiti simulator might offer some benefits over more traditional approaches. For example, the students practise writing the scripts on extremely large canvases that appear to the user to be several meters in length and height. In order to 'paint' the characters in appropriately large fonts, they must use their entire bodies, reaching high, and squatting down low. However, measuring the physical interaction with the VR technology poses some significant challenges for data collection. To address these challenges, we turned to multimodal learning analytics and machine learning.

1.2. Multimodal learning analytics

Blikstein and Worsley (2016) consider multimodal learning analytics to be a central issue in the long-running educational battle between behaviourism (or neo-behaviourism) and constructivism. When it comes to measuring outcomes, they assert that the behaviourist side has traditionally had the advantage. That is because it relies on relatively easier approaches to data collection, including psychometrics and standardised testing, compared to the ones employed by researchers who study constructivist approaches. Blikstein and Worsley (2016) point out that many educators have spent decades calling for constructivist methodologies that are student-centred and focus on student autonomy – including such luminaries as Dewey, Freire, Montessori, and Barron and Darling-Hammond. However, the widespread adoption of such methodologies has been hampered by the challenge of data collection for research.

Fortunately, advances in hardware and machine learning may hold the promise of making constructivist approaches considerably easier to evaluate. Schneider and Blikstein's (2014) research progressed with the use of the Microsoft Kinect™ – a sensor that uses infrared light for full-body 3D motion capture and simple facial recognition – to investigate the correlation between changes in body posture during a learning activity and learning outcomes. Also, eye-tracking technology and computer vision machine learning algorithms were used to explore the pedagogical implications of joint visual attention (see Schneider & Blikstein, 2014; Schneider & Pea, 2013, 2014; Schneider et al., 2015). The outcomes of this chapter are closely connected to the aforementioned studies.

The present study was conducted as a graduate student research project overseen by Schneider. We took particular inspiration from one of his studies mentioned in the previous paragraph (Schneider & Blikstein, 2014). In that study, the researchers collected approximately 1 million data points regarding the X, Y, and Z cartesian coordinates of their test subjects' body movements using the Kinect™. To make sense of this huge amount of data, the researchers utilised an unsupervised machine learning algorithm called K-means. The K-means algorithm does not sort data into predefined categories (that would be called supervised machine learning); instead, it clusters data into novel groupings (see Bahnsen & Villegas, 2017 for an accessible introduction to K-means clustering). Through this clustering, the researchers were able to identify three prototypical body positions: active, semi-active, and passive. They were then able to draw correlations between the subjects' posture and the learning outcomes (e.g. surprisingly, there was a positive correlation between the number of transitions between active and passive positions and better learning outcomes).

We employed a similar approach in the present study. We collected 3D motion capture data with the Kinect™ sensor, clustered the data using the K-means algorithm, and attempted to identify prototypical body positions that might correlate to learning gains.

2. Method

2.1. Hypotheses

Some assumptions underlie the hypotheses of this study. First, language teachers can improve their students' motivation and attitude by providing activities that are enjoyable and interesting, which may in turn improve learning outcomes (Lightbown & Spada, 2013). Second, increased physical movement in the context of embodied cognition will result in improved learning outcomes (Kiefer et al., 2015; Repetto et al., 2016). Third, learners will exhibit greater body movements (the head and arms specifically) when using VR to learn a new script versus pen and paper. Based on these assumptions, we developed two hypotheses for this study:

- Learners will exhibit greater excitement and engagement using VR to learn a new script versus pen and paper.

- Learners who use VR are able to reproduce the script characters more accurately as compared to those who learned using pen and paper.

We hoped to find some indication of prototypical body positions that might correlate to learning gains.

2.2. Participants

The participants for this study were three female students in their 20's at the Harvard University Graduate School of Education. We used convenience sampling to recruit the participants for this study; participants were people that the researchers knew and informally recruited. Because of the demographic makeup of the school, it was easier to recruit female participants. Participants had no previous experience using VR and no previous experience with the Japanese kanji script. All of the participants were native English speakers from the United States of America. Participants were not compensated for participation. Two participants (labeled as VR001 and VR002) were assigned to the experimental group that received the VR treatment. The third participant (labeled as VR003) was assigned to the control group which studied using traditional pen and paper.

2.3. Materials and data collection tools

We tested all participants on their ability to remember and write the seven basic logograms for the days of the week written in the Japanese kanji script as well as their English transliteration (e.g. 月= 'getsu'=Monday; 火= 'ka'=Tuesday; 水 = 'sui'=Wednesday; etc.).

The participant (VR003) in the control group was given a paper-based list of the seven target Japanese Kanji characters which included their stroke order and their English transliteration. She was also given a desk, a blank notebook, and a pen to use for studying.

The participants (VR001 and VR002) in the experimental group took part in the study individually. Each person used an Oculus Rift VR system which was running on an Alienware X51 personal computer and playing the VR app, Kingspray Graffiti Simulator (Figure 1). Within the virtual environment, a list of the seven target Japanese kanji characters which included their stroke order and their English transliteration was pre-painted on the graffiti wall. To get a better sense of what the experience of painting in the Kingspray Graffiti Simulator is like, we recommend that readers watch a short demonstration video (https://youtu.be/dhIxY6G-UHE).

All participants' sessions were video recorded with a smartphone to facilitate behavioural observations. Participant's motions during the session were tracked using a Microsoft Kinect™ (Figure 1) using a data collection tool developed by Dr Bertrand Schneider (which can be found at https://github.com/hgse-schneider/Kinect_Data_Collection_Tool). This data was then analysed and clustered using the K-means machine learning algorithm within the data visualisation software, Tableau. Although the data collection tool records information about multiple body parts, including the X, Y, and Z cartesian coordinates of the participant's hands, wrists, elbows, shoulders, and spine, for the purposes of this chapter, we will focus on the analysis and clustering of the X and Y coordinates of the participants' heads.

After the alloted length of study time, all participants took a paper-based posttest to assess their ability to remember and correctly reproduce the kanji logograms for the days of the week and the phonetic spelling of each symbol in English. After that, each participant filled out an online survey to assess their engagement in the learning activity. The follow-up survey included 28 Likert scale evaluations of statements that covered engagement-related topics such as novelty, aesthetics, involvement, and endurability.

Figure 1. Set-up of the experimental VR condition (**Left**: The Oculus Rift sensor in the foreground and the KinectTM sensor in front of the monitor; **Right**: A participant (background) squatting down while using the Oculus Rift and the KinectTM software (foreground) recording her movement on a laptop)

2.4. Procedure

The participant (VR003) in the control group was given a paper-based list of the seven target Japanese Kanji characters which included their stroke order and their English transliteration. She was also given a desk, a blank notebook, and a pen to use for studying. One of the researchers/authors was on hand to answer any questions she had regarding character shape, stroke order, or English transliteration. The participant was not given any further guidance on how to study. The participant then had up to 20 minutes to study the seven kanji characters. After the study period, the participant took the paper-based posttest and then filled out the online learning engagement survey.

The participants (VR001 and VR002) in the experimental group received the treatment separately. Each participant was first set up in the Oculus Rift VR headset by the researchers and led through a five- to ten-minute tutorial on how to navigate and interact with the VR environment. After this tutorial, on a wall in the virtual environment the participants could see a pre-painted list of the seven target Japanese Kanji characters, including their stroke order and their English transliteration. One of the researchers/authors was on hand to answer any questions the participants had regarding the kanji characters or about the VR system. The participants were not given any further guidance on how to study. The participants then had up to 20 minutes to study the seven kanji characters within the VR graffiti simulator (Figure 2). After the study period, each participant took the paper-based posttest and then filled out the online learning engagement survey.

Figure 2. The participants' view in experimental condition while using the Oculus Rift and Kingspray Graffiti Simulator[5] (**Left**: Wide 'screenshot' of the alley environment used within the Kingspray simulator; **Right**: Close-up 'screenshot' of the brick wall where the participants practiced their kanji characters)

5. Reproduced with kind permissions from Kingspray.

3. Results

3.1. Descriptive statistics

The participant (VR003) in the control condition performed better than the participants (VR001 and VR002) in the experimental condition, with a posttest score of 19 (out of 21) compared to the VR001's score of 7 and VR002's score of 12. However, the VR participants reported higher engagement in the online follow-up survey than the control participant, with VR001 rating the activity as 1.19 on a scale of -2 to 2, VR002 rating it a 1.67, and VR003 rating it 0.98.

In addition to the posttests and the online follow-up survey, all of the participants' sessions were video recorded with a smartphone to facilitate behavioural observations (Table 1). The experimental group expressed more interest and excitement than the control did, but it also expressed more discomfort and asked more questions. As an example of an expression of excitement, experimental group participant, VR002, exclaimed, "Ok! This is so fun!".

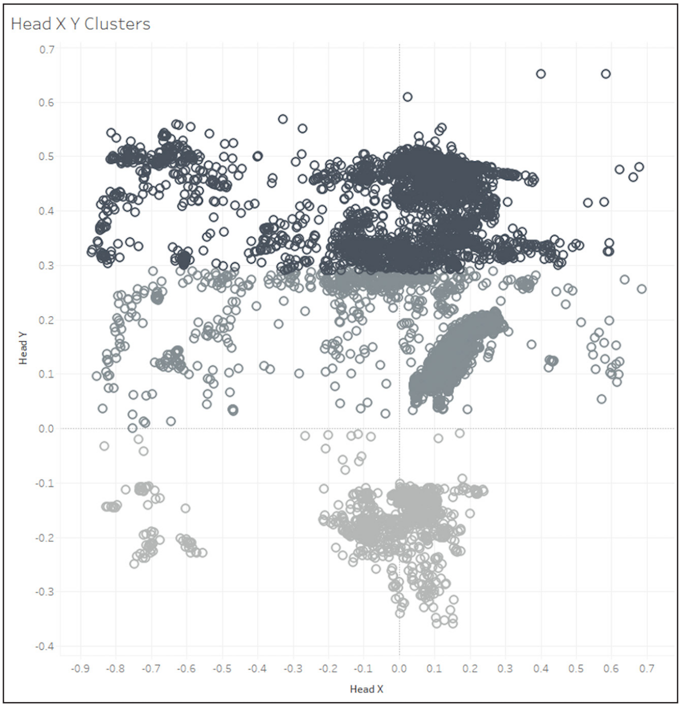
Table 1. Behavioural observations from videos of participants' sessions and results from posttests and follow-up surveys

	VR001 (experimental)	**VR002 (experimental)**	**VR003 (control)**
Number of times verbally expressing interest in activity	2	1	0
Number of times verbally expressing excitement	3	9	0
Number of times verbally expressing discomfort	2	2	1
Number of times participant asks experimenter a question	8	4	1
Paper-based posttest results (out of 21 points)	7	12	19
Online follow-up survey measuring engagement (on a scale from -2 to 2	1.19	1.67	0.98

3.2. Clusters

As mentioned earlier, because we were inspired by the phenomenon of embodied cognition and Schneider and Blikstein's (2014) research, we hoped to find some indication of prototypical body positions that might correlate to learning gains. To do this we used a Microsoft Kinect™ (Figure 1) data collection tool to gather three-dimensional body movements and position data. We chose the X and Y cartesian coordinates of the participants' heads to serve as a simple proxy for physical movement and position. We then clustered that data using a K-means unsupervised algorithm in the Tableau data visualisation software. This clustering performed in Tableau resulted in three clusters: high, medium, and low (Figure 3).

Figure 3. Head X and head Y clusters (Tableau)



In the high cluster, participants had high head Y values, demonstrating that they were standing upright or even reaching up. In the medium cluster, participants were leaning over, producing lower head Y values. In the low cluster, participants had extremely low head Y values, which were indicative of crouching or squatting. See Table 2 for examples of what each prototypical body posture looks like.

Table 2. Head X and head Y cluster prototypical postures

High Cluster	**Medium Cluster**	**Low Cluster**
		

Figure 4. Timelines of clusters of relative head position based on head X and head Y values, by participant

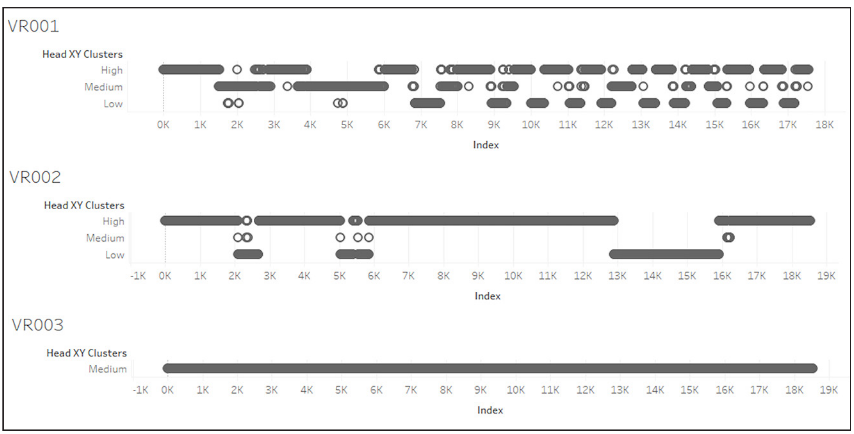


Figure 4 shows the times that each participant was in one of these three clusters. Note that because the data collection tool was recording data points 15 times per second for approximately 20 minutes, the timeline shown in the X axis in Figure 4 runs to 18,000 (15 data collection points per second times 60 seconds times 20 minutes equals 18,000). This shows that the participant in the control group (VR003), who was sitting down at a desk working in a notebook, stayed exclusively in the medium cluster. In comparing the participants in the experimental group, VR001 switched frequently between all of the clusters, whereas VR002 spent most of her time in the high cluster and less frequent, but relatively longer times in the low cluster.

4. Discussion

Our first hypothesis was that participants in the experimental group would exhibit greater excitement and engagement using VR to learn a new script versus pen and paper. Based on the results of the current study, this was found to be true. This is not particularly surprising because the Kingspray Graffiti Simulator is designed as a commercial off-the-shelf VR video game, and its primary purpose is to entertain and engage its users. Furthermore, the participants had no prior experience with VR, so as a novel experience, it was likely to be more exciting than traditional pen and paper study.

Our second hypothesis was that participants in the experimental group would be able to remember and reproduce the kanji characters more accurately as compared to the participant in the control group. However, the posttest contradicted this hypothesis. The control participant achieved a much higher posttest score than either of the VR participants.

These unexpected results have many possible explanations. First, it is possible that the medium of the test might have played a role in the outcome. Namely, the posttest was administered as a written paper test. The control participant practised in the same medium as the test was administered in, while the VR treatment group did all of their practice within an immersive digital environment which was considerably different than the medium in which they were tested. It is possible that this raises issues of transfer. Perhaps in future iterations of this study, the posttest could be administered in the virtual environment for both groups to see if the testing medium has an effect on learning outcomes.

A second explanation is that the novel VR environment itself hindered the learning of the VR treatment group. This could either be attributed to a steeper learning curve (and thus increased cognitive load) involved in becoming acclimated to the VR control scheme or possibly because the digital environment contains many distractions. For instance, the virtual alleyway in which the participants practised featured realistic elements like birds and passing trains which might have drawn the participants' attention away from the target task.

The findings of this study suggest one particularly interesting route for future investigation. Because the control participant utilised well-known study techniques such as spaced repetition but the participants in the VR condition reported higher engagement, perhaps it would be valuable to test a new experimental condition in which both the paper and VR approaches were combined.

In this proposed future study, participants would begin by practising on paper using established and proven techniques (like spaced repetition) for a period of time. Then, once they feel comfortable with the material that they were studying, they would then enter the VR environment for a shorter period of time. Within the digital environment, the participants would be asked to make a large-scale, artistic visualisation of the characters that they had been studying.

This approach would address challenges which might arise when using VR in second language learning. It leverages the best elements of both treatment approaches; the participants get high-quality and high-volume practice on paper, supplemented by the novel and engaging experience of the VR treatment. There is also a practical benefit to this mixed approach. VR equipment is currently expensive, so it would be unlikely that most classrooms would have enough equipment for each student to have their own headset. In the mixed approach, the students would do the majority of their practice on paper and only a short amount (i.e. five or ten minutes) within the VR environment. This would be a more feasible approach for classrooms with limited access to VR technology.

One final point that is worth considering for future VR and multimodal learning analytics research is that the better performing participant in the experimental group (VR002) had a movement pattern that was more similar to the control participant (VR003); she had long periods of less movement and fewer transitions between clusters (Figure 4). It is possible that this is an indication of increased concentration and focus during practice. More research is needed to determine if this pattern of movement does, in fact, correlate with improved learning outcomes.

5. Conclusions

This pilot study introduces a possible way that multimodal learning analytics can supplement an evaluation of a language learning intervention using VR. Although the results of the study did not support the hypothesis that participants studying Japanese kanji with VR would outperform a participant using a more traditional method, it suggests ways in which this line of inquiry can be expanded. Larger sample sizes and the mixed approach (VR plus pen and paper) described in the previous section are both promising avenues for future research.

Acknowledgements

We would like to thank our professor, Bertrand Schneider, and our teaching assistants, Joseph Reilly and Jeffrey King, for their support throughout this study.

References

Bahnsen, A. C., & Villegas, S. (2017). *Machine learning algorithms explained – K-means clustering*. https://blog.easysol.net/machine-learning-algorithms-3/

Bialystok, E. (2015). Bilingualism and the development of executive function: the role of attention. *Child Development Perspectives*, *9*(2), 117-121.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220-238. https://doi.org/10.18608/jla.2016.32.11

Devlin, K. (2015). *Learning a foreign language a 'must' in Europe, not so in America*. http://www.pewresearch.org/fact-tank/2015/07/13/learning-a-foreign-language-a-must-in-europe-not-so-in-america/

Jackson, A. (2013). *Foreign language policies around the world*. http://blogs.edweek.org/edweek/global_learning/2013/03/foreign_language_policies_around_the_world.html

Kiefer, M., Schuler, S., Mayer, C., Trumpp, N. M., Hille, K., & Sachse, S. (2015). Handwriting or typewriting? The influence of pen- or keyboard-based writing training on reading and writing performance in preschool children. *Advances in cognitive psychology*, *11*(4), 136-46.

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping, 32*(6), 982-998. https://doi.org/10.1002/hbm.21084

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard. *Psychological Science, 25*(6), 1159-1168. https://doi.org/10.1177/0956797614524581

Repetto, C., Cipresso, P., & Riva, G. (2015). Virtual action and real action have different impacts on comprehension of concrete verbs. *Frontiers in Psychology, 6*, 1-9. https://doi.org/10.3389/fpsyg.2015.00176

Repetto, C., Serino, S., Macedonia, M., & Riva, G. (2016). Virtual reality as an embodied tool to enhance episodic memory in elderly. *Frontiers in Psychology, 7*, 1-4. https://doi.org/10.3389/fpsyg.2016.01839

Schneider, B., & Blikstein, P. (2014). Unraveling students' interaction around a tangible interface using gesture recognition. In J. Stamper, Z. Pardos, M. Mavrikis & B. Mclauren (Eds), *Proceedings of the 7th International Conference on Educational Data Mining (EDM'14)* (pp.320-323). http://educationaldatamining.org/EDM2014/proceedings/EDM2014Proceedings.pdf

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 375-397. https://doi.org/10.1007/s11412-013-9181-4

Schneider, B., & Pea, R. (2014). Toward collaboration sensing. *International Journal of Computer-Supported Collaborative learning, 9*(4), 371-395. https://doi.org/10.1007/s11412-014-9202-y

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, A. D. (2015). 3D tangibles facilitate joint visual attention in dyads. In O. Lindwall, P. Hakkinen, T. Koschmann, P. Tschounikine & S. Ludvigsen (Eds), *Proceedings of the International Conference on Computer Supported Collaborative Learning 2015: Exploring the Material Conditions of Learning (CSCL'15)* (Vol.1, pp. 158-165). The International Society of the Learning Sciences.

esearch-publishing.net

Published by Research-publishing.net, a not-for-profit association
Voillans, France, info@research-publishing.net

Professional development in CALL: a selection of papers
Edited by Christina Nicole Giannikas, Elis Kakoulli Constantinou, and Salomi Papadima-Sophocleous

Typeset by Research-publishing.net
Illustration cover: © apinan / stock.adobe.com
Cover design: © Raphaël Savina (raphael@savina.net)

ISBN13: 978-2-490057-28-3 (Ebook, PDF, colour)
ISBN13: 978-2-490057-29-0 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-27-6 (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'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, UK: British Library.
Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: mars 2019.