# Binet Kamat Test of Intelligence: Administration, Scoring and Interpretation - An In-Depth Appraisal

B.N. Roopesh

Additional Professor, Department of Clinical Psychology, National Institute of Mental Health and

Neurosciences, Bengaluru.

Corresponding author: B.N.Roopesh Email – bn.roopesh@gmail.com

# **ABSTRACT**

Binet Kamat Test of Intelligence (BKT) is one of the widely used test in India since several decades, especially in clinical settings. Compared to other popular and comprehensive IQ tests, such as Wechsler's tests, BKT is simple to administer, score and interpret; economical in terms of cost; and still a valid measure of intelligence despite the test was standardized several decades ago. These are the few of the reasons due to which many clinical psychologists use BKT in their routine institutional-hospital setting. However, using BKT without understanding some of the key issues can result in arriving at erroneous IQ and/or wrong conclusion about the intelligence of the subject. Some of the issues and concerns that are relevant and needs consideration in using BKT are: ratio IQ, Flynn effect, higher standard deviation, profile analysis and so on. The current article tries have an in-depth look at the issues and concerns regarding administration, scoring and interpretation, as well as tries to provide possible solution. The detailed explanations along with simple to understand figures, tables, examples and analogies used in the article can potentially benefit thousands of postgraduate psychology students and professionals who use BKT in theory and their practice.

Key words: Binet Kamat test, BKT, intelligence, IQ, Flynn effect.

(Paper received – 15<sup>th</sup> July 2020, Peer review completed – 10<sup>th</sup> August 2020; Accepted – 11<sup>th</sup> August 2020)

#### INTRODUCTION

Bombay-Karnatak version of Binet-Simon Intelligence Scale, otherwise popularly known as Binet Kamat Test or just BKT [1-3], is one of the old tests of intelligence that has been in use since several decades in India. Many Masters in Psychology courses have BKT or Stanford Binet test of intelligence in their syllabus and almost every M.Phil. in Clinical Psychology training institute in the country have BKT on their syllabus one way or the other. Many institutes have it as one of their main routine intelligence assessments tests, whereas few others have it in their theory. BKT was considered as one of the gold standards tests of intelligence. Only in recent times, its use has been marginally decreased, due to the introduction of other tests, for example, Wechsler's tests with Indian norms. However, with respect to the assessment of children, some of the recent tests also have limitations, making them difficult to use across the child population. It is not the scope of the current article to list out the limitations of the other tests of intelligence. However, given the limitations of the other tests, many clinical psychologists still stand by BKT. Further, BKT fulfils the face validity in the institutional-hospital setting, where majority of the population belong to middle to lower end of the socioeconomic status and/or from rural background. That is, compared to other tests, which apparently seem to yield lower IQ and/or percentile rank, BKT IQ somehow seem to match the child being tested. However, the question still remain that is this alone justifies the use of BKT in the current era.

While administering, scoring and interpreting BKT, students and clinicians still face several issues and concerns that require to be addressed and possible solutions proposed. For example, what adjustments and modifications are required for some of the items? The rationale for and how to 'adjust' (for convenience only 'adjust/adjusting/adjusted' terms will be used in this article from now on instead of prorate/convert) BKT IQ to match with Wechsler's/World Health Organization (WHO) guidelines of IQ reporting? Similarly, the rationale for and how to adjust Mental Age to match that of adjusted IQ? Does adjusting BKT IQ solve the inherent problem of that exists in BKT? Should you perform the commonly practiced profile analysis of BKT performance? What additional care and concern is required while arriving at and reporting the BKT IQ? Surprisingly, according to the available knowledge, there are no published literature in this regard. As several decades passed since BKT standardization and/or revision, and in the absence of any updating literature, it is possible that some students/clinicians/institutions follow incorrect method of administration, scoring and interpretation. This article attempts to address some of the above issues and concerns in depth and tries to provide solutions wherever possible.

### **ISSUES AND CONCERNS**

# 1. BKT and RATIO IQ

The term IQ, stands for Intelligence Quotient, which is a ratio of Mental Age and Chronological Age multiplied by 100 (i.e. MA / CA x 100 = IQ). Therefore, a score derived by the above process is referred to as ratio IQ. However, currently there are hardly any intelligence tests that use this ratio IQ, except in cases where one uses BKT. Almost all the intelligence tests currently available use 'deviation score', where a score obtained by a person is usually compared with the same age group. However, the term IQ is still used in majority of the intelligence tests even if they use deviation scores. This is due to the popularity of the term IQ.

However, it was not the case earlier. BKT was modified and standardized to suit Indian population in 1934. Ratio IQ was the standard practice at that time. Not until 1939 when David Wechsler introduced his Wechsler tests in US (Wechsler, 1939; 1946) did the concept of deviation IQ existed. Before this (and until Wechsler tests became popular), the method to arrive at IQ was clever. When an 8 years old child passes all the items of 8 years, the child's IQ would be 100 (*i.e. Mental Age (MA) 8 years divided by Chronological Age (CA) 8 years and multiplied by 100*). If a 10 years old child manages to pass only the items of 8 years, the child's IQ will be 80 (*i.e. MA 8 years divided by CA 10 years and multiplied by 100*). The calculation of IQ in this way has the appeal of clarity and simplicity. However, the use of it fell out of favor for several reasons. First being, it does not clearly differentiate among same age group of people as good as deviation IQ. For example, 1 or 2 years of growth in intellectual ability in the age range of 4-5 years corresponds to significant growth, in contrast to age 16 – 17 years. The whole concept of ratio IQ falters for older age groups, where one cannot differentiate the growth of intellectual ability. For example, how can one assess and/or differentiate the growth of intelligence between 25 to 27 years or between 35 to 40 years.

Further, ratio IQ method, on the lower end of the spectrum can yield biased IQs, and on the higher end of the spectrum can yield exorbitant IQs (which is not possible if one use deviation IQ). On the lower end, 7½ years mental age obtained by a 16 years old adolescent, yields an IQ of 47. Unless the social and adaptive functions are considered, the score of 47 will usually be categorized as moderate mental retardation/intellectual disability (to maintain the current usage practices, in this article the term 'intellectual disability' will be used instead of 'mental retardation). Which does not do justice to the particular adolescent in terms of what s/he can or cannot do in day-today life. On the higher end, there is theoretical and practical possibilities that a child as young as 8 years, can get a mental age of 22 years. In this case, the child's IQ can be 250, which is a totally an astronomical figure. One such case clearly depicts the perils of ratio IQ. In 1956, When Merilyn vos Savant was 10 years, she was assessed on Stanford Binet 1937 version, (19 years old with respect to 1956) which uses the ratio IQ. vos Savant obtained a mental age of 22 years 10 months, which yielded an IQ of 228. Due to which

she was considered (for a short period) as having the highest ever recorded IQ in the Guinness Book of World Records. This error of administering old version with ratio IQ was committed even when it was discouraged to concur high IQs above 170 by Binet test manuals [4]. If the appropriate test with recent standardization and/or with deviation IQ would have been used, vos Savant surely would not have obtained such a high score.

On the other hand, surprisingly, many students/professionals do not notice a statistical limitation inherent with the use of ratio IQ. Ratio IQs falls on the 'ordinal scale of measurement'. In ordinal scale of measurement, the difference of 10 points between 50 to 60 IQ points, is not the same as the difference of 10 points between 110 to 120 IQ points. This has significant implications for any study with ratio IQ that involves statistical analysis. Many students / professionals use statistics that are meant for 'interval scale of measurement', such as Pearson's correlation on BKT IQ, which is incorrect. Given the above, IQs derived using ratio method and with a test that is standardized decades ago, requires extreme caution in administration, scoring and interpretation. The following sections will discuss those issues and concerns separately.

### 2. MEAN, STANDARD DEVIATION OF THE TEST and THE NORMAL PROBABILITY CURVE:

Bombay-Karnatak version of Binet-Simon Intelligence Scale of Kamat (1934), i.e. BKT, was based on Stanford Binet Intelligence Scale 1916 - Terman version/edition (*For the purpose of brevity from now on this will be referred to as SB-1916 version*). The mean of BKT is 99.8 which is close to 100, matched SB- 1916 version. However, as against the SB-1916 version which had the standard deviation of 13 (Kamat, 1963; 1967), BKT had 18.7 as the standard deviation (SD). Before discussing how such a wide SD affect the performance and reporting of the IQs, it is intriguing to analyze how such a difference resulted in BKT.

Kamat [1] followed almost same test items as Terman [5]. However, with relatively extensive rearrangements of items to particular age and few modifications in test items to suit Indian sociocultural milieu. In terms of standardization there were differences between Terman's and Kamat's sample. According to Terman, for his 1916 edition, he had 2300 subjects which included "1700 normal children, 200 defective and superior, and 400 adults". According to Kamat, his Bombay-Karnatak sample consisted of "advanced Hindus 653, intermediary Hindus 333, backward Hindus 17, Mohammedans 39 and Christians 32". Further, Terman's version had 90 items and 12 age groups, whereas Kamat's version had 99 items and 13 age groups [2-3].

Another interesting difference between the two tests is the placing of the items in a particular age. Information about developmental pattern of abilities and psychometric properties were used to place each item to its respective ages in the test in Binet – Simon test of 1911 [6], SB-1916 and BKT editions.

- Binet used 2/3<sup>rd</sup> to 3/4<sup>th</sup> criteria to place items in their respective age levels. That is, if 2/3<sup>rd</sup> to 3/4<sup>th</sup> of the children of a particular age (e.g. 6 years old) passed an item, then that item was placed in that year (year VI).
- In comparison, Terman's (1916) version followed slightly different criteria, where he allocated a test item to a particular age, when 65% of children (*within 2 months of that particular age*) pass that item.
- Cyril Burt [7] on the other hand, followed different criteria, where if 50% of the children who are between 6 to 7 years old pass an item, the item was placed in year VII (year seven). That is, for example, out of 200 children who have completed their 6<sup>th</sup> birthday and yet to reach 7<sup>th</sup> birthday, if 100 of them (i.e. 50%) can pass a particular item, then that item was allocated to year VII [3].
- However, Kamat found it difficult to exactly follow 50% pass criteria. Based on the results that he obtained with his sample, his placement followed an average of 42% [3]. That is on an average, for example, out of 200 children who have completed their 6<sup>th</sup> birthday and yet to reach 7<sup>th</sup> birthday, if on an average 84 of them (i.e. 42%) can pass a particular item, then that item was allocated to year VII (year Seven).

Given these, it appears that all the above factors (i.e. differences in the sample characteristics, geographical differences, differences in the number of items and 42% pass criteria for item assignment to particular year) would have played a role

in BKT's SD being wider compared to the Terman's version. Given these differences, based on the estimation obtained from data of both the tests/versions,

- Terman's showed about 0.5 percent of people score above 140 IQ and 1 percent score below 70 IQ.
- Kamat's showed that 2 percent of the people score above 140 IQ and only 0.93 percent between 40 to 60 IQ [3].

Currently, almost all the intelligence tests try to have a mean of 100 and an SD of 15, to match with that of Wechsler's scale [8-9]. Even World Health Organization also recommends a mean of 100 and SD of 15 to classify intelligence, especially the intellectual disabilities [10]. There is no any particular sanctity with mean being 100 and standard deviation being 15. It was felt that due to popularity of the Intelligence tests, 100 IQ came to be known as the exact normal IQ. Therefore, the test developers tried to achieve a mean of 100 in their tests, especially if it is an intelligence test. Similarly, in terms of standard deviation, Wechsler opted for 15, as it is a good number which can be easily divisible by 5 and easy to calculate different IQ range [8-9].

However, the cut-off points to determine the intellectual disability or intellectual superiority are strictly based on the normal distribution characteristics in any population. The values are usually explained with mean and standard deviations. The percentage of people is said to fall between ±1 SD is 68.26, between ±2 SD is 95.44, and between ±3 SD it is 99.73. Further, it is always fixed that a cut-off of minus 2 SD, i.e. a score less than 2 SD is considered as intellectual disability. Going by this, about 2.28 (if 2 SD is 95.44; then 100 - 2 SD = 4.56; so, for each side it will be 4.56 divided by 2 results in 2.28) percent of the population is categorized as intellectually disabled, irrespective of the intelligence tests used.

Given the normal distribution and standard score, test constructors during the standardization of their tests design; select, adapt and tweak their test items to get an overall exact mean of 100 and SD of 15. This is the reason, irrespective of increase in the IQ of a generation, a good standardization will still result in about 2.28 percentage of people being categorized as intellectually disabled. An advantage of normal distribution curve standard scores is that it adheres to the 'interval scale' of measurement, compared to 'ordinal scale' of ratio IQ. That is 10 IQ points forms an exact interval whether the IQ is between 66 to 76, or between 123 to 133 [4]. Due to this deviation IQs can be subjected to the statistical analysis meant for interval scales.

Going by the fact that 2 SD less than the mean forms the cut-off for the intellectual disability, BKT's mean of 99.8 (for practical purpose, it can be considered as 100) and SD of 18.7, result in about an IQ of 62.6 (for practical purpose we can consider it 63), below which, it can be diagnosed as intellectually disabled (1 SD is 18.7, and 2 SD is 18.7 x 2 = 37.4; therefore 100 - 37.4 = 62.6). These intermediate values (non-divisible by 5) has the potential to create lots of confusion among professionals as well as parents and educationists. It becomes even worse, if the category of average/normal, low/high average, mild/moderate/severe disability is involved. Table 1 clearly shows the difference between BKT and WHO classification of IQ according to the mean and SD of BKT and Wechsler's tests.

Table 1: showing the difference between BKT and Wechsler's/WHO IQ classification							
Normally used categ	gories	BKT IQ	Wechsler's/WHO IQ				
Very superior		Above 137	Above 130				
Superior		125 – 136	120 – 129				
High/Above avera	age	113 – 124	110 – 119				
Average		87 – 112	90 – 109				
Low / Below aver	age	75 – 86	80 – 89				
Borderline		64 – 74	70 – 79				
	Mild	38 – 63	50 – 69				
Intellectual Disability/	Moderate	19 – 37	35 - 49				
Mental Retardation	Severe	Below 19	20 – 34				
	Profound	(BKT does not	Below 20				
		differentiate between					
		severe and profound)					

In addition to Wechsler's /WHO IQ range being easy to understand, follow and communicate among fellow professionals (*compared to BKT*), sticking to 100 mean and 15 SD has further benefits. As standard scores and with respect to normal distribution, they can also be represented and explained in terms of Percentile points, i.e. percentage of the people obtaining lower score than yours. For example, Wechsler's IQ of 130 corresponds to about 98 Percentile point. This 98 Percentile point will be same irrespective of the test administered, as long as the test has a mean of 100 and an SD of 15.

If in case, the intelligence test has a different mean and/or SD then the above interpretation will not hold good. For example, as BKT SD is 18.7, it is understood that the normal probability distribution curve (NPC) will have a different characteristics and shape, making it different and difficult to use the IQ categorization and IQ interpretation. For example, figure 1 and 2 clearly depicts the differences of having different SDs. In figure 1 both person A and B obtain an IQ of 70. However, both do not show same/similar intelligence, due to the differences in the distribution. Therefore, in figure 1, person B has more intelligence compared to person A. Similar logic applies to figure 2, where on the other hand, in spite of both person A and B getting the same IQ of 130, it is person A who actually has more IQ.

Figure 1: Showing the IQ distribution with respect to two different SDs on lower end of the NPC

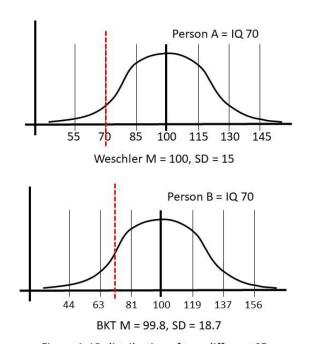


Figure 1: IQ distribution of two different SDs

Further, many professionals and students in India, are not aware about the fact that whenever the SD is wide/high, it is easy and have more chances to score extreme scores on either side. That is, it is easier to score high IQ if the SD is 18.7 compared to if SD is 15. Figure 3 clearly illustrates this phenomenon, where for example, one person out of hundred can get an IQ of 145 on a test that has 15 as SD, compared to one person out of hundred getting an IQ of 166 if the test has 18.7 as SD.

Figure 2: Showing the IQ distribution with respect to two different SDs on upper end of the NPC

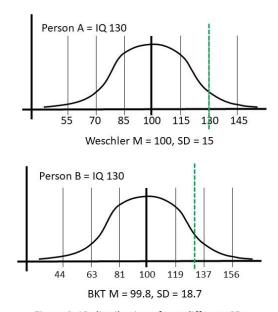


Figure 2: IQ distribution of two different SDs

For some students / professionals, the above might be difficult to comprehend as to how it might increase the IQ if a test has wide SD. Practically having wide SD implies at least two simple things.

- First, is that the population studied in that particular aspect (i.e. intelligence here) show wide/extreme variation on both sides.
- Second, is that the test items are easier to pass/complete on the upper side and vice versa.

Usually if the SD is wide, it would indicate both first and second (*factors mentioned above*) might interact and both might be responsible if one obtains extremely IQ.

Figure 3: Showing the difference between two SDs and their distribution of IQs

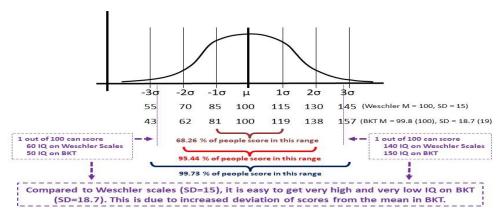


Figure 3: Showing the difference between two Standard Deviations

For some professionals / students, all the above might appear too much complex in terms of test standardization process, and due to which, they might think that it might not be relevant for their everyday clinical practice. But it is not so. As we all know that, the type of test we use has enormous implications in terms of clinical diagnosis, understanding the person, selection for a course/training/job, and planning for the future of the child/person. To further explain, for example, using BKT (compared to say Wechsler's tests) results in two unfair advantages, one is that BKT uses ratio IQ (Case of Merilyn vos Savant, as explained above), and the other is BKT has high/wide SD (as explained in figure 3).

Given the above, if we assume that two candidates with exact same intellectual abilities apply for a higher degree. But there is only one vacancy and the criteria for selection is that whoever has higher IQ will get the seat. 'Person A' goes to a psychologist who uses BKT. On the other hand, 'person B' goes to another psychologist who uses Wechsler's test. However, despite both having equal intellectual abilities, person A's results show that she has higher IQ compared to person B. Which is erroneous and will be injustice to person B. However, one may argue that for the clinical purpose we don't see much difference and many times the obtained BKT IQ seem to match with the patient's abilities. This intriguing and interesting aspect will be discussed in detail below in a separate section (section VIII).

# 2a. Adjusting BKT IQ to match with that of Wechsler's/WHO recommended IQ

The above two sections clearly highlight the main limitation of the BKT, which is that of ratio IQ and the SD being 18.7, which does not match the current global practice. Further, due to the SD being high it is almost impossible to differentiate between severe and profound intellectual disability in BKT (refer table 1). The best way to overcome this limitation is to re-standardize the BKT to suit the current standards, which is to have a deviation IQ with SD being 15. However, till not it has not happened. This is probably can be attributed to two reasons. First, re-standardization is a herculean process, which require enormous amounts of time, financial and human resources. Second, easy availability of other intelligence tests, such as Wechsler's tests and Raven matrices.

However, one of the possible ways to overcome some of this limitation is to mathematically convert the BKT IQ to IQ that matches WHO's requirement of SD of 15. This involves adjusting the SD of 18.7 to SD of 15. In this regard, according to the limited available information, few decades back in India, a small group with Craig Gonsalvez was assigned to find a fix to this problem. Accordingly, the team came up with a formula to adjust the BKT IQ to IQ that adheres to SD of 15. The simple formula for this is to, 'Subtract the BKT IQ from 100, divide the result by 18.7, multiple the output with 15, and again subtract the result with 100', to obtain the adjusted IQ. For example, if the BKT IQ is 50, then the adjusted IQ would be 60. Here one has to report the IQ as "adjusted IQ 60" for all the practical purpose.

Below step by step converting process will explain this.

```
100 - 50 = 50
50 / 18.7
2.6738 \times 15 = 40.0
100 - 40.0 = 60
```

On the other hand, the following formula can also help

Figure 4 – Formula to adjust the 18.7SD to 15SD

One should remember that the above formula adjusts IQs both side of the NPC. That is, if the BKT IQ is less than 100, the adjusted IQ will be higher than the BKT IQ. If the BKT IQ is higher than 100, then the adjusted IQ will be lesser than the BKT IQ. For example, if the BKT IQ is 115, then the adjusted IQ will be 112.

```
100 - 115 = -15

-15 / 18.7 = -0.80

-0.80 \times 15 = -12.0

100 - (-12) = 112
```

# 2b. Adjusting Mental Age (MA):

For few decades, the above formula to adjust BKT IQ (figure 4) was adopted by several institutions and clinical psychologists. However, there was one short coming about the reporting of this adjusted IQ that nobody had noticed. BKT provides both MA and IQ. Given this, if the child / client scores anything that is less/more than the average IQ, it was observed that the psychologists, while providing feedback about IQ results, mainly used to explain the MA (to the family members of the children/client and/or clients / patient themselves). This was due to the fact that, in the absence of percentile points, MA is relatively easy to explain/understand compared to IQ, when discussing about the child's/client's abilities and where do they stand among their peers. For example, if a child of 10 years obtains a MA of 6 years, her BKT IQ would be 60, and her adjusted IQ would be 68. However, while giving the feedback, it was observed that both professionals and students still resorted to the MA of 6 years. This was because, the formula was only to adjust the IQ, but not to adjust MA. In this regard, Roopesh came up with a formula to adjust the MA. Initially, to adjust the MA, he used to follow his technique that involved 3 steps. However, Kumble, while using the technique suggested that instead of 3 steps it can be done with just 2 steps (refer figure 5). This made the formula simple and easy to remember [10]. One has to remember that adjusting MA, should be done only after adjusting BKT IQ.

Figure 5: Formula to adjust BKT MA to match the Adjusted IQ



Figure 5: Formula to adjust MA

# 2c. New proposal/recommendation for a better process of adjustment of BKT IQ

One of the important facts is that, BKT has different SDs for different age groups. It varies from about 16 SD for 2 to 8 years, about 20 SD for 8 to 14 years, and about 15 SD for 14 years and above. The SD of 18.7 is for the entire BKT standardization sample (refer table 2). Given this, using a common SD of 18.7 irrespective of the age of the child, can yield erroneous results. In addition, apart from the variation in SD, the Mean IQ also shows variation from 95.5 to 104.8, which is a difference of almost 9 IQ points (Table 2).

Though not frequently, but a student/professional can ask, 'What difference does these small variations in the SD and the Mean IQ can make?'. However, one should remember that, these are not small variations. It is a 9 points variation in Mean IQ and 8 points variation in SD (refer table 2). This can actually lead to misdiagnose/misclassify the intelligence of the child, i.e. between borderline intelligence and mild intellectual disability or between average to above average intelligence, or between presence and absence of SLD (though the latter is not based on BKT results). Further, however small it is, it makes a lot of difference for that one particular child (and the parents) on which the assessment is being done.

Chron			
From	To	Mean IQ	SD
2 years	3 years 11 mths	104.8	15.0
4 years	5 years 11 mths	99.1	17.1
6 years	7 years 11 mths	95.5	16.0
8 years	9 years 11 mths	100.4	20.5
10 years	11 years 11 mths	99.8	23.0
12 years	13 years 11 mths	102.9	18.8
14 ye	ars & above	98.8	15.2
For the	99.8	18.7	

Table 2: Showing the Mean IQ and SD of the BKT standardization sample

Given the extent of the differences in mean and SD for different age groups, it can be questioned that whether using the whole sample mean and SD (mean of 99.8 and SD of 18.7) for all ages, will do justice for the child/person. A figure depicting how does the IQ distribution vary with respect to mean and SD in different age groups are shown in figure 6.

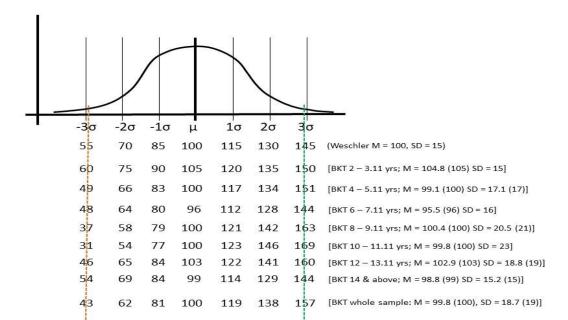


Figure 6: Showing the difference of mean and SD at different age groups in BKT

Figure 6: Showing the difference of mean and SD at different ages of BKT standardization sample. Numbers within the round brackets are rounded to the nearest whole number

As depicted in figure 6, it can be clearly understood that how IQs can vary depending upon the mean and SD. That is, approximately at -3 SD on the NPC, one person out of 100 can get the following IQ (brown dotted line). Here, depending on the mean and SD, the child's IQ can vary from 31 to 60, which is a very huge difference. Similarly, on the other side, approximately at 3 SD, one person out of 100 can get the following IQ (green dotted line). It can be seen that the IQ vary from 144 to 169, which again is a huge difference.

If there are no options or possibilities to correct the above limitations, then one can resort to the usual known processes of using the formula given in figure 4. However, when there can be a better option, to reduce these limitations, then one has to explore those options. This will benefit almost every child on whom BKT has been done. The better option is explained below.

# The better option

Instead of using the whole sample BKT IQ mean and SD (100 and 18.7) for proration/adjustment (as given in figure 4), one can use the particular mean and SD of that particular age level for the proration. That is substitute the whole sample BKT IQ mean and SD with that particular age level IQ mean and SD in the adjustment formula. This better, separate for each age group formulas are given in the figure 7. One should remember that, along with adjusting the SD, the better-option-formulas given below also adjusts the mean IQ. The rationale for the adjustment of the mean IQ is explained in subsection IId below.

Figure 7: Formulas to adjust according to specific age group

From 2 to 3 years 11 months 
$$\rightarrow$$
 100 —  $\boxed{(105 - BKT IQ)}$  = Adjusted IQ

From 4 to 5 years 11 months  $\rightarrow$  100 —  $\boxed{\frac{(99 - BKT IQ)}{17}}$  x 15 = Adjusted IQ

From 6 to 7 years 11 months  $\rightarrow$  100 —  $\boxed{\frac{(96 - BKT IQ)}{16}}$  x 15 = Adjusted IQ

From 8 to 9 years 11 months  $\rightarrow$  100 —  $\boxed{\frac{(100 - BKT IQ)}{21}}$  x 15 = Adjusted IQ

From 10 to 11 yrs 11 months  $\rightarrow$  100 —  $\boxed{\frac{(100 - BKT IQ)}{23}}$  x 15 = Adjusted IQ

From 12 to 13 yrs 11 months  $\rightarrow$  100 —  $\boxed{\frac{(103 - BKT IQ)}{19}}$  x 15 = Adjusted IQ

For 14 years and above  $\rightarrow$  100 —  $\boxed{(99 - BKT IQ)}$  = Adjusted IQ

It is actually surprising that for years 14 and above, the mean IQ and SD obtained by Kamat, is almost similar to the current WHO/Wechsler's standard of 100 mean and 15 SD (refer table 2). Due to it the formula (refer figure 7) for age 14 and above will yield just one IQ difference. Among the formulas given in figure 7, the first formula, i.e. for the age 2 to 3 years 11 months, differs compared to other age level formulas in the same figure. This is due to the fact that, for this age group (as well as to 14 and above age level) the BKT SD is 15, which is equal to that of WHO recommend SD. So, therefore, there is no need to adjust the SD (mathematically speaking, if you divide a value by 15 and then multiply it 15, the value will be the same).

For those examiners who find it difficult to use mathematical formulas, they can adopt the following procedure.

Assume that if your client is a 7 years old child and obtains the IQ of 72. Then as the BKT mean will be 96 and SD is 16 for a child of 7 years (refer table 2).

- BKT mean for age group 7 is 96, so 96 minus BKT IQ 72 = 24
- 24 divided by BKT SD 16 = 1.5
- 1.5 multiplied by WHO SD 15 = 22.5
- 100 minus 22.5 = 77.5

Therefore, the adjusted IQ will be 77.5, i.e. 78. Which is 8 points higher than what the child originally got. Detailed justification to use the separate formulas for each age group is provided in section 2f below.

After using the formula (based on the age level of the child/person) given in figure 7, the examiner has to further use the mental age adjusting formula given in figure 5.

# 2d. Rationale for adjusting the mean IQ:

Those who are used to adjusting BKT IO to Wechsler/WHO standard, usually know that they are doing it so to adjust for the difference in SD only. There was no need to worry about adjusting for the mean IQ difference, as the total mean IQ of the BKT standardization sample (i.e. 99.8) almost matched that of Wechsler/WHO mean recommended IO (i.e. 100).

However, as observed in 2c that there is a more efficient way to adjust than having only one standard adjusting formula. It is clearly seen in table 2 that, mean IO differs along with SD with respect to different age groups. Given this, adjusting only for SD difference is not the correct procedure.

An example would explain better. For time being let us assume that the SD of BKT is same as Wechsler's/WHO standards, but the mean differs. This is actually observed in BKT for ages 2 to 3 years 11 months where the SD of both is 15, but the mean of BKT is 105 (refer table 2). Therefore, if we consider this example, i.e. BKT IQ being 105 (and SD being the same 15). It presents a different scenario, where the classification of intelligence changes (refer figure 6 and 8). For example, at minus 1 SD, the IQ will be 90 in BKT distribution (i.e. as the mean IQ is 105 and SD is 15), instead of 85 which is the standard practice. Given this the whole cut-off for the intelligence categorization would also change, for example, instead of 90 to 109 as the 'average' intelligence range, it would be 95 to 114 (if BKT mean IO of 105 is followed). That for every IO category there will be a shift of 5 IQ points, for example for 'Borderline' intelligence it would be 75 to 84. Therefore, to avoid this big error, one needs adjust mean IQ along with SD. This adjustment process is depicted in figure 8. The top part of the figure 8 shows NPC has shifted to the right when the mean IQ is 105 (red-dashed-NPC). Associated with it, the IQs for each SD range also shift to the right. Notice the blue-dot-vertical line which is where the IQ of 80 would be in the NPC (on the right-dashed-NPC). Therefore, to fix the error, all one has to do is to adjust the NPC. In this case, visually/figuratively speaking, to slide the red-dash-NPC back towards the black-solid-NPC (towards the mean of 100) as depicted in the lower part of the figure 8. Again, visually/figuratively speaking, once it is moved to the left, the red-dash-NPC will be same as black-solid-NPC. Notice the blue-dot-vertical-line also moved back got adjusted, so now the IQ would be 75. This figure is only to visually show how it can be done and how it will appear once it is done. However, to actually achieve this, one has to use the formula shown in the figure 7, which adjusts both mean IQ as well as the SD. Detailed justification to use the separate formulas for each age group is provided in section 2f below.

# 2e. Limitations of adjusting IQ

It might appear that by adjusting the BKT IQ to the more acceptable IQ and then by adjusting the MA will solve the main limitation of the BKT. Even though adjusting make it acceptable, it is still far from perfect. This is because, the very adjusting process is outside of the standard BKT testing-scoring procedure and not inherent to the BKT testing itself. There are many reasons that this simple mathematical adjustment might not fully compensate for the limitations to the BKT. That is, a mathematical formula cannot substitute the test process or the items in the test. Closely associated with the above is that the difference between linear and nonlinear data. In addition, BKT's assignment of items to different age levels and usage of ratio IQ complicates the whole adjusting process. These are explained in detail below.

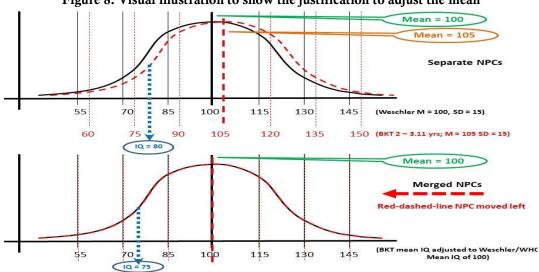


Figure 8: Visual illustration to show the justification to adjust the mean

It is understood that like all routine mathematical/statistical analysis, the above mathematical formula for adjusting BKT IQ applies mainly to linear data/values. That is, the adjusting formula applies only if the performance on BKT follows a linear path/approach. However, unlike Wechsler's tests which has test items that sequentially tests the same cognitive function/s from easy to difficult on one dimension (*Block Design from simple to difficult level*), BKT has uneven distribution of the test items that tests different/multiple cognitive functions randomly. Given these issues, it can be said that performance on BKT does not follow linearity.

Associated with this, BKT uses age scale format (where passing an item will yield scores in terms of months) where the items are not sequentially assigned according to particular ability being tested (more on this is discussed in the section on 'profile analysis in BKT'). Further, as mentioned earlier, ratio IQ falls on the 'ordinal scale of measurement', which has its own limitations in terms of what type of analysis can be done.

Whether linear or non-linear, whether age scale or point scale; irrespective of all the above mentioned limitations in adjusting, a psychologist has to remember that it is very important that 'ADJUSTING IQ (both mean and SD) HAS TO BE DONE', as it is the only possible way to overcome the limitations inherent in BKT scoring and interpretation.

# 2f. Justification for the usefulness to adopt different adjusting formulas for different age groups

Some might argue that there is no need to adjust based on different age groups as suggested above. They might insist that a standard adjustment formula is enough, as according to them, it serves the purpose. Well it can be understood why some might feel or think that way. As it applies to majority of the things and events, significant number of people are risk averse and resist change. Especially if the expected changes require them to learn/do something new and different. It is easier to continue to follow what we are used to do, as it does not tax us. Associated with this is the possible implicit assumption that, what has been followed for decades might be correct, and/or what might be followed by everybody must be correct. On the other hand, there might be fear that what might happen if one alone adopts to the new changes.

All these fears and apprehensions are largely understandable. Hence, the following justification is provided (*in addition to those provided in the section IIa to IId*) to adopt to the proposed adjusting method.

- Following just one standard-whole-sample mean IQ and SD for the adjusting might yield incorrect results and interpretation of one's intelligence.
- As earlier, adopting to the better adjusting method does not change the number of calculations. Earlier there was one formula and the better method also involve one formula.

- The earlier adjusting formula/method and the proposed adjusting formula/method follow the exact same principle and procedures. The only change in the new proposed system will be substituting different values according to the age group of the child/person.
- Therefore, the effort required will be the same.
- Separate mean IO and SD are clearly listed in BKT standardization manual. At the time when BKT was standardized, the concept of how different SDs can affect the interpretation, perception and communication of intelligence (among common public, school psychologists and fellow professionals), was not understood clearly. Actually, the adjusting formula that is being currently used was not even the product of BKT.
- The tests that yield deviation IQ (Wechsler's scales general and current Stanford Binet scales) and/or percentile points (Raven's progressive Matrices) almost always compare the child/person's score among their own age group. Therefore, at least to follow similar procedure, BKT can use specific age- related mean IQ and SD for adjusting.
- The proposed age-related adjusting method (figure 7), takes into account the developmental changes and/or growth spurts that occurs during adolescence.

The above are the sufficient reasons to adopt to the separate age specific formula/method to adjust the BKT IQ to match the current global standards.

### 3. VERBAL Vs NONVERBAL ITEMS and SPECIFIC LEARNING DISABILITY ASSESSMENT

It is a common knowledge that BKT predominantly has verbal items compared to nonverbal items. Similarly, psychologists who have done intelligence assessments will know that children present with various abilities and in different degrees of that abilities/disabilities. Some children are good in verbal abilities, some in nonverbal abilities and some excel in both.

SB 1916 version had only about 16% of non-verbal items. This percentage increased only marginally in the later versions of Stanford Binet, where 1937 version form L and form M had about 26% and 24% of nonverbal items respectively [11]. Depending on how the 'nonverbal' is defined, BKT has only about 20% items which can be considered as nonverbal. This is far less compared to other popular intelligence tests such as Wechsler's scales, where half the items are nonverbal.

Apart from BKT being standardized only on few Indian vernacular languages (Marathi, Kannada and Gujarati), and that it cannot be used when the child's verbal ability is compromised (for eg., speech delay, language incompatibility and extreme shyness/inhibition), it has one major limitation. This limitation of BKT is that it cannot be used to rule out 'subnormal intelligence' as it is required, in a child who is suspected of having Specific Learning Disability (SLD).

Children with SLD are known to have problems with spelling, reading, vocabulary, comprehension, and arithmetic. BKT has high number of test items that involves these functions. Therefore, if BKT is used to rule out subnormal intelligence, then the test might yield biased results, which are far lower than the actual ability of the child. That is, when a child has a disability in a particular function, say, reading, then this particular function (reading) cannot be used to assess the child's intelligence. If in case, if it (reading) is used, then the results might show that the child has lower intellectual ability compared to his/her age group.

An analogy can explain this better. To test a person's athletic ability swimming can be considered as a good test. However, for a person with one hand, swimming ability is not a correct test to test his/her athletic ability. If still swimming is used to test her/his athletic ability, the results will be biased, where the person will get less score. Instead of swimming, one can use tests that does not involve active use of hands, such as running, walking, or jumping. Similarly, as children (or suspected of) having SLD can have problems in spelling, reading, comprehension, vocabulary and mental calculations, it is not advisable to use BKT.

### 4. APPROPRIATENESS OF SOME ITEMS

Lot of time has elapsed since the standardization of BKT and due to which lot of things have changed in terms of usage of things, terminologies and the perception and viewpoints of people and the society.

The most commonly agreed upon items that automatically require replacement are the ones that deal with coins and counting paise. The test requires different denomination of coins that are less than a rupee. In the current circumstances, hardly children see coins of different denominations that are less than a rupee. Therefore, this can be substituted with higher denominations coins that are in circulation or paper currency of lower denominations if coins are not available.

# Aesthetic comparison

This item asks the child to "identify the face that is good looking". There are 3 pairs of faces, and the child is expected to pick one face each that is "good looking". Though this item is simple to administer and score, and appears to have face validity, it presents with ethical dilemma regarding the concept of 'good-looks'.

The issue is not about, whether the item assesses what it needs to assess and that whether it has enough discriminatory power to differentiate people on their intelligence. The issue is not about, whether the beauty is in the eyes of the beholder (i.e. beauty is subjective and varies from person to person) or there is such a thing as universal beauty (i.e. objective view that somethings are beautiful compared to others). The issue is about whether such kind of items are appropriate to ask a child of 5 years. Associated with it is the dilemma that, is this satisfies the criteria of social appropriateness and/or is it ethical. The answer to this question is a simple 'no'. A psychologist has to always be aware, understand and respect the sociopolitical and ethical/moral expectations. Given this, a psychologist inadvertently cannot encourage/propagate any idea/belief that some faces are good looking and some faces are not good looking. Given the above, the opinion of this author is that this item can be omitted and instead an alternative item can be used.

# Pictures for description

There are four pictures that are shown to the child/person in order for them to describe what might be happening in the pictures. Out of the four pictures, the first picture is meant to depict a scene of people waiting in a railway station. However, the picturization of the scene resembles several decades old railway station, which many children in present generation might not have seen/know. Given this, many children (based on the personnel experiences of performing BKT assessment on hundreds of children) fail to register/report it as a railway station. Therefore, if the children do not explicitly report it as a railway station, the examiner has to consider the performance of the child on other picture-cards and take a balanced opinion.

# Ball and the field

The manual provides the examples of several search patterns for scoring at two age levels. Figure 9a shows one of the best search patterns that gets the correct score, compared to the search pattern that is shown in figure 9b. That is, if a child/person is expected to actually search for the ball in the field, no child will go in a continuous circular pattern as shown in Figure 9a. They will usually follow search patterns that resembles figure 9b. However, if in case any child/person draws the pattern similar to pattern 9b, it can be considered for the correct full score.

# Items with gruesome details

There are at least two items that have gruesome / violent details. One is a finding a girl's body with 18 pieces and another is finding a body being hanged on a tree. Many psychologists find it difficult in asking these questions, especially to children. Therefore, if in case it is difficult to administer these specific items, it has to be balanced in the scoring.

#### Patel and Kulkarni

One of the items, expect the person to tell the differences between Patel and Kulkarni. However, the current generation of kids have hardly heard/know the words 'Patel' and 'Kulkarni', let alone knowing the differences between them. Therefore, instead of asking this, the examiner can ask about the differences between 'Manager and an Accountant'

Figure 9: Showing two different ball search patterns

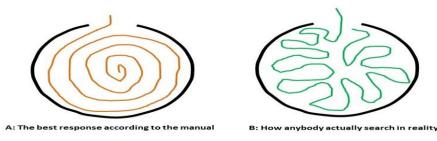


Figure 9: Ball search patterns

#### 5. THE USE OF ALTERNATIVES

There are some confusions regarding when to use alternative items. Some professionals use alternative items as another item, i.e. when a person fails in any of the 6 standard items of that particular year, they automatically use an alternative item, and if the person passes that (alternative) item then s/he is given the respective credit for that item.

Terman [5] and the Kamat [2-3] clearly describe when to use an alternative item. They state that the alternative items are substitutes only to be used on certain occasions only when regular items cannot be used due to reasons, such as, when

- regular item is spoiled/disfigured (eg. broken puzzle box blocks)
- required material is not available during testing (eg. missing picture description cards)
- the subject is already familiar with some tests

Only in exceptional cases, the above authors permit the use of alternative tests to be substituted automatically. These conditions are

- Reading test could not be given when a child has not gone to school and/or when the child has SLD
- English vocabulary test could not be given when English is not the mother tongue of the child or the child is not well versed with English

An examiner should remember that, according to the manuals, alternative items have inferior values (i.e. do not have the same discriminatory power as that of the regular items in that particular age) and that some alternative items assess similar function as that of regular items in the same year, making them redundant and less important. Hence, alternative items should be used with caution.

### 6. THE MAXIMUM CHRONOLOGICAL AGE TO BE CONSIDERED FOR CALCULATIONS

The problem of what chronological age (CA) to be considered for IQ calculations is unique to the Indian subcontinent and mainly restricted to two of the most commonly used intellectual ability tests, such as BKT and Vineland Social Maturity Scale (VSMS; Doll, 1935). This is due to the fact that only these two tests still use antiquated ratio IQ formula of MA / CA x 100 (SA/CA x 100 for VSMS).

Before shifting away from the ratio IQ method, Terman's 1916 version of Stanford Binet had 16 years as the maximum CA, and his 1937 version had 15 years as the maximum CA. According to them this was the reason because "a MA of 15 years represents the norm for all subjects who are 16 years of age or older" and in this regard they further state that "beyond 15 the mental ages are entirely artificial and are to be thought of simply numerical scores" [12]. However, this again appears to be a guess, because their test standardization sample did not involve anybody older than 18 years of age when they revised the Binet scales [4].

BKT uses 16 years as the maximum CA, because it was standardized on Terman's 1916 version. The very question of what should be the maximum CA arises due to the fact that the test items are there till age XXII (i.e. MA can go up to 22 years). Further, research suggest that the intelligence can increase even after 16 years [4]. It is inherent in the above statement of Terman and Merrill, that the MA/CA x 100 formula might not work for efficiently beyond the age of adolescence (i.e. for adults).

Given this, as long as one uses BKT with MA/CA x 100 formula there is no escaping using 16 years. However, as the intelligence is more complex than the simple MA/CA formula, the psychologist has to be careful and take the above factors into consideration while interpreting BKT results of an adult.

#### 7. PROFILE ANALYSIS in BKT: To Perform or Not to Perform?

Though not part of the original scoring and interpretation of BKT, many professionals use the following table for profile analysis (refer table 3). This has become a practice due to the fact that some of the premier institutions (in India) teach their students to carry out this type of analysis. The origin of the classification of items into particular categories as shown in table 3 is not clear. However, looking at the terminologies uses, such as 'social intelligence', and 'non-meaningful memory', it appears that the classification was done at least few decades back.

Table 3: Showing the profile analysis that is followed by some of the institutions/professionals

22	6, A1	3	2, 5		1, 4				
19	A1	5		3	6	1	2	4	
16	A1		4, 6	1, 3	2	5			
14	2, A1			5	6		1, 3, 4		
12	3, A1		4, A2	5	2	1			6, A3
10	5, A1	6	4		1		2		3, A2
9	4, 6, A1	5	1	3			2		
8	4	6	2		A2		1	A1	3, 5
7		1	3	6			5	2	4, A1
6	A1		1		3, 6		4		2, 5
5	2	3					5	6	4
4		4	1	2, A1			5	6	3
3	1, 2	5	3	6					4, A1, A2
Age	Lang	M	NM	СТ	NV	V	Num	Vis-spa	Soc-int
		me	mory		Reasoning				
*I and I and an M - Manningful NM - Non-manningful CT - Consentual Thinking									

\*Lang = Language, M = Meaningful, NM = Non-meaningful, CT = Conceptual Thinking, NV = Nonverbal, V = Verbal,

Num = Numerical, Vis-spa = Visuospatial, and Soc-int = Social intelligence

Many young professionals perform such profile analysis due to several factors, such as:

- They are taught to do it, and hence many continue to do it.
- Many professionals use it. This works as, when many known, senior professionals are doing it, it might be correct.
- As BKT gives mainly mental age and an IQ, it might appear to some that there is not much information to tell the client and/or to write in the report. Therefore, there might be a perception/belief that adding profile analysis might make the report seem voluminous and/or seem more scientific and/ore important.
- They might be influenced by Wechsler's scales, as the latter gives different sub-test level scores/interpretation depending on the version of the tests (such as, verbal-performance or different indexes). However, the main difference here is that the Wechsler's tests follow point scale and hence, it is possible and acceptable to provide sub-test level scores/interpretation, as compared to Binet scales which mainly follow age scale with unequal distribution of items across age levels.

- Apart from the above, one of the important reasons, why profile analysis is being carried out, might be that the belief that profile analysis can recognize the specific abilities as well as specific limitations of a particular child; so that, special inputs/training could be given to improve those functions.

Irrespective of the apparent benefits it seems to offer, profile analysis should not be carried out, due to the reasons listed below:

- The categories / factors that was used currently with respect to table 3, does not exactly match some of the items. For example, both digit span - forward and backward has been classified as 'nonmeaningful memory'. However, current neuropsychological literature suggests that digit span forward mainly taps 'attentional component', whereas digit span – backward mainly taps 'working memory component'. Similarly, several items that taps 'personal and/or general knowledge' has been categorized as 'social intelligence', and so on.
- Several items are wrongly been categorized into different functions without any basis. Such as item 5 (table 3) in year VII, is classified under 'numerical reasoning', but it actually more suitable under 'nonmeaningful memory'. Similarly, in year X, item 2 is under 'numerical reasoning', and item 4 is under 'nonmeaningful memory, but both (items 2 and 4) should be under 'meaningful memory'. Several more such items are wrongly categorized into other functions.
- Some of the items are not categorized into any functions at all. For example, item 1, Alternate items 1, 2 and 3 in year V is not categorized into any function.
- Many items assess more than one function, however, in the above table, items are classified under only of the functions. For example, 'Reading and reporting' is classified under 'language'. However, 'reading' involves more than the language per se.
- Another important reason is that the items in the profile are not evenly distributed across the age levels. There are three types of uneven distribution in this regard.
  - First, there are several gaps where items that test particular function does not exist in particular years. For example, there are no items to assess meaningful memory in year XII, XIV, AND XVI; there are no items to assess visuospatial reasoning in year IX to XVI (green color boxes in table 3).
  - Second, in some categories, the items start only after some age. For example, items to assess nonverbal reasoning starts only at age VI, and items on verbal reasoning start only at age XII (blue color boxes in table 3).
  - o Three, there are lots of random gaps in items distribution in the whole test itself (grey color boxes in table 3).
- Contrary to not having items in certain functions in a particular age level, in some age levels there are more than 1 item to assess a specific function. For example, there are more items to assess in socialintelligence in age III, and language in age IX. Though not a serious limitation, this does pose confusion in arriving at a decision, especially when the child pass in one or two items and fails in one or two (within the same function/ability).
- Another major limitation, is the fact that, not all functions/specific abilities are assessed equally. For example, it was found that the SB-1916 version (of which BKT is based on), has uneven distribution of items across the specific abilities/functions. That is, about 38% of the items assess knowledge, 17% assess fluid reasoning, 11% assess visuospatial processing, 8% assess quantitative reasoning, 9% working memory, 11% short-term memory and 6% other functions [11-13]. Given this uneven distribution of the functions assessed, it is not appropriate to perform profile analysis.
- Similarly, another major limitation is that, the profile analysis incorrectly reports the abilities and limitations of the person. This limitation is inherent to the BKT itself. This is the 'ceiling age' of the child. Ceiling age refers to the age level at which the child fails in all the items, and at which point the examiner stops the assessment. Given this, if profile analysis is carried out, then it is understood that the abilities below this ceiling age will only be reported. However, as already mentioned above, as the items are not evenly distributed, it may so happen that sometimes a child can have a particular ability, but which did not get tested in that ceiling age level. This might do injustice to the child. An example

might explain this phenomenon better. It can be seen in table 3, that in the function 'meaningful memory' there are no items at age VI. If a child passes the 'meaningful memory' item in age V, then fails in all the items of age VI. Given this, the examiner stops the testing. However, it might so happen that the child might actually have good meaningful memory, and if in case, age VII 'meaningful memory' item is given, there might be a chance that the child can pass. So here the child's ability in 'meaningful memory' is actually gets under-reported (due to stopping of the test / ceiling age).

Given all the above limitations, it is strongly advised not to carry out profile analysis.

# 8. WHY BKT STILL SEEMS RELEVANT EVEN AFTER SEVERAL DECADES SINCE ITS STANDARDIZATION?

Whoever familiar in using BKT on a relatively regular basis for their clinical practice, will observe that often times the results obtained in BKT seem to match the child's current level of functioning.

One of the main reasons for this might be that, some of the developmental functions of children as assessed by the test items (eg. digit span, copying diamond) are universal and stand the test of time. These functions can hardly improve to the level expected by the Flynn effect (3 points increase in IQ per decade). Further, the items of year 3, 4, and for some items in year 5, the pass percentage of the standardization sample was almost 100% or upwards of 80% in Kamat 1934 version [3]. This actually overestimates the IQs in lower ages.

In this regard, an important study was carried out in 1999 with a large sample of 759 (random sampling) in an Indian city, to reappraise the appropriateness of the BKT 1934 version [14]. The study found some of the items need to rearranged, i.e. items required downward alignment as they were easy. Interestingly, the study found high similarities in terms of mean and variance with the 1934 version. Based on the results, the author emphasized that BKT is still relevant in the current times. However, there were few confusions/issues with the article. The study used Seguin Form Board (SFB) to assess the concurrent validity, as the use of SFB is debatable with respect to its being a good test of intelligence. Another issue, is that, there is no clear description about how chronological age (CA) was recorded. The article mentions that it was taken/recorded at one particular point/date. However, it is clearly known that even few months of CA matter in the calculation of IQ.

Surprisingly, a small study (between the age groups of 6 to 8 years) conducted by the same author in 2019 [15]. The authors found that, compared to BKT 1934, the mean IQs of the children in their small study had 12 IQ points more. However, the increase in the IQ in this study can be attributed to the sample, which was taken from upper socioeconomic families from private English medium schools in a metropolitan region (Bangalore, India). This actually supports what has been given as an explanation below.

Coming back to the argument of the relevance of BKT in current times, the question is not whether BKT or Stanford Binet are 'valid' tests. There is no doubt about the validity of BKT test. The question is that, how it still seems to have validity even after about 8 decades since its standardization.

This question becomes relevant especially when one considers the Flynn effect. According to the Flynn effect, the IQ scores of the masses increases by about 3 per decade [16]. Given this, technically whatever the score a child gets in BKT (after adjusting IQ) will be an inflated score. For example, just for a very short time and just for the sake of argument, let us consider that a child Master Moah gets an IQ of 80 in BKT (after adjusting). Therefore, as 8 decades have passed, and as each decade 3 IQ points should have increased, the actual IQ of Master Moah would be somewhat equivalent to an IQ of 56 (8x3 = 24; 80 24 = 56). However, it is not that simple and one cannot do such reduction in IQ, based on Flynn effect.

The reasons for BKT results still appear to have validity even after several decades are multifactorial. That is, it cannot be attributed to any one single cause, but to many factors. The following discussion is an attempt to analyze why it might be so.

# Socioeconomic conditions

India is considered as the dual economy, where it is often quoted that, it is the country where rockets, jet planes and bullock carts coexist. This clearly highlights the socioeconomic differences among the

population. Further, India took more time to catch up, and in many areas it is still trying to catch up with Western countries (where the 3 IQ points increase has seen) with respect to improvement in nutrition, accessibility to better education and child based parenting practices, that are hypothesized to be responsible for increase in IQ (i.e. Flynn effect). Given these, one can expect that Flynn effect would have been different across the population with respect to different socioeconomic conditions, i.e. IQ would have increased for some category of population, remained same, or marginally increased (less than 3 IQ points per decade) for some category of the population. Hence, for the latter category of the population, when administered BKT, the obtained (adjusted) IQ might appear to represent their current intellectual functioning.

### Standardization sample

The period i.e. 1930s during which the BKT was standardized, was completely different with respect to the social class/caste-education composition. This article is not about socioeconomic-educational issues of that era. However, for the purpose of the evaluation of a test in terms of its appropriateness, one has to look at the standardization sample. According to the BKT manual, the standardization sample consisted of 1074 subjects, and was categorized into 5 groups [3]. These are 'socioeconomically-advanced Hindus (61%)', 'socioeconomically-intermediate Hindus (31%)', 'socioeconomically-backward Hindus (1.5%)', Muslims (3.5%)', and Christians (3%)'. The manual mentions that socioeconomically-backward Hindus and socioeconomically-forward Hindus constituted extreme ends in terms of the performance, respectively. This can be understood in terms of prevailing socioeconomic conditions of that time. If one keeps aside the issue of whether socioeconomic conditions affected performance, then one can observe few things

- The standardization sample was not equally distributed with respect to prevailing sociodemographic characteristics, at least with respect to various groups within Hindus.
- Each group representation might be appropriate with respect to some criteria at that time, but it does not represent the sociodemographic characteristics of the current population.
- Despite the differences, the standardization sample's mean relatively matched the mean of the Terman's 1916 version of Stanford-Binet.
- The extreme differences in the socioeconomic conditions of the sample, might have contributed to the wide standard deviation (SD = 18.7 in BKT compared to SD of 13 in SB-1916 version).

Given these, if we consider the largest chunk of 61% from a group with advanced socioeconomic conditions (of BKT sample), it can be expected that the normative scores obtained were already high compared to the prevailing circumstances of that time. Therefore, this high score would have masked any improvements in the growth of intelligence in later decades (i.e. Flynn effect). An analogy 'sampling effects' given in the box below might help to understand this better.

# Sampling effects: An analogy

Let us assume that the population of the country 'Barth' consist of 3 groups. The group 'Tall' consists of 100 people with each having height of 7 feet, group 'Medium' consists of 600 people with each having height of 5 feet, and group 'Short' consists of 300 with each having height of 3 feet. Given this, the average height (mean) of the entire population will be 4.6 feet.

A researcher by name 'Vitta' realizes that there is no height chart to measure the height of the people of Barth. He decides to develop a standard chart and for this, he takes about 60 people from group Tall, 35 from group Medium and 5 from group Short for her standardization sample (Notice the sample selection errors with respect to the actual numbers in the population above). Given this, the sample's average would be 6.1 feet. He uses this yardstick to compare the height of the people and based on a value of SD (exact value not relevant to this example at this point) he says height between 5.5 to 6.7 feet would be considered as normal. Let us again assume, that after few decades, another researcher by name 'Nylf' notices that, over the years

the height of the population is increasing by an average of 0.3 feet per decade depending on the nutrition

and exercise. Therefore, after about 8 decades, the Barth people's height has increased to an average of 7 feet (i.e. 8 decades x 0.3 feet = 2.4 feet; add this to the mean height of Barth, which is 4.6 feet, i.e. 4.6 feet + 2.4 feet = 7

Again, if we assume that, a person by name 'Rukma' grows only till the height of 5.8 feet (Even though the population's average height has grown to 7 feet, there will be usual variation on both sides of the NPC). It can be clearly seen that the Rukma's height of 5.8 feet is clearly lesser than the population average of 7 feet.

However, despite knowing that the height of 5.8 feet as less, we cannot certify that it is lesser unless we compare with any yardstick. According to our assumption, we have only one yardstick, which was developed by 'Vitta'. According to Vitta's yardstick, Rukma's height of 5.8 feet falls within the 'normal range' of 5.7 to 6.7 feet.

Therefore, realistically speaking, despite Rukma's height being far lesser than the current population average, we need to erroneously categorize/certify his height as 'normal', because we used 8 decades old Vitta's yardstick.

#### Wide Standard Deviation

Even though the mean of BKT was almost similar to the SB-1916 version, the standard deviation of BKT i.e. 18.7 was almost 1 ½ times that of SB-1916 version i.e. 13. Given this wide SD there are more chances of people falling into a particular category of IQ classification. For example, according to the SB-1916 mean (100) and SD (13), technically (i.e. strictly going only by the NPC characteristics) speaking the normal IQ range would be the IQ 91.5 to 108.5 (i.e. SD of 0.66 on both side of the mean; similarly, if the mean is 100 and SD is 15, the normal range IQ would be the 0.66 on both side of the mean which will be about 90 to 109 IQ points). However, going by the same NPC characteristics the normal IQ range for BKT (refer table 1) would be about 87 to 112 (It should be noted that, even though according to the NPC the percentage of population that falls within the normal IQ range would be 50% irrespective of the SD. BKT scores might not exactly follow this due to it having age scales and ratio IQ. This is explained in the following sections).

# Age scale and ratio IQ

One of the interesting things about age scale (without using ratio IQ) is that, the item placement to particular year can be used as a raw score to derive a standard score that can be matched with the normative data (similar to how it is done with respect to deviation IQ). However, the combination of age scale with ratio IQ do not allow such calculations/comparisons. As mentioned above in earlier sections, ratio IQ has inherent limitations. However, given the wide differences in socioeconomic- educational conditions in the population, wide SD of the normative sample, and the 42% average passing criteria for item placement (explained below) the use of ratio IQ does appear to match the child/person's IQ with relative accuracy (as explained in the 'sampling effects' analogy above). That is because, even after so many decades after standardization, all a child has to do is pass particular items and it will get an IQ based only on his/her age. Hence, here actually the child is **not accurately/appropriately compared** with his/her age level peers as done with deviation IQ. The comparison is just that the child could pass an item or not based on decades old item placement to particular age levels.

#### Average pass criteria of 42% for item placement

As mentioned in the above sections, Kamat placed items to particular age levels, based on the average minimum of 42% pass. Theoretically/technically speaking, this can be attributed to the wider standard deviation obtained for BKT (compared to SB-1916), and also this can contribute to children obtaining slightly higher IQs (compared to SB-1916). For example, Terman used 65% pass criteria (i.e. 65% of the people should pass that item) to place an item for a particular age level. However, Kamat's placement averaged only about 42% pass criteria (i.e. only 42% of the people is required to pass) to place an item for a particular age level. Given this, it becomes easy to pass a particular item in BKT, because the item required less stringent criteria to be placed there, and as it is easy to pass an item, the scores obtained (IQ) seem to match the current levels even after decades of growth (as mentioned in 'sampling effects' analogy). On the whole, all the above said reasons might be contributing collectively, to the feeling that BKT (adjusted) IQ matches to that of the child's actual ability.

#### **CONCLUSIONS**

Though it appears that it so many limitations, BKT is still one of the good tests of intelligence that can be used in routine day-today practice as it still has several advantages. The advantages are that it can be easily administered, takes relatively shorter time compared to Wechsler scales, involves simple scoring system, keeps the interest and motivation of the child/subject as varied items are presented that taps different modalities / functions, the concept of mental age appeals to many parents in the Indian setting, the cost of procuring BKT is very economical compared to other established tests, and many test items if lost, can be easily replaced.

This article has discussed some of the issues and concerns with BKT and try to recommend few suggestions on what can be done to overcome such issues and concerns. The examiner should keep all the above points discussed in this article while administering, scoring and interpreting to minimize the mistakes at all these stages. Even though it is just a number, IQs have enormous significance for any child/person, its future and their family. This can range from admission to schools, school performance, career, training and rehabilitation, disability benefits as well as legal implications.

It was observed that some of the professionals adjust the BKT IQ to that of Wechsler's/WHO standards, but white interpreting, erroneously still use BKT IQ classification. This should not be done. Once the IQ is adjusted, only WHO standards of IQ classification into categories (i.e. average, below/above average, and so on). It is strongly advised that a psychologist need to administer other intellectual ability tests in addition to BKT, as no one intelligence tests will yield comprehensive information about the abilities of the child/person. In addition, a test to assess social and adaptive functions should always accompany these tests. It is always good to consider all the results to arrive at a conclusion that benefits the child/person.

A psychologist should remember that while performing a BKT with a child who is intellectually compromised, that they are first and foremost a clinician. Given this, the administration, scoring and interpretation should be based on clinical judgement, rather than psychometrician approach. Any confusion at which ever stage with respect to any aspect of the BKT, the child/person should get the benefit of doubt. That is, the child/person should not be penalized for any mistakes of the test administration and scoring. Further, it is highly essential that appropriate revision of the BKT and/or the recent Stanford-Binet versions, with adequate representative sample need to be carried out in the Indian subcontinent.

# **REFERENCES**

- 1. Kamat VV. A revision of the Binet scale for Indian children:(Kanarese and Marathi speaking). Br J Educ Psychol 1934;4(3):296-309.
- 2. Kamat VV. Sex differences among Indian children in the Binet Simon tests. Br J Educ Psychol 1939:9(3):251-6.
- 3. Kamat VV. Measuring intelligence of Indian children. 3rd ed. Bombay: Oxford University Press; 1967.
- 4. Kaufman AS. IQ Testing 101. USA: Springer; 2009.
- 5. Terman LM. The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale. Houghton Mifflin; 1916.
- 6. Binet A, Simon T, Simon T. The intelligence of the feeble-minded. Williams & Wilkins; 1916.
- 7. Burt C. Family size, intelligence and social class. Population Stud 1947;1(2):177-86.
- 8. Wechsler D. The Measurement of Adult Intelligence. Baltimore: Williams and Wilkins; 1939.
- 9. Wechsler D. The Wechsler-Bellevue Intelligence Scale: Form II. Manual for administering and scoring the test. New York: The psychological Corporation; 1946.
- 10. World Health Organization. International Classification of Diseases and Related Health Problems. 10<sup>th</sup> edition; 1992.
- 11. Roopesh BN, Kumble CN. Binet Kamat Test of Intelligence -- Issues with scoring and interpretation. Indian J Mental Health 2016;3:504-5.
- 12. Becker K. A. History of the Stanford-Binet intelligence scales: Content and psychometrics (Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 1). Ilinois: Riverside Publishing; 2003.

- 13. Terman LM, Merrill MA. Measuring intelligence: A guide to the administration of the new revised Stanford-Binet Tests of Intelligence. Boston: Houghton Mifflin; 1937.
- 14. Venkateshan S. Reappraisal of the Bombay-Karnatak version of Binet-Simon Intelligence Scale (1964). Indian J Clin Psychol 2002;29:72-8.
- 15. Gopalkrishnan IK, Venkateshan S. Normative congruence between 1967 and 2002 adaptations of age scale for Indian urban children. Int J Indian Psychol 2019;7:579-90.
- 16. Flynn JR. The mean IQ of Americans: Massive gains 1932 to 1978. Psychol Bull 1984;95:29–51.

\*\*\*\*\*\*\*\*\*

Acknowledgements - Nil Conflict of Interest – Nil Funding - Nil