

CALCULATOR PROVISION AS AN ACCOMMODATION FOR THE CANADIAN FORCES APTITUDE TEST (CFAT)

Joseph W. Berry¹

1. Department of National Defence

ABSTRACT

KEYWORDS

testing accommodations,
cognitive ability testing,
calculator provision,
calculator dependency,
accommodation policy

The impact of calculator provision on the reliability and validity of a version of the Canadian Forces Aptitude Test—Problem Solving subtest was investigated in order to inform testing accommodation policy. Two hundred and fifty-four Canadian Armed Forces recruits undergoing basic training participated in the experimental research design, which consisted of a calculator and a no-calculator condition. Results supported that the convergent validity of the test was maintained in the calculator condition, as indicated by similar validity coefficients with other measures of cognitive ability in the two conditions; however, several items showed increased correct responding, and there was mixed support for criterion-related validity when calculators were provided. Additionally, the calculator condition moderated the relationship between calculator dependency and test performance, such that the relationship was positive in the calculator condition and negative in the no-calculator condition. Implications of these results with respect to accommodation policy are discussed, as well as additional research that would be beneficial in this area.

As a large organization of over 100,000 members, the Canadian Armed Forces (CAF) has a significant recruitment and selection component that must screen and select a considerable number of applications each year. As part of this process, the CAF uses the Canadian Forces Aptitude Test (CFAT), which assesses cognitive ability, as one of its main tools to identify applicants who are most likely to succeed in the organization. The CFAT is a 60-item test, composed of three subtests, a 15-item Verbal Skill subtest, a 15-item Spatial Ability subtest, and a 30-item mathematics-based Problem Solving (PS) subtest. Applicants are given 45 minutes to complete the test, and currently no aids (e.g., dictionaries, calculators) are permitted during its completion.

As the CAF strives to increase the diversity of its members and abide by employment equity policies that protect persons with disabilities (e.g., [The Americans with Disabilities Act \[1990\]](#) in the United States and the [Accessible Canada Act \[Government of Canada Justice Laws Website, 2019\]](#) in Canada), there has been growing interest in the feasibility and impact of allowing various types of accommodations for applicants completing selection assessments. One such accommodation that is being considered is the provision of

calculators during completion of the CFAT-PS subtest. This study aims to investigate how such a provision would impact CFAT-PS subtest scores, its reliability, and its validity as a tool for selection into the CAF.

Selection Test Accommodations

Testing accommodations refer to any modification in the testing process that allows for persons with disabilities (or other groups protected by employment equity policies) to participate in the selection process ([Michel & Bonner, 2017](#)). The ultimate aim of any such accommodation is to remove barriers that may be limiting participation in the testing process but that are not relevant to the construct being measured by the test. At the same time, it is important to note that test accommodations are not meant to provide any type of advantage for test takers for whom they are intended ([Lovett, 2010](#)). In other words, the accommodations should not offer an advantage but rather remove the disad-

Corresponding author:
Joseph W. Berry
Author Email: joseph.berry@forces.gc.ca

vantage that is caused by construct-irrelevant factors.

Although the need to establish that any type of accommodation does not change the construct being measured has been clearly stated in various professional guidelines (e.g., *Principles for the Validation and Use of Personnel Selection Procedures* [Society for Industrial and Organizational Psychology, 2018], *Standards for Educational and Psychological Testing* [American Educational Research Association et al., 2014]), it is often the case that organizations will blindly implement them because it is more difficult and time consuming to go through the steps of gathering empirical evidence supporting the accommodation than it is to simply implement and advocate for it (Lovett & Lewandowski, 2015). Such a practice could negatively impact the benefits provided by the selection tools being used and open the door to legal challenges in the future. As such, it is important to gather as much information as possible before implementing any type of accommodation in the selection process.

Calculator Provision as an Accommodation

Despite the common use of calculators as an accommodation in educational settings, the research supporting its appropriateness and impact is still somewhat limited (Bone & Bouck, 2018) and is even less prevalent in an organizational context. In the field of education, some research does exist that looks at this subject. For example, several studies investigating the impact of calculator provision on middle school students (Bone & Bouck, 2018; Bouck & Bouck, 2008; Bouck et al., 2015) found that calculator provision increased mathematical test scores for students both with—and without—disabilities. Scarpati et al. (2011) also looked at a sample of over 70,000 middle school students completing state mathematics exams, of which over 12,000 were identified as students with disabilities who were provided with a calculator. Using differential item functioning analysis to explore differences in math performance between the two groups, they found that after matching students in the two groups on ability level, students provided with a calculator did better on eight of the 34 items, whereas students without a calculator did better on six of the items. Bridgeman et al. (2008) looked at the impact of calculator provision during completion of the Graduate Record Exam (GRE) quantitative section and found small increases in correct responding for individuals provided with a calculator. The authors attempted to break these differences down at the item level by predicting which items would be most likely to show increases or decreases in correct responding based on the degree of calculation required in the item; however, they were only modestly successful with only 20 of the 37 predicted items showing increases and none of the four predicted items showing a decrease in correct responding.

Taken together, it does seem that calculator provision does increase scores, at least by a small amount overall, for

mathematical-based tests. This increase is shown for both those identified as having learning disabilities and those without, a result that suggests caution should be used, and more evidence gathered, when deciding whether to provide calculators to only one group, as it may be that such a provision provides a performance advantage to all test takers. Other results from these studies, such as which items are most impacted and how calculator provision impacts performance after individuals are matched on ability level, are less clear, and require more investigation.

Current Study

After considering the extant literature regarding calculator provision, there are still some important questions that remain unanswered. First, given that all of the studies mentioned take place in an academic context with the majority assessing students at the middle school level, it would be useful to gather evidence on how calculator provision impacts tests used in an organizational selection context, where all applicants are adults and have already completed secondary education requirements. In addition, because random assignment is not typically used in calculator provision studies in the educational accommodations literature (i.e., calculators are often provided only to students with learning disabilities; see Bouck & Bouck [2008] and Bridgeman et al. [2008] for exceptions), it can be more difficult to determine which results are due to calculator provision or learning disability status. The current study uses an experimental design, which randomly assigns a relatively large number of adults into a calculator and no-calculator condition, in order to more clearly understand the specific impact of calculator provision.

Overall, we do not expect the context to vary the general relationship to a large extent, and therefore, our first hypothesis is as follows:

Hypothesis 1: Overall, scores will be higher in the calculator condition compared to the no-calculator condition.

Second, when considering differences at the item level, we would expect items that are more computationally involved to demonstrate a larger score increase when calculators are provided. Therefore:

Hypothesis 2: Items that are more computationally involved will be more likely to have higher rates of correct responding in the calculator condition compared to the no-calculator condition.

Reliability and validity of selection tests are critical to the ability to make appropriate hiring decisions using those tests. Although score differences have been assessed in the research previously mentioned, it is also critical to know

how the validity of the test is impacted by calculator provision. Because the test of concern in this paper was designed to measure cognitive ability or general reasoning ability, and was intended to be completed without the use of calculators, it is possible that the provision of calculators could introduce some construct-irrelevant factors (e.g., calculator proficiency, distractibility) that would decrease the reliability and validity of the test as a cognitive ability measure. Therefore the following three hypotheses are formulated:

Hypothesis 3a: Reliability estimates of the test will be significantly lower in the calculator condition.

Hypothesis 3b: Convergent validity of the test will be lower in the calculator condition.

Hypothesis 3c: Criterion-related validity of the test will be lower in the calculator condition.

Another potential impact of calculator provision could be the reduction of test anxiety for applicants. Test anxiety comprises both a cognitive component, related to negative thoughts and worries about the test, and an emotionality component, related to physical and emotional tension brought on by testing (Blankstein et al., 1992). For individuals who are dependent on calculators, it is possible that these negative thoughts, worries, and emotions would be more prevalent when taking a mathematical test without a calculator compared to when one is provided to them. Additionally, previous research has shown students generally feel positive about using calculators, and their provision contributes to a more positive attitude toward mathematics (Bone & Bouck, 2018). As such, we formulate the following hypothesis regarding calculator provision and test anxiety:

Hypothesis 4: Test anxiety will be lower for those in the calculator condition compared to those in the no-calculator condition.

Finally, it would be beneficial to know how the degree to which individuals report that they are dependent on calculators impacts their test performance when a calculator is provided versus when one is not. As calculators are becoming more and more prevalent in classrooms (Waits & Demana, 2000), so too may individuals' dependency on these tools when completing mathematical tasks. This could very well translate into an increase in requests for calculator provision as a test accommodation for the CFAT in the future. Assessing calculator dependency directly will also provide a new perspective, different from many previous studies (e.g., Bouck & Bouck, 2008; Bouck et al., 2015), as it will more directly identify those most likely to request a calculator accommodation compared to methods often used

in previous studies that simply compared groups with and without any type of learning disability (i.e., not dyscalculia specifically). We expect that those higher in calculator dependency will, in general, have lower test scores; however, the provision of calculators is expected to weaken this negative relationship:

Hypothesis 5: Calculator condition will moderate the relationship between calculator dependency and test scores, such that the relationship will be less negative in the calculator condition.

An experimental design was used to test these hypotheses and thereby help inform the appropriateness of calculator provision for the CFAT.

METHOD

Participants

Roughly 300 recruits completing their basic training for entry into the CAF were informed of the study and invited to participate in the research. Of these recruits, 254 agreed to participate; however, an additional seven participants were removed from the data due to nonresponses on the majority of the assessments. Average age of the participants was 24.1 years ($SD = 5.7$ years) and 78% identified as male, with 17% identifying as female, 1% identifying as other, and 4% not indicating their gender.

Materials

Practice CFAT-PS

Because of ethical concerns of having participants complete the actual CFAT-PS, but not having the results recognized on their service record, the practice version of the CFAT-PS (PCFAT-PS) was used instead. The PCFAT-PS is a parallel form of the CFAT-PS, constructed using item response theory. It has the same number of items (30) as the CFAT-PS and similar types of items, although the number of each item type is not identical and the PCFAT-PS does not contain any digit/letter counting problems, a problem type that is found on the actual CFAT-PS. The PCFAT-PS has four general types of items: word problems, arithmetic operations, number series, and spatial relations, with the word problems and arithmetic operations items being more computationally involved than the other two item types. Consistent with the CFAT-PS, 30 minutes were allotted for the completion of the PCFAT-PS. Psychometric analysis of the PCFAT has shown that all three subtests are psychometrically similar, in terms of means, standard deviations, and raw score distributions, to the actual CFAT (Kemp, 2021). The PCFAT-PS demonstrated an internal consistency of $\alpha = 0.77$.

Shipley-2

The Shipley-2 vocabulary and block patterns scales (Shipley et al., 2009) are measures of general cognitive ability and were used in the convergent validity analyses. The 40-item vocabulary scale is meant to be a measure of crystallized intelligence, whereas the 12-item block pattern scale is a measure of fluid intelligence. An internal consistency of $\alpha = 0.85$ for the vocabulary scale and $\alpha = 0.90$ for the block patterns scale were found in this study.

Cognitive Reflection

The six-item version of the Cognitive Reflection Test (CRT; Frederick, 2005; Primi et al., 2016) was used as an additional measure of cognitive ability and decision making. The internal consistency of the CRT was $\alpha = 0.68$.

Calculator Dependency

A four-item measure developed specifically for this study was used to measure calculator dependency. Participants answered items, such as “I needed a calculator to answer questions on this test,” on a five-point Likert-type scale ranging from strongly disagree to strongly agree. The internal consistency of this scale was $\alpha = 0.73$.

Test Anxiety

Test anxiety was measured using the 10-item scale developed by Arvey et al. (1990). An example item is, “I usually get very anxious about taking tests,” and items are answered on a five-point Likert-type scale ranging from strongly disagree to strongly agree. The internal consistency of the scale was $\alpha = 0.85$.

Task Performance

The seven-item, in-role performance measure (Williams & Anderson, 1991) was used as a criterion measure of task performance. An example item is, “I adequately complete assigned duties,” and items are answered on a five-point Likert-type scale ranging from strongly disagree to strongly agree. The internal consistency of the scale was $\alpha = 0.84$.

Procedure

All participants were randomly assigned into experimental and control groups by issuing tickets from a randomized ticket book that indicated either a calculator or no-calculator condition. Testing was administered simultaneously by six proctors who followed a standardized script and timing schedule. For both groups, demographic and academic history questionnaires were completed first, followed by the PCFAT-PS, measures of test anxiety and calculator dependency, Shipley-2 vocabulary and block pattern scales, the CRT, and finally the self-reported job performance scale.

Responses were analyzed for completeness and careless responding (e.g., same response option selected for all

items on multiple measures), which resulted in the removal of seven participants. The efficacy of our randomization procedure of the participants to conditions was checked by assessing if any differences in age, gender, or calculator dependency existed between the conditions. Independent samples *t*-tests were used to test score differences for Hypothesis 1 and Hypothesis 4, two-proportion *z*-tests were used to test Hypothesis 2, correlations were used for Hypothesis 3b and Hypothesis 3c, and moderated regression analysis was used to test Hypothesis 5.

RESULTS

Results from our randomization check showed a small but statistically significant difference in age ($t(222) = 2.45$, $p = 0.01$) between conditions, with the no-calculator condition having a slightly higher age ($M = 25.14$, $SD = 4.94$) than the calculator condition ($M = 23.30$, $SD = 6.26$). However, no significant differences between groups with respect to gender ($\chi^2 [1, 247] = 2.25$, $p = 0.13$) or calculator dependency ($t(236.54) = 0.70$, $p = 0.48$) were found, indicating that the participants were, for the most part, effectively assigned into conditions randomly.

Both Hypothesis 1 and Hypothesis 2 relate to increased performance on the PCFAT-PS due to calculator provision, at the test and item levels, respectively. Consistent with Hypothesis 1, there was a significantly higher overall PCFAT-PS score ($t(244.66) = 2.27$, $p = .02$) in the calculator condition ($M = 19.14$, $SD = 4.73$) compared to the no calculator condition ($M = 17.83$, $SD = 4.28$). For Hypothesis 2, relating to score differences at the item level, eight of the 30 PCFAT-PS items showed a significant difference in correct responding between conditions, with seven items showing an increase in correct responding in the calculator condition and one item showing a decrease. Response differences were expected in the word problems and arithmetic operations items, as these were assumed to be more computationally involved than the other two item types. Table 1 shows the breakdown of item-level differences broken down by item type.

As can be seen in Table 1, all item differences occurred with the word problem and arithmetic operation item types. No differences in correct responding were found in the number series or spatial relations problems.

With respect to Hypothesis 3a, no significant difference between the reliability estimates in the calculator ($\alpha = .79$) and no-calculator ($\alpha = .74$) conditions was found. However, there were some mixed results regarding the convergent and criterion-related validity of the test between conditions. Table 2 contains the validity coefficients, or correlations, between the PCFAT-PS and the cognitive and job performance measures.

As can be seen in this table, validity coefficients remained stable across conditions for all convergent validity

TABLE 1.
PCFAT-PS Item-Level Changes in Correct Responding With Calculator Provision

| Item Type | Correct response increase | Correct response decrease |
|---------------------------------|---------------------------|---------------------------|
| Word problems (21 items) | 5 items | 1 item |
| Arithmetic operations (4 items) | 2 items | 0 items |
| Number series (3 items) | 0 items | 0 items |
| Spatial relations (2 items) | 0 items | 0 items |

measures ($p < .05$), providing support for Hypothesis 3b; however, differences can be seen between conditions for the criterion-related validity coefficients (i.e., associations with academic performance and task performance). Validity coefficients were significant in the no-calculator condition but not in the calculator condition. Additionally, although the difference in validity coefficients was not statistically significant between conditions for task performance ($z = 1.19$; $p = .23$), the difference was statistically significant for academic performance ($z = 2.12$; $p = .03$), providing partial support for Hypothesis 3c.

Contrary to the prediction of Hypothesis 4, there was no significant difference ($t(226.76) = 1.47$, $p = .14$) in reported levels of test anxiety found between the no-calculator ($M = 2.78$, $SD = 0.68$) and calculator ($M = 2.66$, $SD = 0.60$) conditions.

Finally, with respect to Hypothesis 5, moderated regression analysis (see Table 3) found that calculator provision did moderate the relationship between calculator dependency and PCFAT-PS scores, such that higher levels of reported calculator dependency were associated with higher PCFAT-PS scores in the calculator condition and with lower PCFAT-PS scores in the no-calculator condition ($F(3, 242) = 4.49$, $p < .01$, $R^2 = .05$).

Figure 1 provides a visual representation of the moderated relationship between calculator dependency and PC-

TABLE 3.
Moderated Regression Coefficients

| Predictor | <i>B</i> | <i>p</i> | 95% CI |
|--|----------|----------|---------------|
| Intercept | 22.08 | <.001 | 18.63, 25.54 |
| Calculator Provision (0 = no calculator, 1 = calculator) | -1.24* | .013 | -2.21, -0.26 |
| Calculator Dependency | -5.28* | .027 | -10.00, -0.56 |
| Calculator Provision x Calculator Dependency | 1.93** | .005 | 0.58, 3.28 |

Note. * $p < .05$, ** $p < .01$.

TABLE 2.
Associations Between the PCFAT-PS Subscale and Other Related Measures

| Measure | No Calculator | Calculator |
|--------------------------|---------------|------------|
| ShIPLEY-2 Verbal | 0.34* | 0.32* |
| ShIPLEY-2 Block Patterns | 0.46* | 0.53* |
| Cognitive Reflection | 0.51* | 0.53* |
| Academic Performance | 0.34* | 0.06 |
| Task Performance | 0.26* | 0.12 |

Note. * $p < .05$.

FAT-PS scores. Interestingly, the provision of calculators is actually associated with lower PCFAT-PS scores for those who report lower calculator dependency, suggesting that the provision of calculators can interfere with test performance.

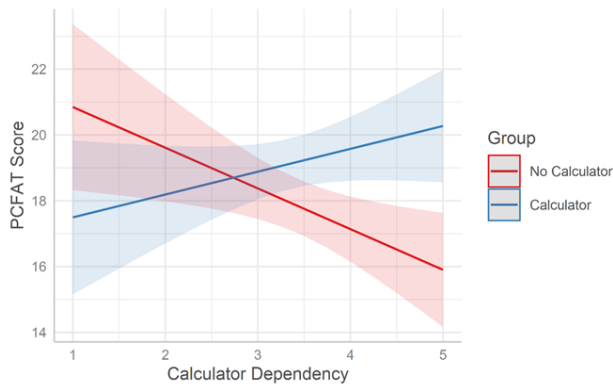
DISCUSSION

The results from this experimental research provide several findings that will be of interest for organizations using mathematics-based problem solving tests in a selection context. First, the provision of calculators for these tests does appear to cause an increase in scores for the overall test and for certain test items. Although this finding is not surprising, as our review of the educational literature does seem to suggest a small increase overall in test scores when calculators are provided, it is important to note that with variability in such factors as who received accommodations, item type, and the construct being measured, not every study found score increases (Scarpati et al., 2011). It is therefore of interest to applied practitioners that these small increases in scores for mathematics-based, problem-solving tests in a selection context do exist for those randomly assigned to receive a calculator.

Although the score increase in the PCFAT-PS subtest

FIGURE 1.

Moderating Impact of Calculator Provision on Calculator Dependency–PCFAT-PS Relationship



was relatively small, roughly one-third of a standard deviation, this difference must still be carefully considered given the high stakes nature of selection tests. Even a small change in scores from the provision of calculators could mean the difference between passing and failing a given selection cut off. For larger organizations, such as the CAF, that test a large number of applicants each year, this takes on even greater importance as it has the potential to impact an even greater number of job applicants.

It is also noteworthy that when this increase in scores is looked at more closely at the item level, not all items appear to be equally affected by calculator provision. All of the items that showed an increase in correct responding were types that are more computationally involved (i.e., word problems and arithmetic operations) as opposed to those that require recognizing patterns or shapes (i.e., number series and spatial relations). In fact, almost 25% of the word problems and 50% of the arithmetic operation items demonstrated significant increases in correct responding with the provision of calculators. Although previous attempts to predict which specific items would be more sensitive to calculator provision were not highly accurate (Bridgeman et al., 2008), it does appear that prediction of which item types are more likely to be impacted is possible. Such a result highlights the importance of considering the types of items contained in selection tests when contemplating the idea of providing calculators. Test developers should also consider this result if they are creating tests for use in a context when only a subgroup of the applicants will be provided with a calculator.

It is also important to consider the moderating impact that calculator provision has on the calculator dependency—CFAT-PS score relationship. Although our expectation that this relationship would be more negative in the no-calculator condition was supported, we had actually predicted that the relationship would be negative for both groups (i.e., those higher in calculator dependency would score lower

on the CFAT-PS, especially in the no-calculator condition). It was therefore surprising to see that the relationship actually became positive when calculators were provided. This means that, as expected, those who are more dependent on calculators score lower on the test in the no-calculator condition; however, when a calculator is provided, those who are more dependent on calculators actually score better on the test than those who are less dependent. Such a result suggests that the provision of calculators may actually inhibit performance for those reporting lower calculator dependency, possibly due to the calculator providing an unnecessary distraction or from users making errors while using tools with which they are not as familiar. Regardless of why this is happening, it is important to note that the provision of calculators seems to be beneficial for those higher in calculator dependency while actually being counterproductive for those low in calculator dependency. Such a finding provides some evidence-based support for providing calculators as an accommodation (i.e., only to those who request it), as those who request it are likely to be the ones with higher calculator dependency. It also qualifies the previous finding of increased scores at the test and item levels. If the increase in these scores is driven mainly by those who report a higher dependency on calculators, and therefore are the ones likely to request the accommodation, then it may not actually be of concern, as the calculators may simply be assisting those with a computation-related disability as opposed to providing an overall advantage to all test takers. This result is also a significant contribution to the existing literature because, to the author's knowledge, calculator dependency has not been considered in previous research. Although calculators have often been provided to students with disabilities, with disability status used as a kind of proxy for calculator dependency, calculator dependency has not been measured directly. As mentioned previously, this can be problematic because many students have disabilities that are not related to dyscalculia.

Our results related to the validity of the PCFAT-PS, specifically the criterion-related validity between the two conditions in Hypothesis 3c does indicate that caution should be used before providing calculators as an accommodation. Although the convergent validity of the test was maintained in our study when calculators were provided, the criterion-related validity dropped significantly. Adding to this the fact that some previous research that assessed construct validity using differential item functioning analysis found that there were in fact differences when calculators were provided (Scarpati et al., 2011), this finding could indicate that the provision of calculators is introducing some construct-irrelevant variance that is decreasing the relationship between the PCFAT-PS and its expected outcomes. These findings are also of particular interest to practitioners and contribute to the existing literature because the assessment of convergent and criterion-related validity, specifically, is

not commonly done in the educational literature with respect to calculator provision research.

Regarding the impact of calculator provision on test anxiety, there was no significant decrease in test anxiety for the calculator group, as we had initially expected. Despite the previous research in the educational domain that found students have generally positive attitudes toward calculator provision (Bone & Bouck, 2018), it appears that this does not translate into reduced test anxiety, at least in the current context.

Depending on the construct assessed by the test, a final important point of consideration for the provision of calculators is whether or not computational facility is required for successful performance in the occupation to which the individual is applying. Even if further support is found that criterion-related validity is maintained with calculator provision, it would not be appropriate to provide them if computational facility is determined to be a necessary requirement for the successful performance of the job (unless numerical facility was assessed using a separate measure). It may be the case that this requirement must be determined, through job analysis, for each occupation and accommodations provided only for those occupations for which (quantitative) reasoning is necessary but computational facility is not.

Limitations and Future Research

Although the experimental design of this study offered many advantages (e.g., ability to establish causality, assume equality of groups), there are a couple of limitations that should be acknowledged. First, our study did not identify participants with learning disabilities, something that is common in many of the previous studies related to test accommodations. We did, however, use a direct measure of calculator dependency, something that we suggest may actually be more appropriate for identifying individuals likely to request a calculator on tests involving basic mathematical operations. Additionally, the random assignment of individuals into calculator conditions assumes that there are no substantial differences between the two groups with respect to the number of participants with learning disabilities.

A second limitation is related to the criterion measures used in this study. Because all criterion data (job performance and academic performance) were collected using self-report measures and at the same time as the other measures in the study, self-report bias and common method variance (Podsakoff et al., 2003) could potentially be impacting our results. It would be useful for future research to see how the criterion-related validity of these measures are impacted when using criterion data collected from different sources and at different times. Despite these limitations, we believe that this research offers new and useful information for practitioners considering calculator provision as an accommodation for mathematical problem solving tests in

employee selection.

Summary

In summary, there is mixed support for the provision of calculators as an accommodation during the completion of cognitive ability tests involving basic mathematical operations. Although it appears that the provision of calculators does not provide an overall advantage for test takers, it does seem that the criterion-related validity of the measure may suffer. Further research to determine if this result is due to using less than ideal criterion measures or if this is a true effect would be fruitful. In the meantime, the risks should be carefully considered before providing calculators as an accommodation for mathematical problem solving tests in employee selection testing.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.
- Americans With Disabilities Act of 1990, 42 U.S.C. § 12101 et seq. (1990).
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43(4), 695-716. <https://psycnet.apa.org/doi/10.1111/j.1744-6570.1990.tb00679.x>
- Blankstein, K. R., Flett, G. L., & Watson, M. S. (1992). Coping and academic problem-solving ability in test anxiety. *Journal of Clinical Psychology*, 48(1), 37-46. [https://psycnet.apa.org/doi/10.1002/1097-4679\(199201\)48:1%3C37::AID-JCLP2270480105%3E3.0.CO;2-F](https://psycnet.apa.org/doi/10.1002/1097-4679(199201)48:1%3C37::AID-JCLP2270480105%3E3.0.CO;2-F)
- Bone, E. K., & Bouck, E. C. (2018). Evaluating calculators as accommodations for secondary students with disabilities. *Learning Disabilities: A Multidisciplinary Journal*, 23(1), 35-49. <https://doi.org/10.18666/LDMJ-2018-V23-11-8437>
- Bouck, E. C., & Bouck, M. K. (2008). Does it add up? Calculators as accommodations for sixth grade students with disabilities. *Journal of Special Education Technology*, 23(2), 17-32. <https://doi.org/10.1177/016264340802300202>
- Bouck, E. C., Bouck, M. K., & Hunley, M. (2015). The calculator effect: Understanding the impact of calculators as accommodations for secondary students with disabilities. *Journal of Special Education Technology*, 30(2), 77-88. <https://doi.org/10.1177%2F0162643415617371>
- Bridgeman, B., Cline, F., & Levin, J. (2008). Effects of calculator availability on GRE quantitative questions. GRE Board Research Report No. 03-09. Educational Testing Service. <https://ets.org/Media/Research/pdf/RR-08-31.pdf>
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25-42. <https://psyc->

- net.apa.org/doi/10.1257/089533005775196732
- Government of Canada Justice Laws Website. (2019). Accessible Canada Act, S.C. c.10. <https://laws.justice.gc.ca/eng/acts/A-0.6/page-1.html>
- Kemp, C. (2021). Psychometric analysis of the Practice Canadian Forces Aptitude Test (Director General Military Personnel Research and Analysis Scientific Report DRDC-RDDC-2021-R152). Defence Research and Development Canada.
- Lovett, B. J. (2010). Extended time testing accommodations for students with disabilities: Answers to five fundamental questions. *Review of Education Research*, 80(4), 611-638. <https://psycnet.apa.org/doi/10.3102/0034654310364063>
- Lovett, B. J., & Lewandowski, L. J. (2015). Testing accommodations for students with disabilities: Research-based practice. American Psychological Association. <https://psycnet.apa.org/doi/10.1037/14468-000>
- Michel, R. P., & Bonner, S. (2017). Blue-collar selection in private sector organizations. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection*, second edition (pp. 760-780). Routledge/Taylor & Francis Group. <https://doi.org/10.4324/9781315690193>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 897-903. <https://psycnet.apa.org/doi/10.1037/0021-9010.88.5.879>
- Primi, C., Morsany, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, 29(5), 453-469. <https://doi.org/10.1002/bdm.1883>
- Scarpati, S. E., Wells, C. S., Lewis, C., & Jirka, S. (2011). Accommodations and item-level analysis using mixture differential item functioning models. *Journal of Special Education*, 45(1), 54-62. <https://psycnet.apa.org/doi/10.1177/0022466909350224>
- ShIPLEY, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *ShIPLEY-2 manual*. Western Psychological Services.
- Society for Industrial and Organizational Psychology. (2018). *Principles for the validation and use of personnel selection procedures* (5th ed.). Cambridge University Press. <https://psycnet.apa.org/doi/10.1017/iop.2018.195>
- Waits, B., & Demana, F. D. (2000). Calculators in mathematics teaching and learning: Past, present, and future. Part 2: Technology and the mathematics classroom. National Council of Teachers of Mathematics, Inc. <https://eric.ed.gov/?id=ED482731>
- Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, 17, 601-617. <https://psycnet.apa.org/doi/10.1177/014920639101700305>