



Mode of Delivery: Paper-pen versus Mobile-Based Testing. Are they Comparable?

Mitchelle R. Juaban^{1*} and Ryan Rommel Dominguez²

^{1*2}Foreign Languages Department, Liberal Arts Faculty, Rajamangala University of Technology Srivijaya, Songkhla 90000, THAILAND

*Corresponding Author E-mail: *mitchelle.j@rmutsv.ac.th

(Article history) Received: 2026-08-06, Revised: 2026-08-12, Accepted: 2026-08-23

Abstract

The ubiquity of mobile technologies in education has brought concerns regarding comparability between paper-based (PBT) and mobile-based testing (MBT) formats. This study examined: (1) the effects of test mode on student performance and (2) the score equivalence obtained via paper-based and mobile-based testing. Participants (126) from a university in the south of Thailand completed two quantitative within-subject counterbalanced tests administered via PBT and MBT through the participants' personal smartphones with a two-week interval between administrations. Descriptive statistics, paired-samples t-tests, Pearson product-moment correlation, and Spearman rank-order correlation were used to analyze the data. Findings showed a statistically significant difference between modes favoring mobile testing indicating a strong mode effect. Moreover, correlation analyses revealed negligible relationships between the two modes, suggesting weak score equivalence and limited rank-order consistency. This research contributes to the growing literature on assessment formats by providing evidence on mode effect and interchangeability. The study provides evidence that MBT may influence student performance and score interpretation; and underscore the necessity for careful validation.

Keywords: Mobile-based Testing (MBT), Mode of delivery, Paper-based Testing (PBT), Score equivalence, English language assessment

1. INTRODUCTION

Post-COVID-19, online testing (frequently referred to as computer-based Test, CBT) has become a widely, if not preferably, used form of assessment. Not surprisingly so, as many have experienced its benefits including: the automated marking, the ease of editing, the comfort of not having to walk to the printing shop and collate papers one by one, and most importantly reduced testing cost. Or, if not all these, then just to continue using the technology to get the most of what it's paid for. In reality though, mobile-based testing (also known as MBT) seems the most practical recourse considering time constraints, insufficient school computers, and availability of mobile devices with internet connections. According to a report made by Dr. KC Dipendra and Dr. Mongkhonvanit Jomphong (2022) to the International Telecommunication Union, there is a 1-to-17 computer to student ratio Thailand, and that 75% of households depend on mobile broadband to access the internet. With the ubiquity of mobile devices, using it for possible testing means increases. However, is it comparable to paper-pen assessment?

After the pandemic, periodic examinations were administered through online testing - specifically mobile phones. This is done via the Learning Management System (LMS) and Google Forms - the versions available did not include lockdown features that could restrict access to browser menus, toolbar options, other applications, right-click functions, keyboard shortcuts, or task switching. After some time, the testing format was modified: midterm examinations continued to be conducted through mobile-based testing, while final examinations were administered using the paper-and-pen format or the paper-based testing (PBT, hereafter). Quizzes and unit tests were oftentimes conducted via MBT.

Teachers, however, have observed that: students score substantially higher on MBT. In fact, memes depicting students' performance discrepancy had made rounds on the internet. On the surface they seem funny but on a closer look, particularly on the assessment level, many critical questions loom.

With this current trend, this study seeks to investigate if delivery mode affects students' academic performance, as well as determine if both assessment deliveries are comparable in terms of equivalence and rank-order. Counterbalanced English achievement tests were constructed to obtain the data. This research provides theoretical and practical significance on mode effect and score equivalence across the two types of assessment delivery.

2. RELATED LITERATURE

The vast advancements in mobile technologies have led to its adoption in education as an alternative tool for testing. Mobile-based testing refers to assessments completed using smartphones or other mobile devices, allowing learners to access tests anytime and anywhere. With the growing popularity of MBT in educational assessment practices, its effects on student performance and if the scores obtained from it are comparable with the PBT in terms of equivalence are being investigated by researchers. However, while there are numerous papers tackling comparability between PBT and CBT, studies on the equivalence of PBT and MBT are scarce, especially for EFL and English language learning contexts.

In alignment with the current study, two major themes will be examined on PBT and MBT: (1) the effects of testing mode on student performance and (2) the equivalence of scores across testing modes.

2.1 Effects of PBT and MBT on Student Performance

Researches on the effects of the employment of MBT as an assessment tool on student performance have revealed mixed results. While some reported no significant differences between PBT and MBT assessments, others suggest that MBT may enhance student performance.

A research conducted by Whattananong (2004) was one of the earliest studies investigating MBT. He compared PBT with two forms of MBT: (1) students listened to questions and answers were submitted via SMS, (2) students viewed questions on a screen and responded through their mobile phones. This study found no significant differences in scores obtained between PBT and MBT. Having no negative effects on performance he proposed that MBT could be an alternative assessment instrument.

In the same way, Gordon's (2015) investigation on PBT and MBT among Korean university students enrolled in English as a Foreign Language (EFL) courses revealed no statistically significant differences, indicating that MBT garnered comparable levels of achievement as PBT. In this study, students completed an equivalent multiple choice test in both PBT and MBT. For the MBT, Socrative mobile application that students can download to their smartphones was used. Gordon additionally reported that there was a general preference for MBT among the participants and that they did not perceive it as more difficult than PBT.

Contrarily, a positive influence between MBT and student performance was found by Nikou and Economides (2013). In the study, 203 students at the Department of Economic Sciences of a Greek university were randomly assigned a 30-item multiple choice test to PBT, CBT, and MBT conditions. The participants took the test in only one mode; and were not rotated through the different modes. The results revealed that students scored significantly higher in MBT than in the PBT. This was accounted for by familiarity with smartphones, reduced anxiety, and increased engagement which may have contributed to the improved performance in the mobile testing condition.

Moreover, Nikou (2018) reviewed 43 studies on MBT, and has given additional support for the potential benefits of MBT. From this review he concluded that MBT often enhances learner motivation, engagement, and achievement. He further mentioned that mobile devices provide flexible assessment practices and encourage positive attitude toward testing. However, he noted that the effects of MBT may vary with conditions like context, subjects, and assessment designs.

Abdillah (2016), similarly reported that students have positive perceptions towards MBT. They found this mode convenient, efficient, and accessible. The study mainly focused on student perceptions, the findings however show that increased comfort and familiarity with this technology may have affected testing behavior and potentially the performance outcomes.

On the whole, the data on MBT suggests that the mode may either be comparable to or slightly higher than the achievement outcomes observed in PBT. But factors such as test design, device characteristics, student familiarity with technology, and testing conditions may greatly influence the magnitude and direction of these effects.

2.2 Equivalence Between PBT and MBT

Aside from the mean score, determining whether PBT and MBT are equivalent measures of the same construct is another important aspect to look into. Equivalence refers not only to similar average scores but also to the consistency of score interpretations across modes. There are a number of studies supporting the equivalence of PBT and MBT.

Whattananarong (2004) reported strong correlations between PBT and MBT, with correlation coefficients ranging from .887 to .910. These indicate that students had similar performance rankings regardless of the testing mode, suggesting a high degree of score equivalence. This finding was corroborated by Gordon (2015) who further suggested that PBT and MBT could be used interchangeably in EFL contexts. The lack

of significant differences in the mean scores was interpreted as comparability between PBT and MBT in measuring the underlying abilities.

Computer-based assessment has reached similar conclusions. Studies by Bergstrom (1992), Boo and Vispoel (2012), Bugbee (1996), Chin and Donn (1991), Choi and Tinkler (2002), and Evans, Tannehill, and Martin (1995) generally reported minimal differences in scores obtained via PBT and MBT. These studies proposed that mode of delivery to a large extent does not affect performance when conducted in controlled settings.

On the other hand, researchers such as Dorans, Pommerich, and Holland (2007) warned against assuming equivalence across modes. They argued that empirical validation through linking, scaling, or equating procedures may be required as scores from different testing modes are not inherently equivalent. For them, score equivalence relies heavily on test design, administration conditions, and contextual factors. In the same way, Yu and Iwashita (2021) underscore the significance of not assuming comparability even when mean differences are minute. Seemingly insignificant differences like screen presentation, navigation, and test-taker interaction may have a slight impact on performance that affect the interpretation of scores.

The findings of Nikou and Economides (2013) also cast doubts on the equivalence of PBT and MBT. Even though the mean scores of the participants were significantly higher in MBT but by reason of lack of rank-order consistency examination across modes; it remains unclear if the scores were reflection of students' improved learning outcomes or mode-related factors unrelated to the construct being measured.

Nowadays, as the adoption of digital assessment in education is rapidly growing, it is essential to examine not only the mean score differences but also the score comparability and validity across delivery modes. Mobile devices are becoming a ubiquitous aid in learning, thus establishing equivalence between PBT and MBT is critical if we are to extend its usefulness to educational assessments.

3. METHODOLOGY

3.1. Research Questions

- 1. Does test mode significantly affect students' English test performance?*
- 2. Are paper-based and mobile-based online test scores comparable in terms of rank-order consistency and score equivalence?*

3.2. Research Participants

One hundred twenty-six non-English majors from Business Administration and Industrial Education were this study's research participants. They were enrolled in the course Communication Skills in English, thus making this population a convenient participant for the study. They first completed a paper-pen test and after a two-week interval, they completed a counterbalanced parallel form via online delivery.

3.3. Research Instruments

The instrument consisted of two counterbalanced tests (version A, paper-pen and version B, online test). Each test comprised 4 sections: Vocabulary in context, Agreement structures, Present perfect vs. past tense usage, and Reading comprehension. Originally a 20-item test was later cut down to 14 to achieve an acceptable internal consistency, Cronbach's $\alpha = .71$.

To maintain construct comparability, the tests were designed to measure similar constructs and contained identical item structure. Although they differ in item and lexical contents, the cognitive demands were equivalent.

Version A was administered in the classroom via paper-pen mode while Version B via Google Forms and conducted in the classroom through participants' mobile phones. Both tests were supervised by the teacher and had the same time limit, instructions, and classroom conditions. It should be noted that during the online delivery, no technical, technological, or log-in detail problems were encountered.

3.4 Data Analysis

Descriptive statistics like mean and standard deviation were used for both modes to identify the mean difference. A paired-samples t-test was conducted to examine whether the difference between PPT and mobile-based test scores was statistically significant. And to examine the consistency across modes the Pearson's r and Spearman's ρ were employed.

4. FINDINGS

4.1. RQ 1: Is there a statistically significant difference in students' English test performance between paper-based and mobile-based online modes of delivery?

Table 1: Paper-pen and Mobile-Based Testing means and standard deviations

Mode	Mean (M)	Standard Deviation (SD)
Paper-pen	6.5	2.9
Online	13.2	2.8

Table 2: Delivery Modes Paired-samples t-Test Results

Parameter	Value
P-value	0
t	18.2256
Sample size (n)	126
Average of differences ($\bar{x}_d$)	6.6587
SD of differences (S_d)	4.101
Normality p-value	0.002454
A priori power	0.9998
Post hoc power	1
Skewness	-0.4205
Skewness Shape	 Potentially Symmetrical (pval=0.051)
Excess kurtosis	-0.4794
Kurtosis Shape	 Potentially Mesokurtic , normal like tails (pval=0.263)
Outliers	

Table 1 shows the mean and standard deviation of PPT ($M=6.5$, $SD=2.9$) and mobile-based test ($M=13.5$, $SD=2.8$), while **Table 2** presents the results of the paired t- test. It indicated that there is a significant large difference between participants' PPT and mobile-based test scores, $t(125) = 18.2$, $p < .001$.

At a closer look, the magnitude of the difference (Mean difference = 6.7 points, Pooled $SD \approx 2.85$), an approximate Cohen's $d \approx 2.3$ suggests an extremely large mode-of-delivery effect.

4.2. RQ 2: Are paper-based and mobile-based online test scores comparable in terms of rank-order consistency and score equivalence?

Table 3: Delivery Modes Correlation Results

Parameter	Value	Parameter	Value
Pearson correlation coefficient (r)	-0.002636	Spearman's rank correlation coefficient (r_s)	-0.03892
r^2	0.000006949	r^2	0.001515
P-value	0.9766	P-value	0.6653
Covariance	-0.0221	Covariance	-50.916
Sample size (n)	126	Sample size (n)	126
Statistic	-0.02935	Statistic	-0.4337

Table 3 presents the results of the Pearson's correlation ($r(124) = .00264$, $p = .977$) and the Spearman's rank correlation ($r(124) = .0389$, $p = .665$). Both indicated that there is a non-significant, very small negative relationship between Paper-pen and online.

5. DISCUSSION

5.1. RQ1: Does test mode significantly affect students' English test performance?

The participants' mean scores in MBT is by far higher than the PBT. Considering that this research design has the same construct (English communication skills), same students, different modes (paper vs online), counterbalanced content, and a 2-week interval that minimizes memory effects, the large increase in scores implies that mode delivery exerts a strong influence on the test performance. This finding aligns with Nikou and Economides (2013) where participants modestly scored higher in MBT (1.827 points mean difference) than in PBT and CBT (1.19 points mean difference). Similar to the current study, the whole test was completed using a mobile phone, but Nikou's and Economides' MBT was conducted via a dedicated Android Quiz Application which had a highly controlled interface, questions were presented one at a time, limited distractions, and most likely prevented easy switching among resources.

Although many researches suggest that CBT and PPT yield equivalent results such as early studies made by Mead and Drasgow (1993) - comparable results between CBT and PBT with minimal differences overall in participants' cognitive ability test; and corroborative studies made by Bergstrom, 1992; Boo & Vispoel, 2012; Bugbee, 1996; Chin & Donn, 1991; Choi & Tinkler, 2002; Evans, Tannehill, & Martin, 1995, Wang et al., 2008 supporting score equivalence between CBT and PBT citing similarity in mean scores, and small and statistically insignificant differences. On a closer look though, the interpretations indicate that such equivalence is conditional - that is under a strictly controlled environment.

On the other hand, other studies observe the two modes are not always directly interchangeable (Godwin and Godwin, 1999; Pommerich & Burden, 2000). The significant mean difference observed in this study provides strong empirical support for this position, indicating that online testing may systematically raise scores or alter performance conditions. This is backed up by Bachman and Adrian (2010, cited by Clark-Gareca, 2010), who argued that test performance is influenced not only by language ability but also by test method facets. In this case, the online mode may have brought in additional variables like the device used, testing environment, students digital skills, or possible external access.

In a real-world testing scenario (real classroom + mobile phone) as in this present study's case such equivalence may not hold. MBT and CBT are both digital technologies, and although many studies found comparability between PBT and CBT in a well-controlled environment, not the same can be said about PBT and MBT. The differences in size, interface, user experience, and other technological distinctness may affect performance. So, the impact of these differences need to be examined since mobile devices are fast becoming a ubiquitous tool in education.

5.2. RQ2: Are paper-based and mobile-based online test scores comparable in terms of rank-order consistency and score equivalence?

The present study shows lack of significant correlation and low rank-order stability between modes which raise concerns regarding score comparability. This

suggests that students who performed well in one mode may not necessarily perform well in the other.

Contrarily, Whattananarong (2004) reported a strong correlation between PBT and MBT. The likely explanation for this contrasting finding is the considerably different implementation of the MBT. Whattananarong's study used the SMS feature of the phone to submit answers. Questions were either read from a projector screen or played from an audio recorder. The current study, on the other hand, employed smartphones as a complete testing platform where questions and answers were presented; and students had to read, navigate, and respond directly to it. This may have presented mode-related factors such as screen reading, navigation, interface interaction, and potential access to online resources that could have influenced the large mode effect and the lack of score equivalence between PBT and MBT. Further, differences in the research design could have also influenced the equivalence result. Whattananarong employed a control and experimental design, whereas the current study used within-subject counterbalanced design and the same group of participants took both the PBT and MBT, which may keep individual differences under control.

Dorans, Pommerich, and Holland (2007) revealed that scores from tests administered through different modes are not inherently comparable. The lack of correlation in the present study indicates that: the test scores were affected by delivery modes and it changes the relative ranking of students. Thus, PBT and MBT are not comparable in terms of rank-order stability and score equivalence.

5.3 Construct-irrelevant Variance

The combination of large mean difference and non-significant correlation, despite the counterbalanced construct, is a strong indication that factors unrelated to language ability may have influenced test performance. According to Messick (1989) construct validity is compromised when unrelated variables affect the test scores. Looking at the present study, the MBT and the PBT may not measure the construct of English Communication in the same way. This is further supported by Kane (2006, as cited by Chapelle 2022), who reported that for test scores to be valid they must be supported by evidence across testing conditions. A generalized interpretation of the scores obtained across the modes may not be possible, thereby weakening the validity argument. Hence, in a blended assessment it can be said that scores obtained may not measure the intended construct in absolute sense unless careful validations have been done prior.

5.4 Practical Implications

In the classroom the reality is, employing MBT is an easy and convenient go-to especially if faced with heavy workload, time and deadline constraints. Another reality is that not all schools and universities are equipped with facilities that can handle a strictly controlled CBT environment, thus most resort to classroom and mobile phone set up. Looking at the current situation, it would seem that this multi-dollar technology is a fixture that is shaping the present's and future's educational arena.

In light of this, the decision by educational institutions and its stakeholders to use digital technologies in terms of what, when, and in what manner is critical. The

notion that it is there for the taking should be re-thought as its impacts on educational settings are not straight-forward.

The findings of this current study are relevant considerations for educators and educational institutions employing mix-method assessment or for those who are thinking of transitioning to pure CBT or MBT. Careful preparations and validation are needed; and in doing so will aid in weighing options that ensure fair and appropriate assessment.

Classrooms that lack facilities for a stricter CBT controlled setting and utilize students' mobile phones instead may have to think twice of using this assessment mode. Switching between tabs and apps during exams, browser search, and peer collaboration (Huynh Thi Trang, et, 2022) are some forms of dishonesty during online tests. Through an online survey, a little over 60% of this current study's participants revealed that they perceived it easier to cheat during online tests. This may be the case if the conditions for a conducive CBT/ MBT are not met: one invigilator, classroom setting, and using students' personal mobile phones.

Thus, the decision to employ MBT or mixed methods should come with intensive prior preparations of test constructions, test validity and reliability for both modes of delivery. Furthermore, the following can also be done to ensure fair assessment: (1) adapting test items and layout for mobile-based testing rather than copying PBT into online format and (2) consider a separate benchmark for each mode.

6. CONCLUSION, LIMITATION, RECOMMENDATION

Assessment is an essential part in education, thus choosing the appropriate mode of delivery is critical. In the classroom, it is easy to opt for MBT over PBT for both convenience and practical reasons. This study shows that the scores obtained through MBT and PBT are not equivalent, thus they are not interchangeable. Therefore; prior preparations in terms of test construction, test reliability and validity are needed.

This study has three major limitations. One is limited generalizability. The number and representation of participants - 126 and EFL context, non-English majors, university students - are minimal in quantity and scope to draw generalization. Two, the two-week interval between tests may be too close to address memory effects. Lastly, the test, although counterbalanced may be limited in length and reliability. Results of future research that addresses the mentioned limitations would be significant contributions to the limited literature on PBT and MBT equivalence.

Further, it is recommended that factors impacting mode effects, mitigating factors, MBT dishonesty preventions, and elements that may contribute to MBT reliability and validity be considered for future research.

REFERENCES

- Abdillah, L.A. (2016). Ujian Online Mahasiswa Ilmu Komputer Berbasis Smartphone. *ArXiv, abs/1701.06428*.
- Bergstrom, B. (1992). Ability measure equivalence of computer adaptive and pencil and paper tests: A research synthesis. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. Retrieved from <https://eric.ed.gov/?id=ED377228>
- Boo, J., Vispoel, W. (2012). Computer versus Paper-and-Pencil Assessment of Educational Development: A Comparison of Psychometric Features and Examinee Preferences. *Psychological reports*. 111. 443-60. 10.2466/10.03.11.PR0.111.5.443-460.
- Bugbee, A. C. (1996). The equivalence of paper-and-pencil and computer-based testing. *Journal of Research on Computing in Education*, 28, 282-299. Retrieved from Education Research Complete, Accession Number: 9605221096
- Chapelle, C. (2022). Argument-Based Validation in Testing and Assessment. *Sage Research Methods*. DOI: <https://doi.org/10.4135/9781071878811>
- Chin, C. H. L., & Donn, J. S. (1991). Effects of computer-based tests on the achievement, anxiety, and attitudes of grade 10 science students. *Educational and Psychological Measurement*, 51, 735. doi: 10.1177/0013164491513025
- Choi, S. W., & Tinkler, T. (2002). Evaluating comparability of paper-and-pencil and computer based assessment in a K-12 setting. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA. 148.
- Coiro, J. (2007). Exploring changes to reading comprehension on the Internet: Paradoxes and possibilities for diverse adolescent readers. (Doctoral dissertation). Retrieved from <http://ctell1.uconn.edu/coiro/coiropencer.pdf>
- Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.). (2007). Linking and aligning scores and scales. *Springer Science + Business Media*. <https://DOI.org/10.1007/978-0-387-49771-6>
- Clark-Gareca, Beth. (2010). Language Assessment in Practice. Working Papers in Applied Linguistics and TESOL. 10. 10.7916/D8CV4HB8.
- Evans, L. D., Tannehill, R., & Martin, S. (1995). Children's reading skills: A comparison of traditional and computerized assessment. *Behavior Research Methods, Instruments, & Computers*, 27(2), 162-165. DOI: 10.3758/BF03204723
- International Telecommunication Union (2022). *E-learning in Thailand: Mapping the digital divide*. Retrieved from <https://www.itu.int/hub/publication/d-pref-them-28-2022/>

- Goodwin, L. D., & Goodwin, W. L. (1999). Measurement myths and misconceptions. *School Psychology Quarterly*, 14(4), 408–427. <https://doi.org/10.1037/h0089017>
- Gordon, A. (2015). “Paper Based Testing vs. Mobile Device Based Testing in an EFL Environment: What’s the Difference?”. *Culminating Projects in English*. 38. https://repository.stcloudstate.edu/engl_etds/38
- Huynh Thi Trang, Tran Kha Han, Phuong Thi Cam Loan. (2022). Cheating Behaviours in Online Exams, *International Journal of Science and Management Studies (IJSMS)*, v5(i6), 32-37. DOI: 10.51386/25815946/ij sms-v5i6p103
- Mazzeo, J., & Harvey, A. L. (1988). The equivalence of scores from automated and conventional educational and psychological tests: A review of the literature (College Board Rep. No. 88-8, ETS R RNo. 88-21). Princeton, NJ: Educational Testing Service. <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2330-8516.1988.tb00277.x>
- Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114 (3), 449–458. <https://doi.org/10.1037/0033-2909.114.3.449>.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York, NY: American Council on Education and Macmillan.
- Naznean, A. (2021). Cheating during Online Examinations– Literature Review. *Journal of Pedagogy*. LXVIII (2):7-21 DOI:10.26755/RevPed/2021.2/7
- Nikou, S., Economides, A. (2013). Student achievement in paper, computer/web and mobile based assessment. *CEUR Workshop Proceedings*. 1036.
- Nikou, S., Economides, A. (2018). Mobile-based assessment: A literature review of publications in major referred journals from 2009 to 2018. *Computers & Education*, vol. 125, 2018, Pages 101-119, ISSN 0360-1315, <https://doi.org/10.1016/j.compedu.2018.06.006>.
- Pommerich, M., & Burden, T. (2000). From simulation to application: Examinees react to computerized testing. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA. <https://eric.ed.gov/?id=ED442813>
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68(1), 5-24.
- Whattananarong, K. (2004). An experiment in the use of mobile phones for testing at King Mongkut’s Institute of Technology North Bangkok, Thailand. In *Proc. International Conference on Making Education Reform Happen: Learning from the Asian Experience & Comparative Perspectives*, Bangkok, Thailand.

- Yu, W., Iwashita, N. (2021). Comparison of test performance on paper-based testing (PBT) and computer-based testing (CBT) by English-majored undergraduate students in China. *Lang Test Asia* **11**, 32. <https://doi.org/10.1186/s40468-021-00147-0>