Bennett Mechanical Comprehension Test® II (BMCT–II)

Technical Manual and User's Guide

MORE INSIGHT
MORE IMPACT™

Pearson | TalentLens

Chapter 1

Introduction

The *Bennett Mechanical Comprehension Test*®**–**second edition (BMCT®**–**II) is an assessment tool for measuring a candidate's aptitude for and ability to perceive and understand the relationship of physical forces and mechanical principles in practical situations. The BMCT may be used to measure aptitude and predict performance in vocational and technical training settings, and to select candidates for mechanical, technical, engineering, and similar occupations. The BMCT–II is a web-based test administered online and may be administered to an individual or to a small group. All items are multiple-choice.

Objectives for the Development of the BMCT®–II

To increase the usability of the BMCT and to respond to customer needs, a more contemporary, item-banked version was developed with the following enhancements:

Item Bank—BMCT–II forms are generated from an item bank that is continually refreshed. The item bank ensures that no two test takers receive exactly the same set of items in their test. Therefore, BMCT–II can be used for candidate screening without a proctor to supervise the test session

Shorter Administration Time—Each BMCT–II form is composed of 55 items that can be administered in 25 minutes.

New Normative Information—New normative data were collected for the most important comparison groups. The comparison groups were selected based on customer feedback.

Contemporary Content and Artwork—New, color art has been created, with contemporary content and scenarios representative of the 21st century workplace.

Updated Evidence of Reliability and Validity—New studies were conducted to demonstrate that the BMCT–II is reliable and valid.

Mechanical Comprehension

The BMCT–II was developed to measure a person's *aptitude* for understanding and applying mechanical principles, from which an employer may infer future performance in jobs that require these skills. *Mechanical comprehension is an aptitude that* is regarded as one aspect of ability, as ability is broadly defined. An individual who scores high in mechanical comprehension tends to learn readily and can apply mechanical principles effectively.

Like other aptitude tests, a person's performance on the BMCT–II may be influenced by environmental factors, but not to the extent that interpretation of his or her performance is significantly affected. An individual's scores on the BMCT–II generally can be improved through training and experience, though it is unlikely that improvement will be dramatic. This is due, in part, to the presentation and composition of items that are simple, frequently encountered mechanisms that neither resemble textbook illustrations nor require special knowledge.

Description of the BMCT–II

The BMCT–II is not a static form test in which every test-taker completes the same questions; instead, an item-banked system generates different forms. This is a much more secure way of administering tests online because examinees are unlikely to have seen the items before or complete the same items as someone else taking the test. An item bank is a large pool of items (i.e., questions) from which tests are randomly generated. Constraints are built into the test to ensure that different tests are relatively equivalent. The BMCT–II scoring is based on *item-response theory* (IRT). This form of scoring can adapt for minor differences in difficulty between test versions so that scores are equivalent across administrations. Each administration of the BMCT–II is composed of 55 items that are illustrations of simple, frequently encountered mechanisms. For each item, the examinee reads a question about an illustration, examines the illustration, and chooses the best answer to the question from among three options. Though there is a 25-minute time limit for completing the test, the BMCT–II is not a *speeded* test. Speeded tests are composed of relatively easy items and rely on the number of correct responses within restrictive time limits to differentiate performance among examinees. The BMCT–II is considered a *timed power* test because items represent a wide range of item-level difficulty and it has a 25-minute time limit. Professionals who are responsible for testing and evaluation can set different BMCT–II cut scores appropriate for different job requirements.

Reading Level

An important design goal in developing the BMCT–II item bank was to keep the readability of the items approximately at the 6^{th} grade level. To achieve this goal, the reading level of BMCT–II items was compared to the *EDL Core Vocabularies in Reading, Mathematics, Science, and Social Studies* (Taylor et al., 1989) and, when possible, words that were judged to be higher than a 6^{th} grade reading level were replaced with similar words of a lower reading level. This comparison showed that both the directions and the test questions fell within the "fairly easy" range, similar to reading levels in popular fiction books or magazines.

History and Development of the BMCT®-II

In 1940, The Psychological Corporation (now Pearson) published the first BMCT form, Form AA. A second form, BB, was copyrighted in 1941 and was appreciably more difficult than Form AA. Form W1, developed using more generic situations and fewer technical terms, appeared in 1942. A number of restricted versions of the test also were prepared under contract with the military during mobilization for World War II. The Psychological Corporation published a variant of the BMCT®, Form CC, in 1949. This form, developed by W. A. Owens and the Bureau of Naval Personnel, had illustrations more nearly resembling mechanical drawings and presented examinees with five, rather than three, response alternatives.

When the *Differential Aptitude Test*® (DAT®) was prepared in 1947, a test of mechanical comprehension was included in the battery. The DAT Mechanical Reasoning Test was created to parallel BMCT Form AA and, though the content differs, the item type and general subject matter clearly identify it as an alternate form of the BMCT. Data on the Mechanical Reasoning test can be found in the *Differential Aptitude Tests for Personnel and Career Assessment*® Technical Manual (Bennett, Seashore, & Wesman, 1991).

The most widely used forms of the original BMCT, Forms S and T, were published in 1969. The last revision of these forms was published in 2006, with updated art and scenarios to reflect current culture.

Development of the BMCT–II Item Bank

The development of the BMCT–II item bank occurred in four main steps: Item writing, artwork creation, data collection, and analysis. The revision goals were to ensure that

- new items provided the broad range of content coverage and difficulty levels as previously established items;
- the item-banked test reliably measured mechanical aptitude; and
- the item-banked test continued to accurately measure the mechanical aptitude construct, as did other versions of the BMCT.

Item Writing

Existing items from the different forms of the BMCT were reviewed for acceptability. Experienced items writers with advanced degrees and backgrounds in physics, mechanical engineering, industrial and organizational psychology, and astronomy created new items. They received training and wrote items according to detailed specifications. The items underwent multiple reviews and revisions by the Pearson research team. Reviewers evaluated items based on content, relevance, and difficulty, and potential fairness issues. Items were rejected if they contained content that would be would not be equally familiar to different groups. A pool of 380 items was accepted.

Pilot

A 100-item subgroup of the item pool, consisting of items from existing forms that demonstrated exceptional item characteristics, was administered to 100 individuals. The goal of this was to ensure that the item properties were still adequate after the re-rendered artwork was introduced. After data collection, 30 of the best items were selected as anchor items for all subsequent standardization forms.

Standardization

Five 100-item static standardization forms were created from the pool of 380 items. The forms were administered in an online test delivery platform to 904 individuals from eight occupational groups (see Table 2.1).

After data collection was completed, the psychometric properties were evaluated based on Classical Test Theory and Item Response Theory (IRT). The 380 items were reviewed and items that were deemed unacceptable were discarded. Differential item functioning (DIF) information was also used to eliminate items that could potentially be biased against protected groups. A total of 302 items comprised the initial BMCT–II online item bank.

Table 2.1 Standardization Groups

Group	Occupations
Installation/Maintenance/Repair	Facility maintenance, HVAC technician, equipment installation technician
Skilled Trades	Electrician, welder, plumber, carpenter
Machine Operator/Machinist	Machine operator, machinist)
General Labor	Laborer, material mover/handler, groundskeeper)
Vocational/Technical Student	Student in a vocational/technical program requiring mechanical aptitude
Mechanic	Automotive, motorcycle, diesel, aircraft
Engineer	Mechanical, electrical, civil, aerospace
Industrial/Technical Sales	Industrial product sales, technical product sales

Item Bank Configuration

Although studies of the factor structure of mechanical aptitude, as measured by tests such as the BMCT, have generally supported a unidimensional factor structure (Trent, Yang, & Rose, 2007), a design goal for the BMCT–II was to cover a broad range of mechanically related scenarios and content. To achieve this goal, 13 broad content areas were identified and used to categorize the item-bank (see Table 2.2). Difficulty levels within each content category were constrained to ensure all that content coverage and item difficulty are equivalent for all examinees.

Table 2.2 Content Coverage in Each Administration of the BMCT–II

Content Area	Number of Items
Pulleys and Levers	7
Hydraulics	6
Resolution of Forces, Centrifugal Force, and Inertia	6
Structures, Planes, and Slopes	6
Gears and Belt Drives	5
Miscellaneous	6
Gravity and Velocity	4
Acoustics and Optics	3
Center of Gravity	3
Electricity	3
Heat	3
Shape and Volume	3

Evidence of Reliability

The *reliability* of a measurement instrument refers to the accuracy, consistency, and precision of test scores across situations (Anastasi & Urbina, 1997) and the confidence that may be placed in those results. Test theory posits that a test score is an estimate of an individual's hypothetical *true score,* or the score an individual would receive if the test were perfectly reliable. In actual practice, administrators do not have the luxury of administering a test an infinite number of times, so some measurement error is to be expected. A reliable test has relatively small measurement error.

The reliability of a test is expressed as a *correlation coefficient,* which represents the consistency of scores that would be obtained if a test could be given an infinite number of times. Reliability coefficients can range from .00 to 1.00. The closer the reliability coefficient is to 1.00, the more reliable the test. A perfectly reliable test would have a reliability coefficient of 1.00 and no measurement error. A completely unreliable test would have a reliability coefficient of .00. The U.S. Department of Labor (1999) provides the following general guidelines for interpreting a reliability coefficient: above .89 is considered "excellent," .80 to .89 is "good," .70 to .79 is considered "adequate," and below .70 "may have limited applicability."

The methods most commonly used to estimate test reliability are internal consistency of the test items (e.g., *Cronbach's alpha coefficient,* Cronbach 1970), test-retest (the stability of test scores over time), and alternate forms (the consistency of scores across alternate forms of a test). The reliability of a test should always be considered in the interpretation of obtained test scores

Internal Consistency

Previous Studies of Internal Consistency Reliability

Cronbach's alpha and the standard error of measurement (*SEM*) were calculated for the previous 2006 norm groups (see Table 3.1). Reliability estimates for these samples were similar to those found in previous studies and range from .84 to .92. Consistent with previous research, these values indicate that BMCT scores possess adequate reliability. In addition, internal consistency reliability estimates were calculated separately for the two forms in an overall sample of examinees who took the BMCT from 2003 to 2005. The internal consistency reliability coefficients were .90 for both Form S (*n* = 725) and Form T (*N* = 754).

Table 3.1 Means, Standard Deviations (*SD*), Standard Error of Measurement (*SEM*) and Internal Consistency Reliability Coefficients (*r*) of Forms S/T

Group	*N*	Mean	*SD*	*SEM*[1]	r_{alpha}
Occupation					
Automotive Mechanic	95	46.3	11.2	3.2	.92
Engineer	105	52.7	8.7	3.0	.88
Installation/Maintenance/Repair	112	49.1	8.4	3.1	.86
Industrial/Technical Sales Representative	133	47.4	7.9	3.2	.80
Skilled Tradesperson	153	47.4	8.9	3.2	.87
Transportation Trades/Equipment Operator	126	42.8	9.4	3.4	.87
Combined Occupation					
Mechanical Trades	388	47.5	9.4	3.3	.88
Automotive and Aircraft Mechanic	123	46.2	10.7	3.2	.91
Industry					
Manufacturing/Production	580	43.8	10.5	3.3	.90
Energy/Utilities	135	48.1	92	3.3	.87

[1]*Computed by the formula SEM = SD* $\sqrt{1 - r_{xx}}$

Current Studies of Internal Consistency Reliability

Because the BMCT–II is delivered via an item-banking system with no static forms, an IRT-based estimated reliability of .91 was generated using the standardization data (*N* = 905). This method, recommended by Dimitrov (2003), provides reliability estimates for forms generated by the item bank as a function of the Rasch difficulty parameters of the BMCT–II items.

Effects of Training and Practice

Scores on the BMCT may be influenced by environmental factors, but not to the extent that important difficulties in interpretation are introduced. Research indicates that, while an individual's scores on the BMCT can generally be improved through training and experience, it is unlikely that dramatic improvement will result (Bennett, 1940). In a study involving applicants for technical defense courses, the average score on the BMCT for individuals reporting previous training in physics was 41.7 (*N* = 220, *SD* = 8.6), while the average score for individuals reporting no previous physics training was 39.7 (*N* = 95, *SD* = 8.9). Similarly, candidates for positions as fire fighters or police officers who reported previous physics training had an average BMCT score of 39.6 (*N* = 488, *SD* = 9.6), while candidates reporting no previous physics training had an average score of 34.0 (*N* = 983, *SD* = 9.8). Point-biserial correlations calculated using data from this sample revealed that physics training was moderately related to years of education (*r* = .26). Further analyses of these data indicated that the effect on BMCT test scores as a result of prior training in physics was to raise the mean by four raw score points, or less than one-half the standard deviation (Bennett, 1969).

Chapter 4

Evidence of Validity

The validity of a test is the single most fundamental and important aspect of test development and evaluation. Validity refers to the degree to which specific evidence supports the interpretation of test scores for their intended purpose (AERA, NCME, & APA, 1999). Although test developers are responsible for providing initial evidence of validity, the test user must evaluate if the evidence supports the use of the test for his or her intended purpose.

Evidence of Validity Based on Test Content

Evidence of content validity is based on the degree to which test items adequately represent and relate to the trait being measured. Test content also includes the wording and formatting of items, and the procedures for administering and scoring the test. Evaluation of content-related evidence is usually a rational, judgmental process (Cascio & Aguinis, 2005). Whether content-related evidence exists depends upon an evaluation of whether the same capabilities are required in both the job performance domain and the test (Cascio & Aguinis, 2005). For the BMCT–II, evidence based on test content should be established by demonstrating that the jobs for which the test will be used require the mechanical aptitude measured with the BMCT–II. The BMCT–II measures tasks, behaviors, knowledge, skills, abilities, or other characteristics necessary to perform a job that involves mechanical reasoning. Content-related evidence of the BMCT–II in training and instructional settings may be demonstrated by the extent to which the BMCT–II measures a sample of the specified objectives of such instructional programs.

Evidence of Validity Based on Test-Criterion Relationships

Selection tests are used to hire or promote those individuals most likely to be productive employees. The rationale behind selection tests is this: the better an individual performs on a test, the better this individual will perform as an employee.

Evidence of criterion-related validity refers to the statistical relationship between scores on the test and one or more criteria (e.g., performance ratings, grades in a training course, productivity measures). By collecting test scores and criterion scores, one can determine how much confidence may be placed in the use of test scores to predict job success. Typically, correlations between criterion measures and test scores serve as indices of criterion-related validity evidence. Provided that the conditions for a meaningful validity study have been met (sufficient sample size, adequate criteria, etc.), these correlation coefficients are important indices of the utility of the test.

Unfortunately, the conditions for evaluating criterion-related validity evidence often are difficult to fulfill in the ordinary employment setting. Studies of test-criterion relationships should involve a sufficiently large number of persons hired for the same job and evaluated for success with a uniform criterion measure. The criterion itself should be reliable and job-relevant, and should provide a wide range of scores. To evaluate the quality of studies of test-criterion relationships, it is essential to know at least the size of the sample and the nature of the criterion.

Assuming that the conditions for a meaningful evaluation of criterion-related validity evidence have been met, Cronbach (1970), characterized validity coefficients of .30 or better as having "definite practical value." The U.S. Department of Labor (1999) provides the following general guidelines for interpreting validity coefficients: above .35 are considered "very beneficial," .21 to .35 are considered "likely to be useful," .11 to .20 "depends on the circumstances," and below .11 "unlikely to be useful." It is important to point out that even relatively lower validities (e.g., .20) may justify the use of a test in a selection program (Anastasi & Urbina, 1997). This reasoning is based on the principle that the practical value of a test depends not only on validity, but other factors as well, such as the base rate for success on the job (i.e., the proportion of people who would be successful in the absence of any selection procedure). If the base rate for success on the job is low (i.e., few people would be successful on the job), tests of low validity can have considerable utility or value. When the base rate is high (i.e., selected at random, most people would succeed on the job), even highly valid tests may not contribute significantly to the selection process.

In addition to the practical value of validity coefficients, the *statistical significance* of coefficients should be reviewed. Statistical significance refers to the odds that a non-zero correlation could have occurred by chance. If the odds are 1 in 20 that a non-zero correlation could have occurred by chance, then the correlation is considered statistically significant. Some experts prefer even more stringent odds, such as 1 in 100, although the generally accepted odds are 1 in 20. In statistical analyses, these odds are designated by the lower case p (probability) to signify whether a non-zero correlation is statistically significant. When p is less than or equal to .05, the odds are presumed to be 1 in 20 (or less) that a non-zero correlation of that size could have occurred by chance. When p is less than or equal to .01, the odds are presumed to be 1 in 100 (or less) that a non-zero correlation of that size occurred by chance.

BMCT Evidence of Validity Based on Test-Criterion Relationships

Studies have generally shown strong support for the validity of the BMCT, based on evidence of test-criterion relationships. Kolz, McFarland, and Silverman (1998) reported that, in a study of

176 manufacturing employees, BMCT scores correlated .40 with supervisory ratings of mechanical understanding, and .27 with supervisory ratings of problem solving ability.

Muchinsky (1993) evaluated the relationships between the BMCT, a general mental ability test, and an aptitude classification test focused on mechanics, and supervisory ratings of overall performance for 192 manufacturing employees. Of the three tests, he found the BMCT to be the best single predictor of job performance (r = .38, p = .01). He also found that the incremental gain in predictability from the other tests was not significant.

Callender and Osbourn (1981) reported that the estimated true validity coefficients for the BMCT were .31 (*SD* =.17) for job performance criteria and .52 (*SD* =.07) for training performance criteria. These correlations were based on a sample of 38 job performance and 11 training performance validity studies. A wide variety of jobs were represented in the Callender and Osburn study, including maintenance crafts, process operators, and laboratory technicians.

Schmidt, Hunter, and Caplan (1981) found an estimated true score correlation of .33 between the BMCT and overall job/training performance for both operator and maintenance positions within the petroleum industry. Their findings were based on data for 1,800 operators and 706 maintenance personnel from 13 different organizations.

In studies conducted by Pearson in 2006, the relationship between BMCT scores and on-the-job performance of incumbents in various occupations, organizations, and industries was examined. Job performance was defined as supervisory ratings on behaviors determined through research to be important to jobs requiring mechanical reasoning ability (e.g., "demonstrates mechanical knowledge and expertise"). Supervisory ratings on a single-item measure of job potential were also obtained. The study found that BMCT scores correlated with job performance for all groups. Specifically, uncorrected correlations with supervisory ratings of mechanical understanding/problem solving ranged from .32 (manufacturing employees) to .50 (transportation trades/equipment operators). Furthermore, in four of the six groups, BMCT scores correlated significantly with supervisory ratings of potential (uncorrected r = .29 to .45).

BMCT–II Evidence of Validity Based on Test-Criterion Relationships

For the development of the BMCT–II, Pearson evaluated the relationship between the BMCT–II scores and on-the-job performance of incumbents at a U.S based tire manufacturing organization. Following completion of the BMCT–II by participants, managers provided ratings on their performance using a rating questionnaire. Each item referred to a particular work based behavior, with managers being required to rate performance on each on a seven-point scale ranging from 1 ("Well Below Job Requirements") to 7 ("Well Above Job Requirements"). The items related to Mechanical Understanding (e.g. "Demonstrates knowledge of machines and tools") or Problem Solving (e.g., "Applies sound logic to choose the best alternative or solution"). A Total Performance variable was calculated by summing scores on the items and separate Mechanical Understanding and Problem Solving variables were created based on scores on the relevant items associated with each. Managers were also asked to rate the participant's overall performance using a single item ranging again from 1 ("Well Below Job Requirements") to 7 ("Well Above Job Requirements"). A single item rating of job potential was also provided, using a scale ranging from 1 ("No Higher Than Current Job") to 5 ("More Than 2 Levels Above Current Job"). Correlations with and without corrections for range restriction were calculated. The results are displayed in Table 4.1 (uncorrected correlations with performance are shown in brackets). The results show significant correlations with both the Total Performance variable and Mechanical Understanding. Whilst the correlation between the BMCT–II scores and the single item performance rating was non-significant, the r value of .22 is above the U.S. Department of Labor (1999) threshold for a "likely to be useful" finding and would probably be significant with a larger sample.

There was no significant relationship found with problem solving or potential. The lack of correlation with potential could be down to the fact that a large proportion of the sample were highly experienced (33% with at least 10 years of experience, 29% with at least 15 years) and the question around potential may not be too relevant. As with the single item performance measure, the correlation between the BMCT–II and the single item potential variable could perhaps be explained by the relatively small sample used in the study.

Table 4.1 presents further criterion-related evidence of validity for various BMCT forms. The table includes statistical information for studies that reported validity coefficients. The first column of the table gives a description of each sample. The column entitled *n* shows the number of cases in the sample. The form of the BMCT administered is then noted, followed by the mean score and standard deviation. The criterion measures include training course grades and supervisor ratings, among others. Means and standard deviations are shown for the criterion measures. The validity coefficient for the sample appears in the last column.

Table 4.1 Studies Showing Evidence of Criterion-Related Validity

Group	N	BMCT			Criterion			r
		Form	Mean	SD	Description	Mean	SD	
Manufacturing employees at a large tire manufacturing company. (Pearson, 2014)	55	BMCT–II	2.1 (theta)	.71 (theta)	**Supervisor's Ratings:**			
					Mechanical Understanding	51.9	9.9	.34(.31)*
					Problem Solving	26.1	4.8	.19(.17)
					Total Performance	77.9	14.4	.30(.27)*
					Overall Performance (single item)	4.9	1.0	.22(.20)
					Overall potential (single item)	2.9	1.2	.07(.06)
Car Haulers at a Major Automotive Transport Company in the USA (Pearson, 2014)	216	S	—	—	Revenue generated	—	—	.36**
Manufacturing employees at a large Midwestern manufacturing company (Kolz, McFarland, & Silverman, 1998)	176	S/T	—	—	**Supervisor's Ratings:**			
					Mechanical Understanding	5.4	2.1	.40**
					Problem Solving	5.3	1.9	.27**

Table 4.1 continued

Group	N	BMCTs Form	BMCTs Mean	BMCTs SD	Description	Criterion Mean	Criterion SD	Criterion r
Manufacturing employees (e.g., machine operators, maintenance mechanics, quality control inspectors) at two plants (Muchinsky, 1993)					**Supervisor's Ratings:**			
	96	S	45.8	8.4	Overall Performance (Plant 1)	139.8	28.2	.40**
	97	S	48.4	8.6	Overall Performance (Plant 2)	141.2	30.1	.33**
	—	—	—	—	Overall Performance (Plants Comb.)	—	—	.38**
Job incumbents across multiple jobs (e.g., maintenance crafts, process operators, laboratory technicians) and industries (Callender & Osburn, 1981)	—	—	—	—	Training Performance	—	—	.52**
					Job Performance	—	—	.31**
Operator positions within the petroleum industry (Schmidt, Hunter, & Caplan, 1981)	1,800	—	—	—	Job/Training Performance	—	—	.33**
Maintenance positions within the petroleum industry (Schmidt, Hunter, & Caplan, 1981)	706	—	—	—	Job/Training Performance	—	—	.33**
Machine Operators at a large paper products manufacturer (Harcourt Assessment, Inc., 2006)	32–41	S/T	43.0	8.0	**Supervisor's Ratings:**			
					Mechanical Understanding/ Problem Solving	93.1	22.6	.45**
					Potential	2.3	1.1	.29

Table 4.1 continued

Group	N	BMCTs Form	BMCTs Mean	BMCTs SD	Description	Criterion Mean	Criterion SD	Criterion r
Job incumbents (primarily engineers) from a large chemical engineering company (Harcourt Assessment, Inc., 2006)	40–42	S/T	53.0	7.7	**Supervisor's Ratings:**			
					Mechanical Understanding/ Problem Solving	42.7	7.8	.32*
					Potential	2.4	0.9	.40**
Manufacturing employees (e.g., equipment operators) (Harcourt Assessment, Inc., 2006)	42–48	S/T	40.3	13.3	**Supervisor's Ratings:**			
					Mechanical Understanding/ Problem Solving	4.8	1.0	.32*
					Potential	2.4	1.1	.29*
Energy/Utilities employees (e.g., maintenance supervisors) (Harcourt Assessment, Inc., 2006)	38–40	S/T	43.8	9.9	**Supervisor's Ratings:**			
					Mechanical Understanding/ Problem Solving	4.8	1.0	.34*
					Potential	2.6	1.1	.01
College students in a work sample laboratory study (Mount, Muchinsky, & Hanser, 1977)	20	S/T	38.8	10.7	Work sample	30.9	7.7	.58**
					Criterion measure	62.8	13.4	.55*
	20	S/T	36.1	11.5	Work sample	27.3	7.4	.51*
					Criterion measure	56.6	11.5	.62**
Operators at a Southern chemical plant (Bennett, 1994)	87	S/T	45.1	7.0	**Supervisor's Ratings:**			
					Reaction to emergency	8.0	1.5	.36**
					Safety rules	7.4	1.0	.21
					Mechanical ability	7.1	1.4	.32**
					Job knowledge	7.7	1.7	.39**

Table 4.1 continued

Group	N	BMCTs Form	Mean	SD	Description	Criterion Mean	SD	r
Operators at a Southern chemical plant (Bennett, 1994)	136	S/T	48.0	7.5	Job knowledge test	64.0	12.8	.63**
Trainees at an Eastern refinery and chemical plant (Bennett, 1994)	54	S/T	51.8	6.9	Training course grades	—	—	.48**
Mechanics at a Northeastern utility company (Bennett, 1994)	35	S/T	47.0	10.4	Weighted criterion score (work samples & trade information test)	50.0	10.0	.64**
Apprentices at a major domestic steel producer (Bennett, 1994)	30	S/T	50.0	6.8	Average course grade	88.2	5.3	.54**
Knitting machine mechanics at a Southern textile plant (Bennett, 1994)	32	S/T	41.2	8.8	Supervisor's ratings	31.2	5.3	.37*
Experienced coal miners at a Midwestern underground coal mining company (Bennett, 1994)	83	S/T	48.5	9.0	Combined rating & ranking index	93.6	16.6	.23*
Inexperienced coal miners at a Midwestern underground coal mining company (Bennett, 1994)	178	S/T	50.9	8.1	Combined rating & ranking index	86.9	15.5	.23**
Technician trainees at a Northeastern utility (Bennett, 1994)	83	S/T	39.6	9.7	Speed in solving training modules	228.1	73.8	.52**
					Number of modules completed	4.4	0.8	.40**

Table 4.1 continued

Group	N	BMCTs			Description	Criterion		
		Form	Mean	SD		Mean	SD	r
Technical assistants at a Northeastern electrical utility (Bennett, 1994)	29	S/T	51.8	7.1	Weighted criterion score derived from performance on specific tasks & trade information test	50.0	10.0	.49**
Equipment operators at a Northeastern utility (Bennett, 1994)	31	S/T	51.6	7.0	Weighted criterion score derived from performance on specific tasks & trade information test	50.0	10.0	.39*
Customer service trainees at a Northeastern electrical utility (Bennett, 1994)	26	S/T	50.5	6.9	Special training exam	18.7	5.3	.56**
	35	S/T	51.1	7.5	Training exams	13.6	—	.41*

*$p < .05$. **$p < .01$. The studies by Callender and Osburn (1981) and Schmidt, Hunter, and Caplan (1981) were meta-analyses. With the exception of the two meta-analyses and the recent Pearson (2014) study sampling tire manufacturing employees, all correlation coefficients reported are uncorrected. Uncorrected correlations, using Thorndike's Formula 2 (Thorndike, 1947) for the Pearson (2014) study are shown in brackets. Pearson (2006) studies reported for manufacturing, energy, and transportation trades employees include some participant overlap with Pearson (2006) organization-specific studies.

Criterion-related validity information for BMCT forms developed prior to S and T (i.e., Forms AA and BB) is included in the 1994 BMCT Manual. Criterion-related validity information for the *Differential Aptitude Tests* Mechanical Reasoning subtest (DAT MR), a statistically equivalent form of the BMCT, may be found in the *Differential Aptitude Tests for Personnel and Career Assessment* Technical Manual (Bennett, Seashore, & Wesman, 1991), and the *Differential Aptitude Tests,* fifth edition, Technical Manual (The Psychological Corporation, 1992).

Published validity coefficients, such as those reported in Table 4.1, apply to the specific groups listed. Test users should not automatically assume that these data constitute sole and sufficient justification for use of the BMCT–II. Inferring validity for one group from data reported for another group is not appropriate unless the organizations and jobs being compared are demonstrably similar.

Careful examination of Table 4.1 can help test users make an informed judgment about the appropriateness of the BMCT–II for their organization. However, the data presented here are not intended to serve as a substitute for locally obtained data. Locally conducted validity studies, together with locally derived norms, provide a sound basis for determining the most appropriate use of the BMCT–II. Whenever technically feasible, test users should study the validity of the BMCT–II, or any selection test, at their own location.

Validity coefficients such as those reported in Table 4.1 apply to the specific samples listed. Sometimes it is not possible for a test user to conduct a local validation study. There may be too few incumbents in a particular job, an unbiased and reliable measure of job performance may not be available or there may not be a sufficient range in the ratings of job performance to justify the computation of validity coefficients. In such circumstances, evidence of a test's validity reported elsewhere may be relevant, provided that the data refer to comparable jobs.

Evidence of Concurrent Validity—BMCT–II to BMCT Form S

To support the equivalent construct being measured in the BMCT–II item bank as compared to the previous Form S, a study was conducted with 54 incumbents in various industries. The results shown in Table 4.2 indicate that the BMCT–II scores correlate very highly with Form S.

Table 4.2 Relationship Between BMCT–II and BMCT Form S

	BMCT Form S
BMCT–II	$r = .81^{**}$

$N = 54$ $^{**}p < .01$

Evidence of Convergent and Discriminant Validity

Evidence of convergent validity is demonstrated when scores on a test relate to scores on other tests or variables that purport to measure similar traits or constructs. Evidence of relations with other variables can involve experimental (or quasi-experimental) as well as correlational evidence (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Evidence of discriminant validity is demonstrated when scores on a test do not relate closely to scores on tests or variables that measure different traits or constructs.

Previous Studies of Convergent Validity Evidence

Evidence of convergent validity for the BMCT has been demonstrated in studies that examined its relationship with other mechanical ability tests. In 2006, Pearson conducted a study of the relationship between the BMCT and the Mechanical Reasoning subtest of the *Differential Aptitude Tests for Personnel and Career Assessment* (DAT MR, Bennett, Seashore, & Wesman, 1991). The DAT MR originally was designed as a statistically parallel version of the BMCT. The study consisted of 107 individuals employed in various roles and industries. As expected, BMCT scores correlated strongly (r. = 85, p = .01) with DAT MR scores.

Other studies have focused on correlations between BMCT scores and scores on cognitive ability tests without a specific mechanical component (e.g., verbal and quantitative abilities tests). These studies typically have reported a pattern of correlations that supports both the convergent and discriminant validity of the BMCT. That is, they generally have found that BMCT scores correlate higher with abilities that are more conceptually related (e.g., spatial ability) than unrelated (e.g., spelling ability). The largest of these studies focused on correlations of DAT MR (the statistically parallel version of the BMCT), with other abilities measured within the *Differential Aptitude Tests* test battery. This series of eight studies included samples sizes of 2790 to 11,120 students, and so provides relatively stable estimates of the relationships among these abilities. Results indicated that Mechanical Reasoning scores correlated highly with Spatial Reasoning (r = .56 to .62) and Abstract Reasoning (r = .57 to .60) scores, moderately with Verbal Reasoning (r = .48 to .53), Numerical Reasoning (r = .40 to .50), and Language Usage (r = .31 to .43) scores, and low with Spelling (r = .15 to .25) and Perceptual Speed and Accuracy (r = .02 to .12) scores.

Current Studies of Convergent Validity Evidence

For the development of the BMCT–II, a study including 50 job incumbents in various industries was conducted to study the relationship between the BMCT–II and the *Ramsey Mechanical Aptitude Test* (MAT). Results suggest that the BMCT–II scores are highly correlated with the Ramsey MAT (r = .81).

Table 4.3 presents correlations between various forms of the BMCT and other tests. Additional studies are reported in the previous versions of the manual. The correlation coefficient (r) indicates the relationship between scores on the BMCT and the comparison test. Higher correlation coefficients generally indicate a higher degree of overlap between the construct measured by the BMCT and the construct measured by the comparison test. Likewise, lower correlation coefficients generally indicate less overlap between the construct measured by the BMCT and the construct measured by the comparison test.

Table 4.3 BMCT Evidence of Convergent Validity

Group	*N*	BMCT			Other Test			*r*
		Form	Mean	*SD*	Description	Mean	*SD*	
Job incumbents across multiple jobs and industries (Pearson, 2014)	50	BMCT–II	0.8 (theta)	0.8 (theta)	*Ramsey Mechanical Aptitude Test*	25.9	5.7	.81**
Job incumbents across multiple jobs and industries (Pearson, 2006)	107	S/T	44.4	10.5	*Differential Aptitude Tests for Personnel and Career Assessment,* Mechanical Reasoning subtest	31.9	7.7	.85**
Adults, ages 16–54 (Kellogg & Morton, 1999)	28	S	45.8	11.1	*Beta III*	109.8	13.6	.46*
College students at a Southwestern university (Lowman & Williams, 1987)	149	S/T	41.8	6.1	*Revised Minnesota Paper Form Board Test*	43.5	9.1	.49**
					Watson-Glaser Critical Thinking Appraisal	56.1	8.9	.39**
					Raven's Standard Progressive Matrices	50.8	5.0	.50**
					Wide-Range Achievement Test, Arithmetic*	27.1	4.9	.37**
Manufacturing employees at 2 plants (machine operators, maintenance mechanics, QC inspectors) (Muchinsky, 1993)	192	S	45.8–48.4	8.4–8.6	*Flanagan Aptitude Classification Test,* Mechanics subtest	7.4–8.2	2.5–3.1	.64**
					Thurstone Language	28.3–31.6	8.7–9.5	.33**
					Thurstone Quantitative	21.6–2.2	6.4–7.2	.37**
College students in a work sample laboratory study (Mount, et al. 1977)	20	S/T	38.8	10.7	*Wonderlic Personnel Test*	30.9	7.6	.62**
	20	S/T	36.1	11.5	*Wonderlic Personnel Test*	31.8	5.7	.46*

Table 4.3 continued

Group	N	BMCT			Other Test			r
		Form	Mean	SD	Description	Mean	SD	
Parents of mathematically gifted students (Benbow, Stanley, Kirk, & Zonderman, 1983)	43 (M)	S/T	40.0	11.0	*Concept Mastery Test*	118.0		.48**
(M =males; F = females)	45 (F)	S/T	26.0	11.0	*California Test of Mental Maturity,* Language	24.1		.53**
					Kirk's Synonym-Antonym Test	35.0		.30
					Kirk's General Information Test	25.0		.67**
					Kirk's Test of Semantic Comprehension	15.0		.50**
					Kirk's Cubes Test	15.0		.53**
					Kirk's Rotation-Inversion Test	12.0		.69**
					Concept Mastery Test	120.0		.46**
					California Test of Mental Maturity, Language	23.5		.66**
					Kirk's Synonym-Antonym Test	35.0		.40**
					Kirk's General Information Test	23.0		.67**
					Kirk's Test of Semantic Comprehension	14.0		.51**
					Kirk's Cubes Test	11.0		.49**
					Kirk's Rotation-Inversion Test	12.0		.64**

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

Comparing Scores to Normative Data

The BMCT–II raw score is the total number of correct responses. Because little can be inferred from raw scores on an item-banked version of a test (i.e., difficulty levels can vary slightly from administration to administration), they are not presented in the profile report. Instead, the scoring algorithms take into account the exact difficulty level of the items each person completed. Specifically, raw scores are converted to estimated theta values (ability scores) using Item Response Theory (IRT—the 1-parameter logistic model, also known as the Rasch model; Rasch, 1980). The examinee's ability score is then automatically compared to the norm group that was selected at the time the administration was assigned.

Normative information (norms) provides a basis for evaluating an individual's score relative to the scores of other individuals who took the test. Norms provide a means by which raw scores can be converted to more useful comparative scores, such as percentile ranks. Typically, norms are constructed from the scores of a large sample of individuals who took a test. This group of individuals is referred to as the normative group or standardization sample; norms represent the performance of the group.

The characteristics of the sample used for preparing norms are critical in determining the usefulness of the norms. For some purposes, such as ability testing, norms that are representative of the general population are essential. For other purposes, such as selection from among applicants to fill a particular job, normative information derived from a specific, relevant, well-defined group may be most useful. However, the composition of a sample of job applicants is influenced by a variety of situational factors, including job demands and local labor market conditions. Because such factors can vary across jobs, locations, and over time, the limitations on the usefulness of any set of published norms should be acknowledged.

When test results are used in making employment selection decisions, the most appropriate norm group is one that is representative of those who will be taking the test in the local situation.

BMCT–II Norms

The BMCT–II Norms Composition Tables are presented in Appendix B. This section provides relevant information about each norm group to assist in selecting the most appropriate norm group. For example, test scores of a person applying for a position as a mechanic may be compared with norms derived from the scores of other mechanics. If a person has applied for an electrician position, his or her test scores should be compared with those of others in similar positions (e.g., Skilled Tradesperson).

Percentile Ranks

The percentile rank indicates an individual's position relative to the norm group. Percentile ranks are not the same as percentage scores, which represent the percentage of correct items. Percentile ranks are derived scores, which are expressed in terms of the percentage of people in the norm group who scored below a given raw score.

Although percentiles are useful for explaining an examinee's performance relative to others, there are limitations. Percentile ranks do not have equal intervals. In a normal distribution of scores, percentile ranks tend to cluster around the 50th percentile. This clustering affects scores in the average range the most because a difference of one or two raw score points may change the percentile rank. Extreme scores are less affected; a change in one or two raw score points typically does not produce a large change in percentile ranks. These factors should be taken into consideration when interpreting percentile ranks.

Using the BMCT–II as an Employment Assessment Tool

The BMCT–II was developed for adult assessment applications. In employee selection, the BMCT–II may be used to predict success in certain industrial, mechanical, and engineering jobs. It is also useful in monitoring the effectiveness of mechanical comprehension instruction and training programs, and in researching the relationship between mechanical comprehension and other abilities or traits.

Employment Selection

Many organizations use testing as a component of their employment selection process. Employment selection test programs typically use cognitive ability tests, aptitude tests, personality tests, basic skills tests, and work values tests to screen out unqualified candidates, to categorize prospective employees according to their probability of success on the job, or to rank order a group of candidates according to merit.

The BMCT has been used in the assessment of applicants for a wide variety of jobs, including electrical and mechanical positions in metals plants and aviation companies, skilled trades in automobile companies, computer operators, mechanics and equipment operators at utility companies, and operators in manufacturing. The BMCT–II may also be used effectively with other tests to evaluate candidates for jobs that require aptitudes other than mechanical comprehension. The relatively low correlation of the BMCT–II with other tests enhances the practical value of the BMCT–II in a test battery. For example, if the job requirements involve the use of critical thinking, a reasonably demanding mental ability test, such as the *Watson-Glaser Critical Thinking Appraisal*–second edition (WGCTA–II; Watson & Glaser, 2009), may also be useful. When job requirements involve critical and analytical thinking skills (e.g., for management-level positions), an instrument such as the *Watson–Glaser Critical Thinking Appraisal–II* may provide valuable supplementary information.

Do not assume that the type of mechanical comprehension required in a particular job is identical to that measured by the BMCT–II. Job analysis and validation of the BMCT–II for selection purposes should follow accepted human resource research procedures and conform to existing guidelines concerning fair employment practices. Local validation is particularly important when a selection test may have adverse impact; this is likely to occur with cognitive ability tests such as the BMCT–II (see Chapter 6). Though it is not unlawful to use a test with adverse impact (Equal Employment Opportunity Commission, 1978), the testing organization must be prepared to demonstrate that the instrument is job-related and consistent with business necessity. If use of the test is challenged, the organization must provide historical evidence of the test's validity in similar situations, evidence of content validity, or evidence of locally obtained criterion-related validity.

Making a Hiring Decision Using the BMCT–II

Human resource professionals can look at the percentile that corresponds to the candidate's raw score in several ways. Candidates' scores may be rank ordered by percentile so that those with the highest scores are considered further. Alternatively, a cut score (e.g., the 50th percentile) may be established so that candidates who score below the cut score are not considered further. In general, the higher the cut score is set, the higher the probability that a given candidate who scores above that point will be successful. However, the need to select high scoring candidates typically needs to be balanced with situational factors, such as the need to keep jobs filled and the supply of talent in the local labor market. Ultimately, it is the responsibility of the hiring authority to determine how it uses the BMCT–II scores. If the hiring authority establishes a cut score, that person should consider examinees' scores in the context of appropriate measurement data for the test, such as the standard error of measurement and data regarding the predictive validity of the test. In addition, selection decisions should be based on multiple job-relevant measures rather than relying on any single measure (e.g., using only BMCT–II scores to make hiring decisions).

When interpreting BMCT–II scores, it is useful to know the specific behaviors that an applicant with a high BMCT–II score may be expected to exhibit. These behaviors, as rated by supervisors, were consistently found to be related to BMCT–II scores across many different occupations requiring mechanical aptitude. In general, applicants who score low on the BMCT–II may not exhibit these behaviors, or they may find it challenging to effectively demonstrate these behaviors. Conversely, applicants who score higher on the BMCT–II tend to display a higher level of competence in these behaviors, and can effectively demonstrate

* knowledge of physical principles and laws,
* knowledge of machines and tools, and/or
* mechanical knowledge and expertise.

Applicants with high scores also may effectively demonstrate the ability to

* diagnose complex problems, involving machinery or equipment;
* install equipment, machines, wiring, or programs to meet specifications;
* imagine how something will look after it is has been moved around or when its parts are moved or rearranged;
* develop new ideas to address work-related problems;
* readily understand the implications of new information; and or

* discern subtle interrelationships among seemingly disparate pieces of information.

Human resource professionals who use the BMCT–II should document and examine the relationship between examinees' scores and their subsequent performance on the job. Using locally obtained information provides the best foundation for interpreting scores and most effectively differentiating examinees who are likely to be successful from those who are not. Pearson does **not** establish or recommend a passing score for the BMCT–II.

Fairness in Selection Testing

Fair employment regulations and their interpretation are continuously subject to changes in the legal, social, and political environments. Therefore, a user of the BMCT–II should consult with qualified legal advisors and human resources professionals as appropriate.

Legal Considerations

Governmental and professional regulations cover the use of all personnel selection procedures. Relevant source documents that the user may wish to consult include the *Standards for Educational and Psychological Testing* (AERA, NCME, & APA, 1999); the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 2003); and the federal *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission, 1978). For an overview of the statutes and types of legal proceedings which influence an organization's equal employment opportunity obligations, the user is referred to Cascio and Aguinis (2005) or the U.S. Department of Labor's (2000) *Testing and Assessment: An Employer's Guide to Good Practices*.

Differences in Mechanical Comprehension Test Scores Based on Sex

It is consistently reported in the literature on cognitive abilities that females perform better than males on tests of verbal ability, though males perform better than females on visual-spatial and mathematical ability tests (Halpern, 1986). The results of studies investigating differences in mechanical comprehension test scores based on the examinee's sex have been consistent with these findings (Bennett & Cruikshank, 1942; de Wolf, 1981; Lunneborg & Lunneborg, 1985; McCall, 1973). Although many studies indicate that sex-related differences on aptitude tests are diminishing, differences in mechanical aptitude are still commonly found to be of sufficient magnitude to be statistically significant (Feingold, 1988; Sapitula, & Shartzer, 2001).

In a review of the literature on gender differences in cognitive ability, Spelke (2005) stated that the gender difference in standardized test performance "likely reflects a complex mix of social, cultural, and biological factors" (p. 956). In a study conducted by Fortson (1991), a multiple regression analysis was used to explain score variance between sexes in BMCT scores. When education factors, work experience factors, and leisure factors were entered into the equation, gender explained only 2% additional unique variance. Regardless of the explanations for group differences in performance on cognitive ability tests, the literature indicates that these differences are consistent across tests and situations. The origins of observed differences in cognitive ability test scores based on sex are external to the test, assuming that the test is reliable and the construct valid.

Fairness of Using the BMCT With Female Applicants

"A lower mean test score in one group compared to another is not by itself evidence of bias" (Guion, 1998, p. 436). Group differences in mean test scores can represent actual differences in the ability being measured, as is the case with the BMCT–II. A more appropriate indicator of whether an ability test is biased is to evaluate whether the test is unequally predictive of job performance based on group membership (i.e., differential prediction).

In a study conducted in 2006, Pearson evaluated evidence of validity based on the relationship between the BMCT scores and on-the-job performance for 53 female job incumbents in various occupations and industries that require mechanical comprehension. *Job performance* was defined as the sum of supervisory ratings on 20 behaviors determined to be important to most jobs that require mechanical aptitude, as well as ratings on single-item measures of "Overall Performance" and "Overall Potential." Results of this study indicated that the criterion-related validity coefficients for the female sample were comparable to, if not higher than, those obtained for the predominantly male samples. A strong relationship between BMCT scores and job performance in the sample of females, was evidenced by high correlations with the sum of job performance ratings (uncorrected $r = .38$, $p = .05$, $n = 29$), ratings of overall job performance (uncorrected $r = .36$, $p = .05$, $n = 53$), and ratings of overall potential (uncorrected $r = .41$, $p = .05$, $n = 52$).

Group Differences/Adverse Impact

Local validation is particularly important when a selection test may have adverse impact. According to the *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission, 1978), adverse impact is indicated when the selection rate for one group is less than 80% (or 4 out of 5) of another. Adverse impact is likely to occur with aptitude tests such as the BMCT–II. Though it is within the law to use a test with adverse impact (Equal Employment Opportunity Commission, 1978), the testing organization must be prepared to demonstrate that the selection test is job-related and consistent with business necessity. A local validation study, in which scores on the BMCT–II are correlated with indicators of on-the-job performance, can provide evidence to support the use of the test in a particular job context. In addition, employers should conduct an evaluation that demonstrates that the BMCT–II (or any employment assessment tool) is equally predictive for protected subgroups, as outlined by the Equal Employment Opportunity Commission, will assist in the demonstration of fairness of the test.

Monitoring the Selection System

An organization's abilities to evaluate selection strategies and to implement fair employment practices depend on its awareness of the demographic characteristics of applicants and incumbents. Monitoring these characteristics and accumulating test score data are clearly necessary for establishing legal defensibility of a selection system, including those systems that incorporate the BMCT–II. The most effective use of the BMCT–II is with a local norms database that is regularly updated and monitored.

Employee Development

Some organizations provide career counseling services to employees as part of a formal career development program or as outplacement counseling as a result of employment termination. These services assist the employee in identifying and exploring career alternatives that are well-matched with his or her interests and aptitudes. In addition, career development counseling assists the organization in identifying candidates for training or other suitable roles within the organization.

Training and Instruction

The ability to comprehend mechanical principles is an important objective of many instructional programs in both business and educational settings. The BMCT–II may be used to measure the extent to which examinees have mastered certain aspects of mechanical comprehension or need training. The availability of an item bank that delivers equivalent, but different sets of items during each administration makes it possible to conduct pre- and post-testing to gauge the efficacy of instructional programs, as well as measure development of these aptitudes over an extended period of time.

References

Aiken, L. R. (1979). *Psychological testing and assessment* (3rd ed.). Boston: Allyn & Bacon.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, N.J.: Prentice Hall.

Antill, J. K., & Cunningham, J. D. (1982). Sex differences in performance on ability tests as a function of masculinity, femininity, and androgyny. *Journal of Personality and Social Psychology, 42,* 718–728.

Bechtoldt, H. E. (1972). [Review of the Bennett Mechanical Comprehension Test]. In J. V. Mitchell, Jr. (Ed.), *The seventh mental measurements yearbook* (pp. 1484–1485). Lincoln, NE: University of Nebraska Press.

Benbow, C. E., Stanley, J. C., Kirk, M. K., & Zonderman, A. B. (1983). Structure of intelligence in intellectually precocious children and in their parents. *Intelligence, 7,* 129–152.

Bennett, G. K. (1969). *Bennett Mechanical Comprehension Test, Forms S and T,* manual. San Antonio, TX: The Psychological Corporation.

Bennett, G. K. (1994). *Bennett Mechanical Comprehension Test, Forms S and T* (2nd ed.), manual. San Antonio, TX: The Psychological Corporation.

Bennett, G. K. (1981). *Bennett Hand–Tool Dexterity Test,* manual. San Antonio, TX: The Psychological Corporation.

Bennett, G. K., & Cruikshank, R. M. (1942). Sex differences in the understanding of mechanical problems. *Journal of Applied Psychology, 26,* 121–127.

Bennett, G., Seashore, H., & Wesman, A. (1991). *Differential Aptitude Tests for Personnel and Career Assessment,* technical manual. San Antonio, TX: The Psychological Corporation.

Callender, J. C., & Osburn, H. G. (1981). Testing the constancy of validity with computer-generated sampling distributions of the multiplicative model variance estimate: Results for petroleum industry validation research. *Journal of Applied Psychology, 66,* 274–281.

Cascio, W. E. (1982). *Applied psychometrics in personnel management* (2nd ed.). Reston, VA: Reston Publishing.

Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Ciechalski, J.C. (2005). [Review of the Bennett Mechanical Comprehension Test]. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook*. Retrieved from http://buros.unl.edu/buros/jsp/reviews.jsp?item=16182568

Cohen, B. H. (1996). *Explaining psychological statistics.* Pacific Grove, CA: Brooks & Cole.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum.

Crawford, J. E., & Crawford, D. M. (1981). *Crawford Small Parts Dexterity Test,* manual. San Antonio, TX: The Psychological Corporation.

Cronbach, L. (1970). *Essentials of psychological testing*. New York: Harper & Row.

Dagenais, F. (1992). Bennett Mechanical Comprehension Test: Normative data for a sample of Saudi Arabian technical trainees. *Perceptual and Motor Skills*, *74,* 107–113.

de Wolf, V. (1981). High school mathematics preparation and sex differences in quantitative abilities. *Psychology of Women Quarterly, 5,* 555–567.

Dimitrov, D. (2003). Reliability and true-score measures of binary items as a function of their Rasch difficulty parameter. *Journal of Applied Measurement, 4*, 222–233.

Equal Employment Opportunity Commission. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*, 38295–38309.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*, 95–103.

Finch, C. R. (1968). Predicting mathematics achievement in technical training. *Vocational Guidance Quarterly*, *1,* 193–197.

Fortson, H. D. (1991). An investigation of gender differences in mechanical aptitude. Unpublished doctoral dissertation, California School of Professional Psychology.

Geary, D. (1989). A model for representing gender differences in the pattern of cognitive abilities. *American Psychologist*, *44,* 1155–1156.

Guion, R. M. (1998). Assessment, measurement, and prediction for personnel decisions. Mahwah, NJ: Lawrence Erlbaum.

Halpern, D. (1986). *Sex differences in cognitive* abilities. Hillsdale, NJ: Erlbaum.

Hambleton, R. K. (1972). [Review of the Bennett Mechanical Comprehension Test]. In J. V. Mitchell, Jr. (Ed.), *The seventh mental measurements yearbook* (pp. 720–721). Lincoln, NE: University of Nebraska Press.

Harcourt Assessment. (2005). *Miller Analogies Test for Professional Selection,* manual. San Antonio, TX: Author.

Harris, A. J., & Jacobson, M. D. (1982). *Basic reading vocabularies*. New York: Macmillan.

Harsh, C. M. (1949). [Review of the Test of Mechanical Comprehension]. In O. K. Buros (Ed.), *The third mental measurements yearbook* (pp. 720–721). Lincoln, NE: University of Nebraska Press.

Hegarty, M., Just, M. A., & Morrison, I. R. (1988). Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning. *Cognitive Psychology*, *20,* 191–236.

Humphreys, L. G. (1949). [Review of the Test of Mechanical Comprehension], In O. K. Buros (Ed.), *The third mental measurements yearbook* (pp. 721–722). Lincoln, NE: University of Nebraska Press.

Kellogg, C. E., & Morton, N. W. (1999). *Beta III,* manual. San Antonio, TX: The Psychological Corporation.

Kolz, A. R., McFarland, L. A., & Silverman, S. B. (1998). Cognitive ability and job experience as predictors of work performance. *The Journal of Psychology, 132,* 539–548.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28,* 563–575.

Lowman, R. L., & Williams, R. E. (1987). Validity of self-ratings of abilities and competencies. *Journal of Vocational Behavior, 31,* 1–13.

Lunneborg, P. W., & Lunneborg, C. E. (1985). Nontraditional and traditional female college graduates: What separates them from the men? *Journal of College Student Personnel, 26,* 33–36.

McCall, J. N. (1973). Birth-order differences in special ability: fact or artifact? *Psychological Reports, 33,* 947–952.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*, 449–458.

Morton, N. W. (1953). [Review of the Test of Mechanical Comprehension]. In O. K. Buros {Ed.), *The fourth mental measurements yearbook* (pp. 720–721). Lincoln, NE: University of Nebraska Press.

Mount, M. K., Muchinsky, P. M., & Hanser, L. M. (1977). The predictive validity of a work sample: A laboratory study *Personnel Psychology, 30,* 636–645.

Muchinsky, P. M. (1993). Validation of intelligence and mechanical aptitude tests in selecting employees for manufacturing jobs. *Journal of Business and Psychology, 7*, 373–382.

The Psychological Corporation. (1980). *Manual supplement for the Bennett Mechanical Comprehension Test, Forms S & T.* San Antonio: Author.

The Psychological Corporation. (1992). *Differential Aptitude Tests* (5th ed.), technical manual. San Antonio: Author.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Roberts, O. H. (1972). [Review of the Bennett Mechanical Comprehension Test]. In J. V. Mitchell, Jr. (Ed.), *The seventh mental measurements yearbook* (pp. 1485–1486). Lincoln, NE: University of Nebraska Press.

Sanders, G., & Ross-Field, L. (1986). Sexual orientation and visuo-spatial ability. *Brain and Cognition, 5,* 280–290.

Sapitula, L. L. & Shartzer, M. C. (2001). Predicting the job performance of maintenance workers using a job knowledge test and a mechanical aptitude test. *Applied HRM Research, 6*, 71–74.

Satter, G. A. (1949). [Review of the Test of Mechanical Comprehension]. In O. K. Buros (Ed.), *The third mental measurements yearbook* (pp. 722–723). Lincoln, NE: University of Nebraska Press.

Schmidt, E. L., Hunter, J. E., & Caplan, J. R. (1981). Validity generalization results for two job groups in the petroleum industry, *Journal of Applied Psychology, 66,* 261–273.

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Spangler, M. (2005). [Review of the Bennett Mechanical Comprehension Test]. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook*. Retrieved from http://buros.unl.edu/buros/jsp/reviews.jsp?item=16182568

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60*, 950–958.

Taylor, S. E., Frackenpohl, H., White, C. E., Nieroroda, B. W, Browning, C. L., & Birsner, E. E. (1989). *EDL core vocabularies in reading, mathematics, science, and social studies.* Columbia, SC: EDL.

Thorndike, R. L. (1947). *Research problems and techniques*. Washington, DC: U.S Government Printing Office.

Trent, J., Yang, Z., & Rose, M. (April, 2007). *Confirmatory factor analysis of mechanical aptitude.* Paper presented at the Society for Industrial and Organizational Psychology Conference, New York.

U.S. Department of Labor. (1999). *Testing and assessment: An employer's guide to good practices.* Washington, DC: Author.

Watson, G., & Glaser, E. M. (2005). *Watson-Glaser critical thinking appraisal short form manual.* San Antonio, TX. Harcourt Assessment.

Wing, H. (1992). [Review of the Bennett Mechanical Comprehension Test]. In J. V. Mitchell, Jr. (Ed.), *The eleventh mental measurements yearbook* (pp. 106–107). Lincoln, NE: University of Nebraska Press.

Test Log

A test log should always be maintained. An example is provided below. This page may be photocopied for use of the BMCT in your organisation.

Testing Details

Organization	
Purpose of Testing	Selection / Development
Test(s) Used	BMCT–II/ BMCT Form S/BMCT Form T
Administration	Paper and Pencil / On-line
Test Administrator	
Test Invigilators	
Date	
Start time	
Finish time	

Candidate List

	Candidate Name		**Candidate Name**
1.		6.	
2.		7.	
3.		8.	
4.		9.	
5.		10.	

Materials Checklist

Materials	**Number Checked Out**	**Number Checked In**
Laptops		
Test Booklets		
Record Forms		
Test Instructions		
Stopwatches		
Pencils		
Erasers		
Pencil Sharpeners		

Disturbances / Unusual Occurrences

Any other comments

BMCT–II Norms Composition Tables

Bennett Mechanical Comprehension Test®–II
Normative Sample Composition

Values represent percentages of the norm group in each demographic category.

	Installation/ Maintenance/Repair	Skilled Trades	Machine Operators/ Machinists	General Labor	Vocational/ Technical Students	Mechanics	Engineers	Industrial/ Technical Sales
Sample Size	**125**	**121**	**109**	**111**	**108**	**111**	**105**	**114**
Industry	**%**	**%**	**%**	**%**	**%**	**%**	**%**	**%**
Aerospace, Aviation	2	–	2	1	1	11	10	4
Construction	14	50	7	13	13	1	9	11
Education	3	2	4	5	12	8	4	1
Energy, Utilities	6	9	4	2	6	1	9	7
Government, Public Service, Defense	5	2	7	5	2	5	10	3
Health Care	5	4	4	15	2	2	1	4
Information Technology, High-Tech, Telecommunications	32	2	59	1	12	8	16	25
Manufacturing & Production	9	8	2	14	5	9	26	11
Natural Resources, Mining	1	1	2	6	3	4	1	1
Pharmaceuticals, Biotechnology	2	2	6	–	1	–	3	4
Retail & Wholesale	2	3	–	–	3	–	1	18
Transportation Warehousing	1	–	–	–	1	–	1	2
Other	18	16	5	38	41	52	10	11
Years in Current Occupation	**%**	**%**	**%**	**%**	**%**	**%**	**%**	**%**
Not Applicable (Never Employed)	–	–	2	–	6	1	–	1
Less than 1 year	6	7	6	12	26	5	3	6
1 year to less than 2 years	10	5	12	16	24	15	6	14
2 years to less than 4 years	12	12	15	17	21	13	10	26
4 years to less than 7 years	14	17	24	22	7	16	22	18
7 years to less than 10 years	14	15	10	11	4	13	20	13
10 years to less than 15 years	17	15	8	16	6	16	20	12
15 years or more	27	30	23	6	6	21	19	9

Bennett Mechanical Comprehension Test®–II **Normative Sample Composition** *Values represent percentages of the norm group in each demographic category.*	**Installation/ Maintenance/Repair**	**Skilled Trades**	**Machine Operators/ Machinists**	**General Labor**	**Vocational/ Technical Students**	**Mechanics**	**Engineers**	**Industrial/ Technical Sales**
Sample Size	**125**	**121**	**109**	**111**	**108**	**111**	**105**	**114**
Education Level	**%**	**%**	**%**	**%**	**%**	**%**	**%**	**%**
Attended 12th grade or less, no diploma	1	1	3	2	6	2	–	–
High School diploma or GED	16	7	22	24	10	18	–	5
Some college credit, no degree	30	31	38	30	51	33	–	20
Associate's degree (AA, AS, etc.)	15	22	16	19	16	23	8	11
Undergraduate degree (BA, BS, BBA, etc.)	34	32	17	22	12	20	52	49
Graduate or Professional degree (MA, MS, JD, MD, PhD, EdD)	5	7	5	4	6	4	40	14
Sex	**%**	**%**	**%**	**%**	**%**	**%**	**%**	**%**
Female	17	32	29	55	29	13	28	48
Male	83	68	71	45	71	87	72	52
Ethnicity	**%**	**%**	**%**	**%**	**%**	**%**	**%**	**%**
Alaska Native	2	–	–	–	–	–	–	–
Asian	3	1	3	5	3	2	18	4
Black	18	17	28	17	19	11	8	16
Hispanic	11	7	7	12	6	12	8	7
Native American	5	–	3	1	1	–	–	–
Pacific Islander	–	2	2	–	–	–	1	–
White	66	73	56	65	72	75	66	72
Other	1	–	2	1	–	1	–	1
Age Group	**%**	**%**	**%**	**%**	**%**	**%**	**%**	**%**
18–20 years	1	2	6	5	32	10	–	2
21–24 years	6	4	6	6	21	12	4	2
25–29 years	10	11	14	13	12	14	21	28
30–34 years	17	15	16	14	13	11	22	15
35–39 years	14	15	14	18	5	5	15	9
40–49 years	27	26	22	22	7	30	23	16
50–59 years	20	25	20	21	8	15	14	20
60–70 years	6	2	4	2	1	4	1	9

Directions for Administration

The BMCT–II can be administered in a supervised or unsupervised setting and is only available online. The BMCT-I forms S and T are equivalent versions; however only Form S is available in paper format. The BMCT-I is suitable for supervised completion. The following directions are for the BMCT–II, which is administered only online.

The BMCT–II is administered through Pearson's online testing platform at pan (https://tara.vitapowered.com/your account name), an online system designed for the administration, scoring, and reporting of professional assessments.

Instructions for administrators on how to order and access inventory online are provided at the Platform Help section of the TalentLens website (http://us.talentlens.com/online-testing).

Although the BMCT–II is intended as a measure of power rather than speed, there is a 25-minute time limit for completing the test. Allow at least 35 minutes for total administration time—this includes setting up the online administration, giving directions, and actual testing.

Accommodating Examinees Who Have Disabilities

Find out about disabilities and special requirements that any candidates have and make arrangements to accommodate these. You should not change the standardized administration procedures without consulting an expert, as this can change the meaning of the scores. Contact Pearson TalentLens for advice if you are unsure about making accommodations.

You must provide reasonable accommodations for examinees with special needs to take the test comfortably. Reasonable accommodations may include, but are not limited to, modifications to the test environment (e.g., desk height) and medium (e.g., having a reader read questions to the examinee) (Society for Industrial and Organizational Psychology, 2003). In situations where an examinee's disability is not likely to impair his or her job performance, but may hinder the examinee's performance on the BMCT–II, the organization may want to consider waiving the test or de-emphasizing the score in lieu of other application criteria. Interpretive data as to whether scores on the BMCT–II are comparable for examinees who are provided reasonable accommodations are not available at this time due to the small number of examinees who have requested such accommodations.

The Americans with Disabilities Act (ADA) of 1990 requires an employer to reasonably accommodate the known disability of a qualified applicant provided such accommodation would not cause an "undue hardship" to the operation of the employer's business.

Test Security

Keep all test materials secure before, during, and after testing. Test scores should be stored in a password-protected folder. Allow only designated test administrators to have access to the BMCT–II tests.

As examinees complete the BMCT–II, their responses are uploaded to the platform scoring servers, which are encrypted. All candidate information is protected by encryption. Scores are confidential and accessible only to authorized individuals. It is unethical and poor test practice to allow individuals access to score information if they do not legitimately need it. Storing test scores in a locked cabinet or password protected file that can be accessed only by designated test administrators will help ensure security. The security of testing materials and protection of copyright must be maintained at all times by authorized individuals. Only disclose test access information, such as usernames or passwords, to authorized individuals.

All the computer stations used in administering the BMCT–II must be in locations that can be supervised easily with the same level of security as with paper-and-pencil administration.

Setting Up an Unsupervised Testing Session

Ensure that you have an appropriate account with TalentLens and that you have access to BMCT–II on the platform.

For Remote Testing: Ensure you have the correct email address for the candidates and send them invitations to the testing session. Go to your online account for the email invitation template. Inform them of the nature of the test, including how and why it is being used, the date, time, location and what they are required to bring with them (e.g. some testing centres require personal identification to be checked).

Familiarize yourself with the test format and platform.

Setting Up a Supervised Testing Session

The person responsible for administering the BMCT–II may not need special training (this differs among countries), but must be able to carry out standard examination procedures. The supervised testing session must be conducted according to standardized procedures to provide examinees with the same opportunity for doing well. To ensure accurate and reliable results, the administrator must become thoroughly familiar with the Directions for Administration and the test materials before attempting to administer the test. The best way for test administrators to familiarize themselves with the BMCT–II is to take the test prior to the administration, complying with the 25-minute time limit.

It is also recommended that the administrator be thoroughly familiar with the organization’s hiring and testing policies.

Schedule the testing sessions. Consider the duration of the session, the number of people to be tested, book an appropriate room, trained test administrators and additional invigilators (we recommend a ratio of at least 20:1 examinees to invigilators/proctors).

Before inviting candidates to a test session, ensure that the organization has sufficient test credits and computers for the number of examinees. Contact Pearson TalentLens customer service if you need to order online tests.

Make sure that the room is suitable for testing—consider size, space, layout, lighting, temperature, noise and possible distractions. Seat examinees apart and not directly opposite each other to avoid cheating and distraction. Ensure that potential disturbances are minimized, e.g., phones are turned off and “Testing in Progress” signs are used.

Make sure that all computer equipment is working and that the candidates have been added to the system.

Prepare the Test Log. This can function as a register and detail any reasonable adjustments to be made for candidates with disabilities, as well as any unusual occurrences. See Appendix A for a sample test log.

Try to engender a friendly, but purposeful atmosphere to put test takers at ease and enable them to work at their best. Start with an informal introduction to the testing session and introduce yourself.

Tell examinees:

- Who you are
- Your relationship to the organisation
- The purpose of the test
- How results will be used
- Who will have access to the results
- How the results will be stored (data protection)
- What will happen after the testing
- The logistics of the testing session: breaks, fire alarms expected, duration, toilets

Examinees do not need pencils or scratch paper for this computer-based test. After you have introduced yourself and accessed TalentLens.com, the BMCT–II initial instruction screen appears. Make sure all examinees are seated at their computers and say,

The on-screen directions will take you through the entire process, which begins with some demographic questions. After you have completed these questions, the test will begin. You will have 25 minutes to work on this test. Do you have any questions before starting the test?

Give examinees the opportunity to ask questions before you give the signal for the test to begin.

Answer any questions the examinee may have and say,

Please begin the test.

During the test, the examinee has the option of skipping items and returning to them later. The examinee may also review test items at the end of the test. Examinees have 25 minutes to complete the test.

If Technical Problems Occur During Testing

If an examinee's computer develops technical problems during testing, move the examinee to another suitable computer location. If the technical problems cannot be solved by moving to another computer location, contact Pearson TalentLens for assistance. The contact information, including phone and fax numbers, is on the TalentLens.com website.

Scoring and Reporting

Scoring is automatic and the report is available a few seconds after the examinee completes the test. A link to the report is available through your online portal (https://tara.vitapowered.com/your account name). You must have Adobe® Acrobat Reader® loaded on your computer to open the report. The test administrator may view, print, or save the candidate's report.