JPSI (Journal of Public Sector Innovations)
Vol. 7, No. 2, May 2023

Journal homepage: https://journal.unesa.ac.id/index.php/jpsi/index

Print ISSN: 2541-4291
Online ISSN: 2348-396X

To link to this article: https://doi.org/10.26740/jpsi.v7n2.p67-75

Application of A/B Testing Experimentation on Government Digital Products to Enhance Teachers' Skills and Capabilities in Indonesia

Bagoes Rahmat Widiarso
PT Metra-NET TELKOM Group Indonesia
E-mail: bagoes.rahmat@wartek.belajar.id

Bhaskoro Abdillah Muthohar
PT Metra-NET TELKOM Group Indonesia
E-mail: bhaskoro.muthohar@wartek.belajar.id

Septi Rito Tombe
PT Metra-NET TELKOM Group Indonesia
E-mail: septi.tombe@wartek.belajar.id

Putri Wikie Novianti
PT Metra-NET TELKOM Group Indonesia
E-mail: putri@wartek.belajar.id

Abstract

The Ministry of Education, Culture and Research and Technology (MoECRT) in Indonesia has developed a digital product named "Microlearning" within the *Merdeka Mengajar* digital Platform. "Microlearning" enables teachers to enhance their skills and capabilities by doing self-determined learning that is followed-up by the submission of *Aksi-Nyata* reports. We improved users' journey flow and provided an alternative of user-interface designs in the platform, in which we hypothesized that these improvements could boost the number of teachers participating in the *Aksi-Nyata* report submissions - the main expected outcome of the "Microlearning" digital product. We conducted a randomized-control-trials experimentation in the digital platform to scientifically quantify its effect on the product, by including 1.6 million users who logged-in to the platform during a two weeks period. This experimentation (known as an A/B Testing) allowed us to simultaneously compare two groups within a specific period and within a controlled-environment. The bayesian-based experimentation revealed that the new developments increased the number of users who started to learn a topic, as well as increased the *Aksi-Nyata* report submission in the "Microlearning" product, as compared to the legacy-designs. A follow-up study showed even more convincing results, which shows the number of submitted *Aksi-Nyata* reports increased by 590% up to two months after implementing the analytics-based initiatives for all logged-in users in the platform.

Keywords: *Microlearning*, *self-determined learning, Bayesian*

Corresponding author: Putri WIkie Novianti; putri@wartek.belajar.id

Article history: Received, 31 December 2022; Revised, 14 April 2023; Accepted, 12 May 2023.

To cite this article: Widiarso, B. R., Muthohar, B. A., Rito, S., & Wikie, N. P. (2023). Application of A/B Testing Experimentation on Government Digital Products to Enhance Teachers' Skills and Capabilities in Indonesia. *JPSI (Journal of Public Sector Innovations)*, *7*(2), 67–75. https://doi.org/10.26740/jpsi.v7n2.p67-75

INTRODUCTION

In the last decade, there has been a rapid global movement towards digitization in the public sector. Countries worldwide have collaborated to shift public services from offline to online environments, referred to as e-government. This trend is quantitatively demonstrated by the increase in the e-Government Development Index (EGDI) across almost all nations between 2020 and 2022 (UN, 2022). Indonesia is not an exception to this trend and has seen a significant improvement in its UN e-government index. The country's rank has risen considerably, from the 116th rank in 2016 and 107th rank in 2018, to the 88th spot in 2020 (Kominfo RI, 2020). This could be an indicator of the Indonesian government's commitment to meeting its citizens' needs in the digital age.

The Ministry of Education, Culture, Research and Technology of the Republic of Indonesia (MoECRT) has been actively involved in such digital transformation by introducing various programs that are supported by digital platforms as its main means of implementation. These programs include: (i) the adoption of "Microlearning" to promote self-determined-learning among teachers in the Implementasi Kurikulum Merdeka program, (ii) the utilization of “Rapor Pendidikan” digital platform to facilitate Perencanaan Berbasis Data program and (iii) the implementation of ARKAS and SIPlah to improve administrative services for school principals (Kemendikbud Ristek RI, 2022).

One of digital products that MoECRT launched in 2022 is “Microlearning” that was developed within Merdeka Belajar Platform (refer to as: PMM). This digital product is specifically designed for teachers as well as school principals (refer to as: user) to enhance their skills and capabilities by doing asynchronous and self-determined learning on the available topics in the platform. It is aimed to be a primary digital-tool for accelerating the understanding of the Merdeka Curriculum (known as KM) in more than 143,000 schools throughout Indonesia (Kemendikbud Ristek RI, 2022). At the end of the learning program, users are strongly encouraged to reflect on their learning process through a final activity called Aksi-Nyata. Users may submit their Aksi-Nyata report through the platform, which is then validated by well-trained assessors. Should the Aksi-Nyata report pass the minimum standard defined by the committee, users will receive a Sertifikat Pelatihan Mandiri. In the “Microlearning” program, the submission of the Aksi-Nyata report is the main indicator to quantify the understanding of learnt-materials.

One of the major challenges faced by the team during the development phase, was ensuring the product to be suitable for fulfilling the needs of teachers in developing their capacity and their teaching quality. At the outset of its introduction, however, a significant number of users did not even complete one learning topic, which indeed led to a low number of users submitting their Aksi-Nyata report. Such conditions raised concerns and warrant further investigations on the root cause of the situation, i.e. whether the users faced challenges related to the substance of KM and/or to the experience in the platform. To address the latter hypothesis, we initiated improvement in the platform, by changing the design interface and improving the users’ experience in the platform.

First, our qualitative research team conducted user interviews towards some teachers from various categories of school levels and ages. The result of the research suggested that teachers would be more inspired to complete a learning package in “Microlearning”, if they were exposed to colleagues’ or other teachers’ Aksi-Nyata report submissions. This qualitative finding drove us to formulate a further hypothesis, i.e: users would be more likely to finish learning materials and further to submit an Aksi-Nyata report, if they were exposed to other teachers’ Aksi-Nyata report submission.

Next, the team created a competitive user-interface design as well as providing an alternative users’ flow, to accommodate the aforementioned-hypothesis and further to develop those designs into the platform. The team evaluated its effectiveness by comparing the new improvements with the legacy designs via in-app experimentation (or widely known in the technology industry as A/B testing).

Despite its widespread use in the private industry such as e-commerce (Westland, 2022) or marketing (Xue et al, 2022) sector, A/B testing is not yet extensively utilized in the public sector. With the aim of making data-driven decisions, particularly in the realm of technology use in the public sector, we chose to apply the A/B testing approach to assess our prior hypothesis. This paper presents the implementation of A/B testing to improve the users’ experience in the “Microlearning” digital products, as part of the effort to accelerate the improvement of teachers’ skills and capabilities in Indonesia.

METHOD

Our analytical approaches are divided into two phases. First, we conducted descriptive data explorations to obtain prior knowledge on the data characteristics before applying the A/B testing. This step is crucial to uncover the unknowns-unknowns and to set the foundational experimental design. Second, we focused on applying A/B testing to test our hypothesis: the number of users who submit Aksi-Nyata reports would increase when users are exposed to other teachers' Aksi-Nyata reports.

1. Data Exploration and Time Series Modeling

On the descriptive data exploration, we observed the demographic distribution of users (teachers and school principals) who logged-in to the PMM and experienced the new page design. To predict the number of users who submit an Aksi-Nyata report, we employ `ARIMA(p,d,q)` model. `ARIMA` is a prediction model that simultaneously combines autoregressive processes and moving averages. Prediction modeling allows us to compare the upward trend between before and after the A/B testing. An ARIMA model equation (p,d,q) is shown as follows (Wei, 2006).

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t,$$

where an autoregressive AR(p) model is indicated by this following equation:

$$\phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p,$$

and an moving average MA(q) model is indicated by this equation:

$$\theta_q(B) = 1 - \theta_1 B - \ldots - \theta_q B^q.$$

A non-stationary model with d=1 is then written as follows

$$(1-B)Z_t = a_t$$
$$Z_t = Z_{t-1} + a_t$$

2. A/B Testing

a. Bayesian A/B Testing

We opted for Bayesian-based A/B testing to analyze the data, as it addresses several limitations in the analysis and decision-making process of frequentist-based testing (outlined in Table 1). With Bayesian-based A/B testing, we can formulate meaningful conclusions on whether the conversion rate probability of an alternative is superior to the legacy based on numerical measurements. This probability can corroborate the prior distribution-based information we possess.

Prior probability is derived from initial belief that an option or a treatment becomes successful or effective before the primary datasets are actually being observed. Along with priors, there are likelihoods that aid in determining the probability of treatment A being superior to treatment B. The likelihood refers to the probability of treatment A that is derived through information (or data) obtained during the experimentation. By combining probability of treatment, A from both priors and likelihoods, the Bayesian approach computes the probability of treatment A being better than another. This probability is referred to as the posterior, also known as the combination probability of priors and likelihoods. In this study, we used the conversion rate of Aksi-Nyata report submission before the experimentation as our prior distribution; while we computed the posteriors from the data gathered during the experimentation.

Despite its complex calculation, the results from Bayesian testing may be reliable. The general process of Bayesian is visualized in Figure 1. Prior, likelihood, and posterior typically follow the beta distribution with parameters (α, β), where both α and β are two parameters that control the shape of the beta distribution and are represented by a continuous number greater than 0. The α is a parameter represented by numerical variables that shows the desired event to happen. Supposed there are 10 users who complete the learning and only 6 of them submit their Aksi-Nyata report, the α is then equal to 6. Meanwhile, β describes how an unexpected event happened. For instance, 4 out of 10 teachers do not submit their Aksi-Nyata report to the platform, the β is then equal to 6. In general, the relationship to the posterior distribution equation is as follows Puza (2015),

$$Beta(\alpha_{posterior}, \beta_{posterior}) = Beta(\alpha_{prior} + \alpha_{likelihood}, \beta_{prior} + \beta_{likelihood})$$

.

Table 1. Comparison between Frequentist and Bayesian Methods (Samaniego, 2010)

Area	Frequentist	Bayesian
Assumption	Requires strict assumptions to get the p-value.	Requires fewer assumptions to meet before making a decision.
Decision making process	P-values are difficult to interpret in a business context, making them vulnerable to be misunderstood by peers who are less familiar with statistical testing.	The Bayesian method calculates the probability that one group will be the best over the other groups.
Uncertainty	Unable to interpret the uncertainty that may occur in the future.	Intuitive interpretation of uncertainty by using the density distribution.
Calculation	Simpler calculation, for sample sizes below 30, it is easy to calculate without computational assistance.	Requires more extensive and complex computation

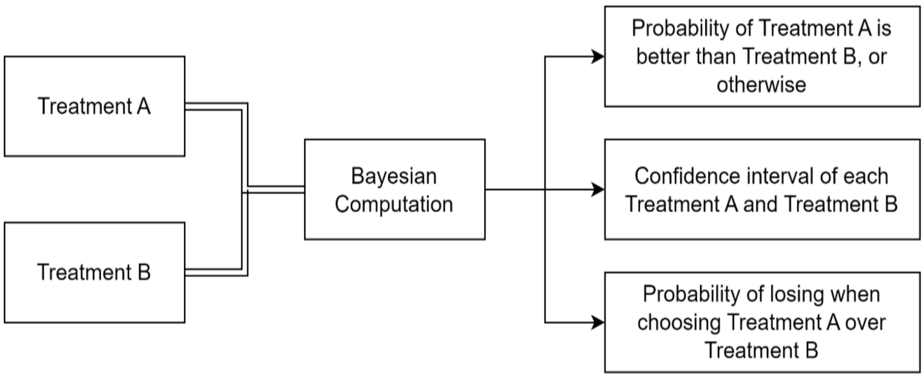

Figure 1. Illustration of Bayesian A/B Test Computational Process - Simplified (Robert, 2007)

b. Experimental designs
For this experimentation setting, we defined our hypothesis as follows:

H_0: $\mu A = \mu B$
(Exposure to other users' Aksi-Nyata report in the platform does not encourage users to submit their Aksi-Nyata report)

H_1: $\mu A \neq \mu B$
(Exposure to other users' Aksi-Nyata report in the platform encourage users to submit their Aksi-Nyata report

To conduct the A/B testing experiment, we divided the participants into two groups. One group received the new proposed design that consisted of a new page in the platform to enable users viewing others' submitted Aksi-Nyata report (refer to as Treatment A group); the second group received the legacy designs (refer to as Treatment B or control group). The users who participated in this experimentation consisted of new-users (i.e. users who already logged-in to the PMM before the experimentation period, but had not yet started learning in the "Microlearning"), and existing-users (i.e. users who had started learning at least one material). During the experimentation period, all logged-in users were randomly assigned to one of the groups. The randomization follows a binomial distribution with a probability of "success" p equal to 0.5, i.e. the chance of a user being assigned to a particular group is 50%. Once a user is assigned, the users would receive a corresponding treatment until the end of the experimentation period. The experimentation ran for two weeks and ended in Q1 2022, with 1.6 million users participating in the experiment.

3. Decision Making

The Bayesian test utilizes probability to determine whether a treatment is superior to another. To assess the effectiveness of the alternative design over the legacy designs, we established a 70% threshold for the goodness of fit. Since there is no standard benchmark for the Bayesian A/B testing threshold in the ministry, we selected this threshold through a team consensus as a reference for future experimentation.

4. Software

R open source software is employed to run statistical packages for the analyses. The `tidyverse` package version 2.0.0 (Wickham, 2022) was used as a foundation for loading other packages. The `dplyr` version 1.0.9 (Vaughan, et al, 2022) was used to make a dataframe from the experiment before continuing to the analysis. The `readxl package` version 1.4.0 (Bryan, et al, 2022) was used to import the excel dataset from experiment into R. The `ggplot2` package version 3.3.5 (Pedersen et al, 2021) was used to perform visualization from the result of the analysis. Further, the primary package to perform the A/B testing analysis was `bayesAB` from Portman (2021).

RESULTS AND DISCUSSION

1. Descriptive and Time Series Analysis

The results revealed that the age distribution of the sample was fairly diverse, along with the school level where the teacher teaches. The chosen sample of teachers has represented different age ranges from the general population. The majority of the participants fell into the 31 to 40 year age group, accounting for 35.5% of the total sample (Figure 2). Teachers in this age range are recognized as productive teachers and possess ample experience in the teaching and learning process in the classroom (Yen et al., 2015). Interestingly, 24.9% of the participants comprises teachers in the age range of 51 to 60 years old. Teachers in this range are senior teachers who will retire in the next 1 to 9 years, but in fact, a number of them are curious enough to try the new digital product offered by MoECRT. It suggests that age is not a dominant factor that motivates teachers to do self-determined learning, as discussed by Han et al (2016).

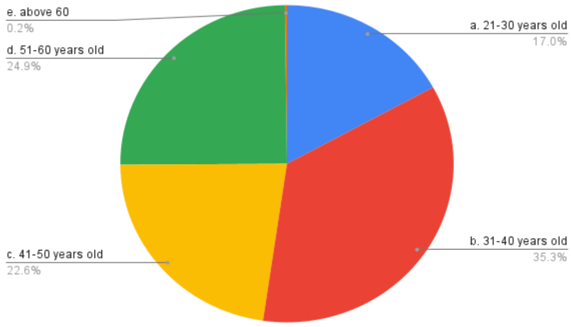

Figure 2. Distribution of the participants based on age range

The majority of participants in the study were elementary school teachers, comprising 50.5% of the total sample (Figure 3). This aligns with the overall distribution of teachers in Indonesia, where elementary schools have the highest number of educators. In contrast, special-needs schools (SLB) have the smallest proportion of teachers. The sample's age and school level distribution suggest that our sampled-participants are representative of the teacher population in Indonesia.

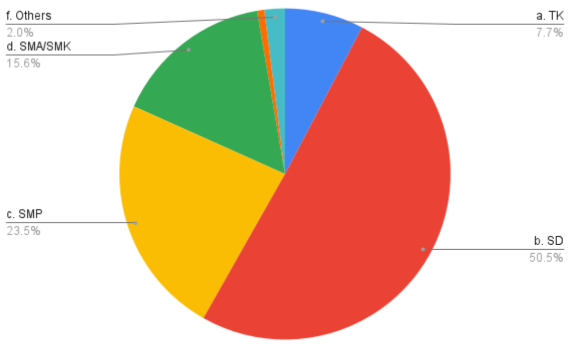

Figure 3. Distribution of participants based on school level

The data exploration continues to examine the number of users submitting Aksi-Nyata reports by employing time-series models. The ARIMA(1,2,0) model predicted that we would achieve 39.06% of the target in the upcoming 12 months after the experimentation period ended, if there was no intervention. However, by implementing the new proposed designs (as suggested by the results of A/B testing), we could achieve 96.83% of the predetermined target. It reflected that the improvement produced by this experimentation was 147%.

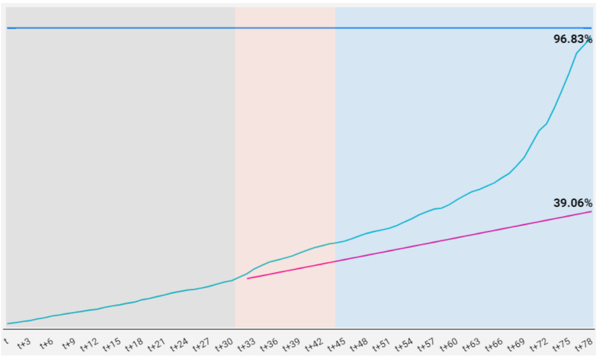

Figure 4. The movement of the conversion of users who submit Aksi-Nyata reports. Gray area: period before experimentation; Pink area: period during experimentation; Light blue area: period after experimentation. Purple line: predicted increase in the number of users submitting Aksi-Nyata report if there was no experiment; Green line: post-experimental increase in the number of users submitting Aksi-Nyata report.

Before the experiment was conducted, the rate at which users converted from watching videos to submitting Aksi-Nyata reports in the "Microlearning" platform was relatively low (2.89%). By the end of the experiment, we found that the alternative designs have the potential to increase conversion rate improvement by 115% in the final learning step of "Microlearning" from 2.89% to 6.22%. The difference between 2.89% and 6.22% is statistically significant with p-value < 0.00001 at 5% significance level.

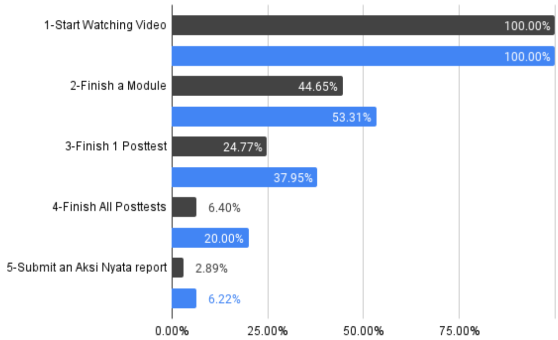

Figure 5: Conversion rate for the number of users for each activity funnel in "Microlearning", comparison for control (black color) and alternative (blue color) designs during the experiment

Ingarianti (2017) explained that one of the external factors influencing a person's career commitment in psychology includes the motivations offered by the organization (offered inducements), as well as work situations that support promotion opportunities. However, at this point of analysis, we learn that teachers and school principals are having self-motivation to complete learning materials provided in the "Microlearning" after seeing other's Aksi-Nyata report submission, although the MoECRT does not offer any incentives to users. We believe that this initiative will be a good foundation for finding the best support to enhance teachers' skills and capability.

Although we were very much interested in increasing the number of Aksi-Nyata report submissions, we also tracked the learning process of users before the submission, i.e. from watching learning videos (stage 1), completing reflections in the learning module (stage 2), to passing all the posttest in one topic (stage 4). The new designs, fortunately, also facilitated users in those learning stages, shown by the significant improvement of conversion rate in all stages (Figure 5). Such results suggest that our users might be more convenient in using the "Microlearning" platform in which the new design was implemented.

2. Bayesian A/B Testing

Having understood the importance of conducting an analysis based on group characteristics (The Signal, 2022), we decided to divide the analysis into two sections: one for new user groups (users who had not started learning any material when they were given either the alternative designs or the legacy ones during the experiment), and the other for existing users who had started learning before the experiment conducted.

Confidence-plots and probability-plots for each group of new- and existing-users are visualized in Figures 6 and 7, respectively.

Figure 6 shows that the posterior distribution of the alternative treatment is skewed towards the right compared to the control group. This indicates that, on average, the conversion rate of teachers submitting Aksi-Nyata reports for the alternative treatment is higher than that of the control group. Table 2 provides further evidence that the alternative treatment has a greater ability to convert teachers to submit their Aksi-Nyata reports, with a 1.43% improvement for the new-user group. Similarly, for the existing-user group, the alternative treatment yielded a conversion rate that was 0.41% higher than the control group.

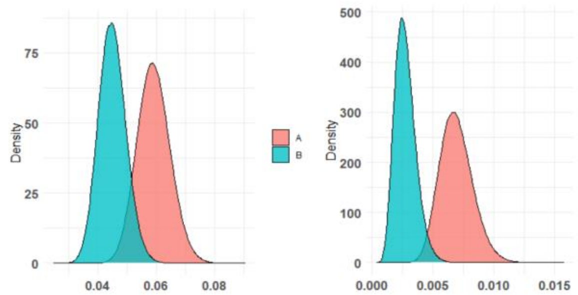

Figure 6(a)-left: Confidence-plot for new user groups, and **6(b)-right**: for existing user groups

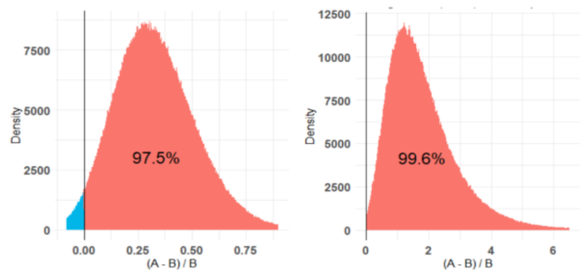

Figure 7(a)-left: Probability-plot for new user groups, and **7(b)-right**: for existing user groups. The vertical line separates the probability distribution of Treatment A and B

The alternative page (i.e. the new page design) has a 90% chance of performing better than the legacy designs in the general population, for both new- and existing-user groups (Figure 7). This finding is also supported by the fact that the expected loss of implementing alternative pages is close to zero.

We calculated the potential increase in the average conversion rate to provide a better insight for MoECRT regarding the effectiveness of the experimentation results, as commonly exercised in many companies during A/B Testing (Harvard Business Review, 2023). The potential increase of the average conversion rates for new- and existing-user groups is 32.06%, and 100%, respectively.

Table 2. Summary of Bayesian A/B Testing Analysis Results

User classification	Treatment	Average Conversion Rate	Difference in Conversion Rates	Probability A Better than B	Expected Loss of Choosing A	Potential Increase of Average Conversion Rate
New-users	A-alternative	5,89%	1,43%	97,5%	0,14%	32,06%
	B-control	4,46%				
Existing-users	A-alternative	0,66%	0,41%	99,6%	0,04%	164%
	B-control	0,25%				

CONCLUDING REMARK

Conclusion

The analytical evidence generated from Bayesian- based A/B testing strongly suggested that the new design should be implemented in the "Microlearning", as it brings enormous potential to better serve the users' experience in doing asynchronous and self-determined learning. The alternative designs succeeded in increasing the conversion rate of submitting Aksi-Nyata reports by 590% up to two months after its implementation. We hold the view that through consistent application of data-driven digital product development, such as the A/B testing method described in this article, we may unleash the optimum potential of the Merdeka Mengajar platform, which further may contribute to the irreversible transformations in the field of education in Indonesia.

Limitation

Despite the success of this experimentation in increasing the conversion rate of Aksi-Nyata report submissions, it is important to note the limitations encountered during the study. We did not take into account other variables, such as schools' level (e.g. elementary-, secondary-, and high-schools), age ranges and other potential factors that may confound the results. However, we believe that these limitations are unlikely to significantly alter the overall conclusions drawn from this study.

Recommendation

To enhance the analytical results presented in this paper, MoECRT may explore other variables related to teacher's attributes to gain a more comprehensive understanding of the impact of the "Microlearning" platform. Such insights may further support continuous improvement in the platform. In addition, A/B testing experimentation has proven to be an effective- and non-invasive method for digital product development. Therefore, we recommend that MoECRT keeps to utilize this approach for any future improvements in PMM or MoECRT's other digital platforms. By doing so, MoECRT can mitigate negative risks associated with changes to digital products and make more scientific- and data-driven decisions.

ACKNOWLEDGEMENT

This study was a join-effort research project (start from the planning to the implementation) that involved multiple team members, from Product Manager (Raden Achmad Zulfikar Hermawan), Product Designers and Qualitative Researchers (Moga Adiangga Aliffian and Lathifah Halim), Software Engineers (Hermanet Lay, M Zahid Rausyanfikri, and Riandy Dimas Banimahendra) to Content Team (Putri Rizki Dian Lestari and Aghnia Mega Safira). We are also thankful and grateful enough to have enormous trust and support from the MoECRT, especially stakeholders at Satuan Kerja Guru dan Tenaga Kependidikan (Prof. Dr. Nunuk Suryani, M.Pd., Dr. Iwan Syahril, Ph.D., Dr. Praptono, and Dr. Drs. Rachmadi Widdiharto, M.A.), to

continuously improve the digital platforms used to support flagship programs in the Ministry.

REFERENCES

Bryan, J., Wickham, H., Posit, P. B. C., Kalicinski, M., Valery, K., Leitienne, C., Colbert, B., Hoerl, D., Miller, E. 2022. Computer Software. [Online]. Available: https://cran.r-project.org/web/packages/readxl/index.html.

Harvard Business Review. 2017. A Refresher on A/B Testing. Cambridge. Last accessed on April 4, 2023 on page https://hbr.org/2017/06/a-refresher-on-ab-testing

Han, J., Yin, H., Boylan, M. 2016. Teacher Motivation: Definition, Research Development and Implications for Teachers. Cogent Education Vol. 3 Issue 1.

Ingarianti, T. M. 2017. Faktor-faktor yang Mempengaruhi Komitmen Karir. Jurnal Ilmiah Psikologi Terapan Universitas Gadjah Mada 5(2): 202-209.

Kemendikbud Ristek RI. 2022. Implementasi Kurikulum Merdeka. Jakarta. Last accessed on December 22, 2022 on page https://kurikulum.gtk.kemdikbud.go.id/.

Kominfo RI. 2020. Hasil Survei PBB, e-Government Indonesia Naik Peringkat. Jakarta. Last accessed on December 22, 2022 on page https://www.kominfo.go.id/content/detail/30024/hasil-survei-pbb-e-government-indonesia-naik-peringkat/0/artikel.

Pedersen, T. L., Wickham, H., Chang, W., Henry, L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Posit, P. B. C. 2021. Computer Software. [Online]. Available: https://cran.r-project.org/web/packages/ggplot2/index.html.

Portman, F. 2021. Computer Software. [Online]. Available: https://cran.r-project.org/web/packages/bayesAB/index.html.

Puza, Borek. 2015. Bayesian Methods for Statistical Analysis. Canberra: The Australian National University eView.

Robert, C. P. 2007. The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation Second Edition. Paris: Springer Science & Business Media.

Samaniego, F. J. 2010. A Comparison of the Bayesian and Frequentist Approaches to Estimation. Paris: Springer Science & Business Media LLC.

The Signal. 2022. Users: New vs Returning. San Francisco. Last accessed on April 8, 2023 on page https://mixpanel.com/blog/new-vs-returning-users/.

UN Department of Economic and Social Affair. 2022. UN E-Government Survey 2022 The Future of Digital Government. Jakarta. Last accessed on December 22, 2022 on page https://desapublications.un.org/publications/un-e-government-survey-2022.

Vaughan, D., Wickham, H., Francois, R., Muller, K. 2022. Computer Software. [Online]. Available: https://cran.r-project.org/web/packages/dplyr/index.html.

Wei, William W. S. 2006. Time Series Analysis: Univariate and Multivariate Methods Second Edition. Philadelphia: Pearson Education, Inc.

Westland, J. C. 2022. A Comparative Study of Frequentist vs Bayesian A/B Testing in The Detection of E-commerce Fraud. Journal of Electric Business & Digital Economics Vol. 1 No. 1/2, pp. 3-23.

Wickham, H. 2022. Computer Software. [Online]. Available: https://cran.r-project.org/web/packages/tidyverse/index.html.

Xue, X., Fu, Y., Xia, S., Li, X. 2022. A/B Test Research on the Selection of Push Scheme in Network Precision Marketing. ISCTT 2022; 7th International Conference on Information Science, Computer Technology and Transportation, Xishuangbanna, China, 2022, pp. 1-6.

Yen, H. S., Lim, H. E., Campbell, J. K. 2015. Age and Productivity of Academics: A Case Study of a Public University in Malaysia. Malaysian Journal of Economic Studies 52(1): 97-116.