

Telepractice norms and validity of performance-based tests

When compared to face-to-face assessment, how does telepractice with performance-based tests measure up?

Understandably, you may not be using assessments at this time, but if you are considering telepractice, please consult your professional organisations, licensing boards, professional liability insurance providers, and state laws and regulations regarding telepractice.

Comprehensive telepractice assessment with performance-based measures is more common today than it was prior to 2020. Some practitioners engaged in telepractice to increase access to assessment for individuals in underserved areas, where travel to the practitioner's location would be impractical but those services are still needed. Others did so for school districts with a shortage of practitioners to make the assessment process more accessible and efficient for all. In large part, telepractice has taken place on proprietary platforms by practitioners that specialized in that type of assessment, and validity studies were collected to provide evidence that telepractice is a reliable and valid way to complete assessment and that performance results are equivalent to those of in-person assessment. The COVID-19 pandemic was accompanied by evolution of the telepractice landscape. Direct-to-home assessment became more common, and access to teleconferencing software became more common in the general population as many educational, social, and community events adapted to an online environment during stay-at-home orders. Telepractice assessment became more widely adopted, and it has played an important part in keeping examinees and practitioners safe.

Practitioners requested assistance with telepractice to provide accommodations and modifications for clients who required assessment and could not wait until face-to-face assessment was deemed safe and feasible to receive those services. All clients (and all practitioners) were suddenly vulnerable to a serious illness, and spreading the virus by meeting face-to-face could worsen a public health emergency. Clients and practitioners alike require protection from the virus, but a substantial number of clients continue to urgently require assessment services to assist with clinical issues and difficulties.

When a test user contemplates an alteration in test administration mode (e.g. from face-to-face to telepractice), the user should have a sound rationale and empirical evidence, when possible, for concluding that the reliability/precision of scores and the validity of interpretations based on the scores won't be compromised (Standard 9.9; Standards for Educational and



Pearson

Telepractice norms and validity of performance-based tests

Psychological Tests [*Standards*]; AERA, APA, & NCME, 2014). The *Standards* indicate that if a modification such as a change in test administration mode is suspected of affecting the validity of score precision and interpretation, such that the change modifies or changes the construct being assessed, or if a given modification becomes widespread, evidence for validity should be gathered.

The *Standards* indicate that separate norms may be needed in the event a practitioner needs to make an accommodation involving a mode change (e.g. to telepractice from face-to-face) if the validity evidence suggests this is necessary. The available validity evidence to date (collected prior to the COVID-19 crisis) does not suggest separate telepractice norms are necessary, but it does not cover all task types, age ranges, nor all clinical conditions. However, gathering new evidence is impractical during the pandemic, because validity evidence for equivalence of two modes involves randomly assigning a sample of examinees to either a standard (face-to-face) or a new (telepractice) condition and investigating the properties of the obtained scores in the two modes.

Several studies have produced evidence of equivalence for tasks administered in telepractice and face-to-face modes for examinees with and without clinical conditions (see telepractice references). A meta-analysis of telepractice studies provides support for telepractice and face-to-face mode equivalence across a variety of neuropsychological tasks with adults (Brearly et al., 2017). Telepractice involves the use of technology in assessment as well as viewing onscreen stimuli. For these reasons, studies that investigate assessment in digital versus traditional formats are also relevant. Several such studies have also produced evidence of equivalence (see Digital references).

The existing support for telepractice assessment spans a wide variety of measures. Conducting demands analysis of the tasks that have been evaluated is useful because it yields information about task input and output demands. That information can then be applied to understand normative information applicability and equivalence for tasks with similar demands. For example, various receptive vocabulary tasks are constructed with nearly identical input and output demands (i.e., input is brief verbal directions and visual stimulus (pictures), and the required output is a brief oral response or a pointing motor response in a multiple choice format). Therefore, validity evidence on one task can be applied to understand the norms' applicability to other tasks with similar demands. A similar approach was applied to establish the validity of the traditional norms for Q-interactive (Pearson's tablet-based administration and scoring platform), for which a series of several equivalence studies demonstrated equivalence of traditional and digital norms. However, while equivalence data on similar measures are relevant, practitioners should be mindful that more research is needed to establish telepractice equivalence in all ages and for all tasks.

According to the *Standards*, normative information applicability is maximized to the extent that the original standardized conditions for the testing environment are maintained. In principle, the goal of telepractice assessment is, therefore, to mirror closely the face-to-face experience. The equivalence studies that provide the basis for telepractice and face-to-face equivalence were most often conducted in controlled environments, and some testing - especially during the COVID-19 pandemic - may occur in examinees' homes, so efforts to replicate a similar environment are important to maximize normative information applicability. If in-home assessment is taking place, it is advisable to prepare a similar environment as much as possible, as described in the telepractice guidance on the Pearson website in the [Telehealth page](#). It should be noted that very little research has been done about remote assessment in private homes, so the environment should emulate a clinic or school situation as much as possible. The examinee's environment should furnish reasonable comfort and involve minimal distractions to avoid construct-irrelevant variance.

The *Standards* also indicate that normative information applicability is also maximized to the extent that examiners maintain the standardized procedures. Most of the equivalence studies that support telepractice and face-to-face equivalence involve the study examiners becoming very familiar with the teleconference platform by using it for its intended purpose for several hours and administering tests (even those that are familiar in face-to-face mode) multiple

Telepractice norms and validity of performance-based tests

times to “practice examinees.” Therefore, the examiner needs to gain competency with the teleconference software and remote assessment. The goal is to administer in the telepractice environment as smoothly and naturally as one would in a face-to-face situation. All items and instructions of each task should be reviewed with an eye toward telepractice to become aware of any administration requirements that may differ in the telepractice environment. For example, the examinee may point with the mouse instead of with a finger.

Using third party facilitators

Studies that have established telepractice and face-to-face mode equivalence generally involve either the examinee or a professional staff member who is handling administrative and technological aspects of the session. However, preliminary research conducted and described by Lana Harder (Stolwyk et al., 2020) with parents serving as in-home facilitators who managed audiovisual needs and response booklets found no significant differences in scores across telepractice and face-to-face modes. Practitioners should evaluate concerns such as threats to validity of conclusions and to test security, as well as any effects on the examinee’s participation or communication between the examinee and examiner, as outlined eloquently by Otto and Krauss (2009). However, none of these issues may be a reason to prohibit the third party’s presence.

Practitioners should consider involving a third party if they conclude that the third party’s presence will result in more valid results than would otherwise be possible. Practitioners engaging in telepractice assessment may train facilitators to work with them on a regular basis in order to provide greater coverage to underserved populations (e.g., only two practitioners within a 500-mile radius, shortage of practitioners within a school district). If such a facilitator is well trained and in a professional role (i.e., a professional facilitator), they can present manipulatives as well as adjust audiovisual equipment. Importantly, there are subtests and entire tests for which a professional facilitator is a necessity or for which telepractice is not recommended. Without a professional facilitator, for example, some tasks with manipulatives cannot be administered correctly. In the presence of such major alterations in administration procedures which considerably alter the construct measured, the existing normative information may not apply.

It is anticipated that the use of telepractice assessment may increase in the future because it has become more widespread and practitioners and clients recognize its benefits. Additional validity studies may be conducted when the COVID-19 crisis wanes and data can again be safely collected using random assignment studies in which examinees are assigned randomly to face-to-face versus telepractice conditions and the test properties can be examined.

Multiple areas that require further study

In some cases where results are higher stakes, studies on full adult batteries may be conducted in the future because the existing equivalence studies for adults focus on smaller screening neuropsychological batteries formed by extracting representative tasks from multiple tests rather than studying full tests. Limited research validates the use of widely-available teleconferencing software and screensharing with children, as most of the available research was collected on proprietary platforms designed for this purpose. Studies validating modified instructions for the examinee and the examiner are also lacking; at present, modified instructions are not provided in detail in most studies. In general, there is a paucity of research on preschoolers. More research is also required for clinical groups in all age ranges. In addition, direct-to-home assessment and situations where clients rather than a professional facilitator independently manage technology, testing materials, or kits require examination.

Telepractice norms and validity of performance-based tests

If future studies identify tasks or age groups for which normative information should be adjusted for telepractice, norms may be established in several ways. For example, scaling and equating procedures can be used to link scores across two modes, formats, or forms of tests that are highly related so that the scaled or standard scores obtained take mode into account. Equating techniques have been used to adjust scores across modes or formats for tests ranging from large scale education testing to clinical tests. For example, college entrance exam scores are regularly equated across digital and paper delivery formats (Li, Yi, & Harris, 2016; Proctor, Chuah, Montgomery, & Way, 2019), and Pearson has previously equated test scores obtained in traditional and digital formats in a similar manner (Raiford et al., 2016).

Conclusion

Test users should review the general and test specific telepractice considerations on Pearson's website, which provides references to the literature supporting equivalence to date. As always, test users should remain mindful to consult professional best practice recommendations, respective ethical codes, telepractice regulations and legal requirements from federal, state, and local authorities, licensing boards, professional liability insurance providers, and payors. Test users should develop competence with telepractice assessment through activities such as practicing, studying, consulting with other professionals, and engaging in professional development. Following these activities, test users should make an informed decision to determine if assessment via telepractice is appropriate for a given examinee, referral question, and situation, and that it is feasible and not contraindicated. It is important to weigh all these considerations against the potential benefits for each individual when determining if they should be evaluated via telepractice during the COVID-19 pandemic.

References

General

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: Author.
- American Psychological Association Services (APA Services). (2020). *Guidance on psychological tele-assessment during the COVID-19 crisis*. (2020). <https://www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-covid-19>
- Grosch, M. C., Gottlieb, M. C., & Cullum, C. M. (2011). Initial practice recommendations for teleneuropsychology. *The Clinical Neuropsychologist*, 25, 1119–1133.
- Interorganizational Practice Committee [IOPC]. (2020). *Recommendations/guidance for teleneuropsychology (TeleNP) in response to the COVID-19 pandemic*. <https://static1.squarespace.com/static/50a3e393e4b07025e1a4f0d0/t/5e8260be9a64587cfd3a9832/1585602750557/Recommendations-Guidance+for+Teleneuropsychology-COVID-19-4.pdf>
- Li, D., Yi, Q., & Harris, D. (2016). *Evidence for paper and online ACT comparability: Spring 2014 and 2015 mode comparability studies* (ACT Working Paper 2016-02). ACT. <https://www.act.org/content/dam/act/unsecured/documents/Working-Paper-2016-02-Evidence-for-Paper-and-Online-ACT-Comparability.pdf>
- Proctor, T. P., Chuah, S. C., Montgomery, M., & Way, W. D. (2019). *Comparability of performance on the SAT suite of assessments across pencil-and-paper and computer-based modes of administration: SAT, PSAT™ 10, and PSAT™ 8/9*. The College Board. <https://collegereadiness.collegeboard.org/pdf/comparing-performance-paper-digital-tests-sat-suite-assessments.pdf>
- Stolwyk, R., Hammers, D. B., Harder, L., & Cullum, C. M. (2020). *Teleneuropsychology (TeleNP) in response to COVID-19*. <https://event.webinarjam.com/replay/13/pyl2nayhvsp09>

Telepractice/Face-to-Face Mode Equivalence

- Brearly, T., Shura, R., Martindale, S., Lazowski, R., Luxton, D., Shenal, B., & Rowland, J. (2017). Neuropsychological test administration by videoconference: A systematic review and meta-analysis. *Neuropsychology Review*, 27(2), 174–186.
- Cullum, C. M., Weiner, M., Gehrmann, H., & Hynan, L. (2006). Feasibility of telecognitive assessment in dementia. *Assessment*, 13(4), 385–390.
- Cullum, C. M., Hynan, L. S., Grosch, M., Parikh, M., & Weiner, M. F. (2014). Teleneuropsychology: Evidence for video teleconference-based neuropsychological assessment. *Journal of the International Neuropsychological Society*, 20, 1028–1033.
- Galusha-Glasscock, J., Horton, D., Weiner, M., & Cullum, C. M. (2016). Video teleconference administration of the Repeatable Battery for the Assessment of Neuropsychological Status. *Archives of Clinical Neuropsychology*, 31(1), 8–11.

Telepractice norms and validity of performance-based tests

- Grosch, M., Weiner, M., Hynan, L., Shore, J., & Cullum, C. M. (2015). Video teleconference-based neurocognitive screening in geropsychiatry. *Psychiatry Research, 225*(3), 734–735.
- Hildebrand, R., Chow, H., Williams, C., Nelson, M., & Wass, P. (2004). Feasibility of neuropsychological testing of older adults via videoconference: Implications for assessing the capacity for independent living. *Journal of Telemedicine and Telecare, 10*(3), 130–134. <https://doi.org/10.1258/135763304323070751>
- Hodge, M., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., Detheridge, J., Drevensek, S., Edwards, L., Everett, M., Ganesalingam, K., Geier, P., Kass, C., Mathieson, S., McCabe, M., Micallef, K., Molomby, K., Ong, N., Pfeiffer, S., ... Silove, N. (2019). Agreement between telehealth and face-to-face assessment of intellectual ability in children with specific learning disorder. *Journal of Telemedicine and Telecare, 25*(7), 431–437. <https://doi.org/10.1177/1357633X18776095>
- Otto, R. K., & Krauss, D. A. (2009). Contemplating the presence of third party observers and facilitators in psychological evaluations. *Assessment, 16*, 362–372.
- Ragbeer, S. N., Augustine, E. F., Mink, J. W., Thatcher, A. R., Vierhile, A. E., & Adams, H. R. (2016). Remote assessment of cognitive function in juvenile neuronal ceroid lipofuscinosis (Batten disease): A pilot study of feasibility and reliability. *Journal of Child Neurology, 31*, 481–487. <https://doi.org/10.1177/0883073815600863>
- Stain, H. J., Payne, K., Thienel, R., Michie, P., Vaughan, C., & Kelly, B. (2011). The feasibility of videoconferencing for neuropsychological assessments of rural youth experiencing early psychosis. *Journal of Telemedicine and Telecare, 17*, 328–331. <https://doi.org/10.1258/jtt.2011.101015>
- Sutherland, R., Trembath, D., Hodge, A., Drevensek, S., Lee, S., Silove, N., & Roberts, J. (2017). Telehealth language assessments using consumer grade equipment in rural and urban settings: Feasible, reliable and well tolerated. *Journal of Telemedicine and Telecare, 23*(1), 106–115. <https://doi.org/10.1177/1357633X15623921>
- Temple, V., Drummond, C., Valiquette, S., & Jozsvai, E. (2010). A comparison of intellectual assessments over video conferencing and in-person for individuals with ID: Preliminary data. *Journal of Intellectual Disability Research, 54*(6), 573–577. <https://doi.org/10.1111/j.1365-2788.2010.01282.x>
- Wadsworth, H., Galusha-Glasscock, J., Womack, K., Quiceno, M., Weiner, M., Hynan, L., Shore, J., & Cullum, C. (2016). Remote neuropsychological assessment in rural American Indians with and without cognitive impairment. *Archives of Clinical Neuropsychology, 31*(5), 420–425. <https://doi.org/10.1093/arclin/acw030>
- Wadsworth, H. E., Dhima, K., Womack, K. B., Hart, J., Weiner, M. F., Hynan, L. S., & Cullum, C. M. (2018). Validity of teleneuropsychological assessment in older patients with cognitive disorders. *Archives of Clinical Neuropsychology 33*(8), 1040–1045. <https://doi.org/10.1093/arclin/acx140>
- Wright, A. J. (2018a). Equivalence of remote, online administration and traditional, face-to-face administration of the Woodcock-Johnson IV cognitive and achievement tests. *Archives of Assessment Psychology, 8*(1), 23–35.

Telepractice norms and validity of performance-based tests

Wright, A. J. (2018b). *Equivalence of remote, online administration and traditional, face-to-face administration of the Reynolds Intellectual Assessment Scales-Second Edition*. <https://pages.presencelearning.com/rs/845-NEW-442/images/Content-PresenceLearning-Equivalence-of-Remote-Online-Administration-of-RIAS-2-White-Paper.pdf>

Wright, A. J. (2020). Equivalence of remote, digital administration and traditional, in-person administration of the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V). *Psychological Assessment*, 32, 800-817. <https://doi.apa.org/fulltext/2020-54568-001.pdf>

Digital/Paper Format Equivalence

Daniel, M. H. (2012). *Equivalence of Q-interactive administered cognitive tasks: CVLT-II and selected D-KEFS subtests* (Q-interactive Technical Report 3). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/005-s-Technical%20Report%203_CVLT_DKEFS_final_rev.pdf

Daniel, M. H. (2012). *Equivalence of Q-interactive administered cognitive tasks: WAIS-IV* (Q-interactive Technical Report 1). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/007-s-QinteractiveTechnical%20Report%201_WAIS-IV.pdf

Daniel, M. H. (2012). *Equivalence of Q-interactive administered cognitive tasks: WISC-IV* (Q-interactive Technical Report 2). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/009-s-Technical%20Report%202_WISC-IV_Final.pdf

Daniel, M. H. (2013). *Equivalence of Q-interactive administered cognitive tasks: Selected NEPSY-II and CMS subtests* (Q-interactive Technical Report 4). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/006-s-Technical%20Report%204_NEPSY-II_CMS.pdf

Daniel, M. H. (2013). *Equivalence of Q-interactive and paper scoring of academic tasks: Selected WIAT-III subtests* (Q-interactive Technical Report 5). Pearson. <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/008-s-Technical-Report-5-WIAT-III.pdf>

Daniel, M. H. (2013). *Equivalence of Q-interactive and paper administration of WMS-IV cognitive tasks* (Q-interactive Technical Report 6). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/0010-s-Technical_Report_6_WMS-IV.pdf

Daniel, M. H., Wahlstrom, D., & Zhang, O. (2014). *Equivalence of Q-interactive and paper administrations of cognitive tasks: WISC®-V* (Q-interactive Technical Report 8). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/003-s-Technical-Report_WISC-V_092514.pdf

Daniel, M. H., Wahlstrom, D., & Zhou, X. (2014). *Equivalence of Q-interactive and paper administrations of language tasks: Selected CELF-5 tests* (Q-interactive Technical Report 7). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/004-s-Technical%20Report%207_CELF-5_Final.pdf

Telepractice norms and validity of performance-based tests

- Drozdzick, L. W., Getz, K. G., Raiford, S. E., & Zhang, O. (2016). *WPPSI-IV: Equivalence of Q-interactive and paper administrations of cognitive tasks* (Q-interactive Technical Report 14). Pearson. https://images.pearsonclinical.com/images/assets/WPPSI-IV/WPPSI-Qi-Tech-Report-14_FNL.pdf
- Raiford, S. E., Holdnack, J. A., Drozdick, L. W., & Zhang, O. (2014). *Q-interactive special group studies: The WISC-V and children with intellectual giftedness and intellectual disability* (Q-interactive Technical Report 9). Pearson. <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/wisc-v/q-i-tr9-wisc-v.pdf>
- Raiford, S. E., Drozdick, L. W., & Zhang, O. (2015). *Q-interactive special group studies: The WISC-V and children with autism spectrum disorder and accompanying language impairment or attention-deficit/hyperactivity disorder* (Q-interactive Technical Report 11). Pearson. http://images.pearsonclinical.com/images/assets/WISC-V/Q-i-TR11_WISC-V_ADHDAUTL_FNL.pdf
- Raiford, S. E., Drozdick, L. W., & Zhang, O. (2016). *Q-interactive special group studies: The WISC-V and children with specific learning disorders in reading or mathematics* (Q-interactive Technical Report 13). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/011-s-Q-i%20TR13_WISC-V_SLDR_SLDM_FNL.pdf
- Raiford, S. E., Zhang, O., Drozdick, L. W., Getz, K., Wahlstrom, D., Gabel, A., Holdnack, J. A., & Daniel, M. (2016). *Coding and Symbol Search in digital format: Reliability, validity, special group studies, and interpretation* (Q-interactive Technical Report 12). Pearson. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/002-Qi-Processing-Speed-Tech-Report_FNL2.pdf
- Zhang, O., & Wang, W. (2015). *Equivalence of Q-interactive and paper administrations of a speech sound task, GFTA-3 sounds-in-words* (Q-interactive Technical Report 10). Pearson. <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/q-interactive/GFTA-3-Equivalence-study.pdf>

