

Introduction

In the field of speech-language pathology, standardized language tests are assessment tools that measure oral language and articulation skills. They also inform the diagnosis of speech and language disorders in Cantonese-speaking children. These tests are supported by robust normative data and have been shown to possess strong psychometric properties, yet their clinical acceptability remains unexplored (Wong et al., 2026). Clinical acceptability refers to “*a multi-faceted construct that reflects the extent to which people delivering or receiving a healthcare intervention consider it to be appropriate, based on anticipated or experienced cognitive and emotional responses to the intervention*” (Sekhon et al., 2017, p. 4).

Sekhon et al. (2017) proposed the Theoretical Framework of Acceptability (TFA), which guides the assessment of the clinical acceptability of healthcare interventions. It covers seven constructs, including *Affective Attitude*, *Burden*, *Ethicality*, *Intervention Coherence*, *Opportunity Costs*, *Perceived Effectiveness*, and *Self-Efficacy* (Sekhon et al., 2017). There are various forms of clinical acceptability, including prospective, concurrent, and retrospective acceptability. Specifically, retrospective acceptability refers to acceptability after participating in the intervention. Assessing retrospective acceptability is crucial since it captures lived experience rather than anticipated emotions. Given that the absence of clinical

acceptability may lead to underutilization of well-developed tools, the use of ineffective clinical methodology, or misdirected future development, it is necessary to evaluate clinical acceptability from the perspective of intervention providers. These findings are essential for guiding future development of standardized language tests.

The Hong Kong Test of Preschool Oral Language (Cantonese) (TOPOL; Child Assessment Service, 2019) is one of the standardized language tests for Cantonese-speaking preschool children between the ages of 3 years 0 months and 5 years 11 months. While it demonstrates acceptable psychometric properties and diagnostic accuracy (Child Assessment Service, 2019), it may be impractical in certain clinical settings due to its lengthy administration time (45-60 minutes) (Child Assessment Service, 2019), given that an assessment session lasts approximately 60 minutes for children in Hong Kong (Wong et al., 2023). Therefore, low clinical acceptability of the TOPOL may be expected for *Affective Attitude*, *Burden* and *Self-Efficacy*. In addition, most test items are delivered along with picture books. A lack of play-based assessment may be less attractive to preschool children, which may compromise *Affective Attitude*. As administrators of standardized language tests, SLPs' attitudes towards the test are crucial as negative perceptions may reduce their willingness to use it, leading to a waste of resources.

This study aims to examine the retrospective clinical acceptability of the TOPOL

from the perspective of speech-language pathologists (SLPs), who are the suggested intervention providers. Guided by the TFA, the associations between clinicians' *Affective Attitudes, Burden, Ethicality, Intervention Coherence, Opportunity Costs, Perceived Effectiveness, Ethicality, and Self-Efficacy* with the TOPOL will also be explored. These factors enable an in-depth understanding of the test's applicability in real-world settings and guide future development of standardized tests. Two research questions were developed:

Research Question 1 (RQ1): What is the overall clinical acceptability of TOPOL?

Research Question 2 (RQ2): Which TFA construct(s) are most strongly associated with overall clinical acceptability?

Methodology

Study Design

An online survey was designed based on the seven constructs of the TFA, including *Affective Attitude, Burden, Ethicality, Intervention Coherence, Opportunity Costs, Perceived Effectiveness, and Self-Efficacy*. There were 102 items in the questionnaire, which were expected to be completed in 15-20 minutes to minimize the possibility of low-quality and unreliable answers or survey dropout due to

respondents' fatigue (Sharma, 2022). For each of three standardized tests, 28 questions required ratings and 6 questions were open-ended questions. A 5-point Likert scale was used, in which SLPs respondents provided ratings on a scale of 1 (strongly disagree) to 5 (strongly agree). The reliability of the questionnaire, based on the mean scores of the seven TFA constructs, demonstrated good internal consistency (Cronbach's $\alpha = .83$; Tavakol & Dennick, 2011).

This study is part of a larger project investigating the clinical acceptability of the three Cantonese standardized language tests and was approved by the Human Research Ethics Committee (HREC reference number: EA260311). The present study focuses on exploring the clinical acceptability of TOPOL.

Participants

SLPs, who have experience with standardized tests for Cantonese-speaking preschool children aged 2-6 years and are fluent in Cantonese and written Chinese were included. Exclusion criteria included SLPs whose experience was limited to standardized assessments for non-Cantonese-speaking children or adults in either local or overseas settings, and or speech-language pathology students. These criteria were applied to ensure that participants had sufficient experience with standardized assessment for Cantonese-speaking preschool children to provide meaning acceptability ratings.

Procedure

The online survey was developed and administered on Qualtrics (Qualtrics, 2026). A recruitment poster containing the survey link was disseminated through social media and distributed via email to local non-governmental organizations (NGOs) and private speech-language pathology clinics that provided preschool speech-language pathology services. Responses were collected over an approximately two-month period between June to August 2026.

Data and Statistical Analyses

Quantitative responses were analyzed using descriptive and inferential statistics, with Likert-scale ratings treated as interval data. Qualitative data were not analyzed in this study.

The RQ1, the overall clinical acceptability of the TOPOL, was addressed using descriptive statistics, which was illustrated by (a) the mean rating of the single overall acceptability item (Item 8.1: *Overall, how acceptable was this intervention to you?*) and (b) the mean rating across 7 TFA constructs (i.e., mean of Item 1.1 to Item 7.4, which contain questions across the seven TFA constructs). For the latter, since the constructs vary in the number of questions, equal distribution to mean should be established. The mean score for each of the seven constructs was first calculated, and

the overall mean across the 7 constructs was then computed. To determine whether the observed mean ratings differed significantly from the neutral reference value of 3, a two-sided alternative hypothesis was set: the mean rating of clinical acceptability differs from 3. A Wilcoxon signed-rank test and a two-tailed one-sample *t*-test were conducted to compare the mean rating of overall acceptability item and the mean ratings across 7 constructs against the neutral score of 3, respectively. All analyses were conducted on IBM SPSS 31.0.2.0 (IBM Corp., 2025).

Results

Thirty-eight responses were analyzed. The mean score across 7 constructs was *M* = 4.13 (*SD* = .37, 95% CI [4.01, 4.25]). For the single overall acceptability item, it yielded *M* = 4.13 (*Mdn* = 4.00, *SD* =.58, 95% CI [3.94, 4.32]). In addition, 89.5% (34/38) of respondents rated the TOPOL as either acceptable or strongly acceptable, whereas 10.5% (4/38) of respondents selected neutral.

Table 1

Descriptive Statistics of Across 7 Constructs and Single Overall Acceptability Item (item 8.1)

Variable	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>SE</i>	Min	Max
Across 7 constructs	38	4.13	—	.37	.06	3.36	5.00

(item 1.1-
7.4)

Single	38	4.13	4.00	.58	.09	3.00	5.00
overall							
acceptability							
item (item							
8.1)							

Note. M = Mean; Mdn = Median; SD = Standard Deviation; SE = Standard Error;
Min = Minimum; Max = Maximum.

The mean rating of the single overall acceptability item (item 8.1) was compared with a neutral score of 3. Since the ratings were not normally distributed, as indicated by the Shapiro-Wilk test ($p < .001$) and the visual inspection of the Q-Q plot, a Wilcoxon signed-rank test was used (Field & Hole, 2003). The median rating was significantly higher than neutral, $Mdn = 4.00$, $Z = 5.36$, $p < .001$, $r = .87$. In addition, the mean score across 7 constructs was statistically significantly higher than the neutral score of 3, $t(37) = 18.86$, $p < .001$, with a very large effect size (Cohen's $d = 3.06$, 95% CI [2.29, 3.82]). The 95% CI for mean difference ranged from 1.01 to 1.25 (Cohen, 1988; Jiroutek & Turner, 2016).

The Pearson correlation coefficients between each of TFA1-TFA7 and the single overall acceptability item (item 8.1) are presented in Table 2. Overall acceptability was significantly and positively associated with *Affective Attitude*, *Burden*,

Intervention Coherence, Opportunity Costs, Perceived Effectiveness, and Self-efficacy. After Bonferroni adjustment, the association between overall acceptability and *Self-Efficacy* remained statistically significant ($r = .47$, unadjusted $p = .003$, adjusted $p = .021$), whereas the association between overall acceptability and *Ethicality* was positive but did not remain statistically significant ($r = .37$, unadjusted $p = .021$, adjusted $p = .147 > .05$). In particular, the strongest associations were observed for *Burden, Opportunity Costs, and Perceived Effectiveness* ($r = .68, .67, .68$).

Table 2

Pearson Correlations Coefficients Among the Seven TFA Constructs and Overall Acceptability Item (Item 8.1)

Variable	Affective Attitude	Burden	Ethicality	Intervention Coherence	Opportunity Costs	Perceived Effectiveness	Self-efficacy
Affective Attitude	—	—	—	—	—	—	—
Burden	.60	—	—	—	—	—	—
Ethicality	.44	.29	—	—	—	—	—
Intervention Coherence	.57	.63	.47	—	—	—	—
Opportunity Costs	.37	.62	.01	.19	—	—	—
Perceived Effectiveness	.54	.56	.38	.53	.51	—	—

Self-efficacy	.47	.60	.36	.66	.26	.32	—
Single Overall Acceptability Item (Item 8.1)	.60†	.68†	.37	.57†	.67†	.68†	.47†

Note. Entries are Pearson correlations. $N = 38$ † $p < .007$, Bonferroni-adjusted significance criterion for the seven planned correlations with overall acceptability.

Discussion

Overall Clinical Acceptability of TOPOL

This study aims to investigate the retrospective clinical acceptability of TOPOL from the perspective of SLPs with two research questions. RQ1, which examined the overall acceptability of TOPOL, indicated a mean rating of 4.13 out of 5. With reference to the interpretations by Goldstein et al. (2017), which revealed ratings of 4.2–5.1 out of 6.0 (70%– 85%), Thiemann-Bourque et al. (2018), which revealed ratings of 4.0–5.0 out of 5.0 (80%), and Wong et al. (2026), showing a mean rating of 3.24 out of 4 (81%), our mean ratings of $M = 4.13$ (82.6%) for both the mean score across 7 constructs and mean score of the single overall acceptability item (item 8.1) were considered high. The results are consistent with previous findings that SLPs prefer norm-referenced, decontextualized assessments administered in clinical context (Denman et al. 2021; Fulcher-Rood et al., 2018; Wilson et al., 1991). This suggested

that TOPOL was perceived as highly acceptable by SLPs as a language assessment and thus underscores TOPOL's practical relevance in Hong Kong.

Multifaceted Relationship Among Factors

RQ2 examined the relationship between TFA constructs and overall clinical acceptability of TOPOL. The correlation analyses revealed a multifaceted relationship between the overall acceptability and the seven TFA constructs. The strongest associations were observed in *Burden*, *Opportunity Costs*, and *Perceived Effectiveness*, ($r = .68, .67, .68$, respectively). Specifically, Sekhon et al. (2017) defined *Burden*, *Opportunity Costs*, and *Perceived Effectiveness* respectively as “the perceived amount of effort that is required to participate in the intervention,” “the extent to which benefits, profits or values must be given up to engage in the intervention,” and “the extent to which the intervention is perceived as likely to achieve its purpose,” respectively. Since all items are positively worded, higher scores indicated greater perceived manageability, more acceptable trade-offs, and more favorable perceptions of effectiveness, respectively. These three constructs appeared to be important correlates of TOPOL's acceptability. However, because the analyses were correlational, they should not be interpreted as demonstrating causal effects.

The strong positive association between greater manageability (i.e., low burden)

and higher overall acceptability is consistent with Fulcher-Rood et al. (2018) who emphasized the practical feasibility of assessment. The authors demonstrated that SLP respondents regarded standardized tests as easy to administer and relatively less time-consuming, particularly when under time-constraints and handling considerable number of caseloads. This association may be explained by TOPOL's straightforward scoring and clear administration instructions, which may enable SLPs to administer the test consistently without using a considerable amount of cognitive effort to comprehend instructions. To illustrate, most items use picture books with simple clinician-delivered questions. The complexity of administration and scoring may be perceived as low to SLPs. Therefore, these attributes may make TOPOL well suited to SLPs' skills and underscore TOPOL's practicability for local clinicians.

Additionally, the primary purpose of TOPOL is to evaluate a child's language ability. Our findings revealed high perceived effectiveness of TOPOL. This is consistent with the finding reported in previous studies. In particular, SLPs value assessment tools that are perceived as effective in supporting clinical judgments, making decisions in ambiguous cases and providing a comprehensive overview of a child's language skills (Fulcher-Rood et al., 2018; Kerr et al., 2003; Wilson et al., 1991). One possible explanation is that clinicians may perceive TOPOL as a valid and reliable assessment tool, which is useful for obtaining a multiple domains of children's

language abilities, including receptive and expressive grammar, story comprehension, and expressive vocabulary. However, the current data did not directly examine the reasons underlying perceived effectiveness. Future investigations are recommended.

Moreover, greater perceived acceptability of trade-offs—that is, more favorable judgments about the opportunity costs associated with using the intervention—was strongly correlated with overall acceptability. This is consistent with findings from Betz et al. (2013), who reported that the benefits of standardized tests in assessing a child’s linguistic system outweighed the burden of a lengthy administration period. Despite its lengthy administration period, TOPOL’s broad coverage of multiple language domains within a single assessment may similarly make this demand acceptable to SLPs. One possible explanation is that clinicians consider the time to administer TOPOL worthwhile because it provides comprehensive information about a child’s language abilities, which can support clinical decision-making.

Limitation and Future Investigations

There were several limitations. First, demographic information (e.g., years of experience, qualification, and clinical setting) of respondents was not collected. These factors might be associated with acceptability since clinicians with different workplaces, levels of experience and education vary in familiarity and tolerance towards TOPOL. Additionally, the number of sections was not indicated at the

beginning of the survey, which may have led some participants to discontinue without completion of the TOPOL section (both quantitative and qualitative parts), leaving 26 incomplete responses, resulting in the current small sample ($N = 38$). These 38 responses are estimated to represent approximately 3.69% of the SLPs in Hong Kong (Hong Kong Association of Speech Therapists [HKAST], 2026), resulting in limited generalizability. The association between demographic information and reported clinical acceptability can be further explored with a larger sample size.

Conclusion

This study aimed to examine the clinical acceptability of the TOPOL from perspectives of SLPs in Hong Kong. The quantitative results showed high overall acceptability and demonstrated a multifaceted relationship—the strong correlations between high overall acceptability and low perceived burden, low perceived opportunity costs, and high perceived effectiveness. These findings highlight the practical relevance of TOPOL in Hong Kong. Investigations on the association between demographic information and clinical acceptability can be explored in the future to guide improvement in standardized tests.

References

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment.

Language, Speech, and Hearing Services in Schools, 44(2), 133–146.

[https://doi.org/10.1044/0161-1461\(2012/12-0093\)](https://doi.org/10.1044/0161-1461(2012/12-0093))

Child Assessment Service. (2019). *The Hong Kong Test of Preschool Oral Language (Cantonese) (TOPOL)*. Department of Health, the Hong Kong SAR Government. Hong Kong.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).

Lawrence Erlbaum Associates.

Denman, D., Cordier, R., Kim, J.-H., Munro, N., & Speyer, R. (2021). What influences speech-language pathologists' use of different types of language assessments for elementary school-age children? *Language, Speech, and Hearing Services in Schools*, 52(3), 776–793.

https://doi.org/10.1044/2021_LSHSS-20-00053

Field, A., & Hole, G. (2003). *How to design and report experiments*. SAGE

Publications. <https://doi.org/10.4135/9781036231200>

Fulcher-Rood, K., Castilla-Earls, A. P., & Higginbotham, J. (2018). School-based

speech-language pathologists' perspectives on diagnostic decision making.

American Journal of Speech-Language Pathology, 27(2), 796–812.

https://doi.org/10.1044/2018_AJSLP-16-0121

Goldstein, H., Olszewski, A., Haring, C., Greenwood, C. R., McCune, L., Carta, J.,

Atwater, J., Guerrero, G., Schneider, N., McCarthy, T., & Kelley, E. S. (2017).

Efficacy of a supplemental phonemic awareness curriculum to instruct

preschoolers with delays in early literacy development. *Journal of Speech,*

Language, and Hearing Research, 60(1), 89–103.

https://doi.org/10.1044/2016_JSLHR-L-15-0451

Guildford, S., & Bird, E. K. R. (2003). Standardized language test use: A Canadian

survey. *Journal of Speech-Language Pathology and Audiology*, 27(1), 1–10.

Hong Kong Association of Speech Therapists. (2026). *Annual report 2025-2026*.

<https://drive.google.com/file/d/1n6SDUKNhkVZIEfOxB8Ckbs9KeLjbYj12/view>

[w](#)

IBM Corp. (2026). *IBM SPSS Statistics* (Version 31.0.2.0) [Computer software]. IBM

Corp.

Jiroutek, M. R., & Turner, J. R. (2016). In praise of confidence intervals: Much more

informative than *P* values alone. *The Journal of Clinical Hypertension*, 18(9),

955–957. <https://doi.org/10.1111/jch.12908>

J. Kerr, A. Guildford, and E. Kay-Raining Bird in the *Journal of Speech-Language Pathology and Audiology* (JSLPA).

Qualtrics. (2026). *Qualtrics XM Platform* [Computer software].

Qualtrics. <https://www.qualtrics.com>

Sekhon, M., Cartwright, M., & Francis, J. J. (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC health services research*, 17(1), 88.

<https://doi.org/10.1186/s12913-017-2031-8>

Sharma, H. (2022). How short or long should be a questionnaire for any research? Researchers dilemma in deciding the appropriate questionnaire length. *Saudi Journal of Anaesthesia*, 16(1), 65–68. https://doi.org/10.4103/sja.sja_163_21

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>

Thiemann-Bourque, K., Feldmiller, S., Hoffman, L., & Johnner, S. (2018).

Incorporating a peer-mediated approach into speech-generating device intervention: Effects on communication of preschoolers with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research*, 61(8), 2045–

2061. https://doi.org/10.1044/2018_JSLHR-L-17-0424

Wong, E. C. H., Chan, M. W. L., & Sheng, L. (2026). Clinical acceptability of the dynamic assessment of narrative in Cantonese: A mixed-methods study.

American Journal of Speech-Language Pathology, 35(4), 1516–1543.

https://doi.org/10.1044/2026_AJSLP-24-00568