Use of the Test Observation Checklist with the Stanford-Binet Intelligence Scales for Early Childhood, Fifth Edition (Early SB5)

Glen P. Aylward
Southern Illinois University School of Medicine
Andrew D. Carson
Riverside Publishing
Presented at the meeting of the National Association of School Psychologists,
Atlanta, GA, April 1, 2005

Behaviors that occur during testing may have a dramatic impact on estimates of a child's levels of development or cognitive abilities. This impact often is negative, producing underestimates of true levels of functioning. Non-compliance and variability in test performance are more frequent in younger children, particularly in preschoolers and those of early school age. In addition to the immediate ramifications problematic test-taking behaviors have on actual scores, there is accumulating evidence that early high rates of refusals are associated with similar behaviors at later ages (Mantynen, Poikkeus, Ahonen, Aro, & Korkman, 2001), and with lower intelligence, visual perceptual, neuropsychological or behavioral scores at 7 to 8-years of age (Bishop & Butterworth, 1979; Langkamp & Brazy, 1999; Ounstedt, Cockburn & Moar,1983; Wolcaldo & Rieger, 2000).

Test refusal, where a child either declines to respond to any items, or eventually stops responding when items become increasingly difficult occurs in 15% to 18% of preschoolers (Bishop & Butterworth, 1979; Mantynen et al., 2001; Ounstedt, et al.,1983, Wolcaldo & Reiger, 2000). Occasional refusals occur in 41% of young children. As a result, a *minority* of children in the 2 ½ to 5-year age range are truly test compliant. This behavior affects the reliability and validity of assessment, and places examiners in a quandary with respect to whether they should count refusals as failures, prorate scale scores, or consider the testing to be invalid.

Non-compliance during testing has been reported to occur in verbal production tasks (Mantynen et al., 2001), gross motor activities (Ounstedt et al, 1983), or toward the end of the testing session (Ounstedt et al., 1983). Generally, it is less likely that children will refuse tests that require no verbal output. Refusals are reported to occur more frequently in children born at biological risk, or those from lower socioeconomic households (Bishop & Butterworth, 1983; Landry, Chapieski, Fletcher & Denson, 1988, Langkamp & Brazy, 1999; Roth, Eisenbert & Sell, 1984; Wolcaldo & Reiger, 2000). Children who refuse any aspect of testing differ from those who refuse some items, or those who are compliant and cooperative to a certain point, and then refuse more difficult items. While intuitively it makes sense to score refusals in the latter situation as failures, a decision regarding how to approach missing scores in the former situation is less clear.

Therefore, a child's test behaviors can be a rich source of clinical information. Qualitative aspects of behaviors observed during testing should be considered by the examiner when formulating impressions about a child's developmental or cognitive status, and when describing the nature of deficits. Behaviors observed during test administration might provide early indicators of subsequent, high frequency/low severity dysfunctions (learning disabilities, borderline cognitive function, neuropsychological deficits, behavioral problems, attention deficit hyperactivity disorders) (Aylward, 2000).

Moreover, early on these dysfunctions may be first manifest as subtle weaknesses or specific problematic behaviors, versus more obvious, clear-cut deficits.

To facilitate documentation of relevant behaviors, the Stanford-Binet Intelligence Scales for Early Childhood, Fifth Edition (Roid, 2005) (Early SB5) contains the *Test Observation Checklist* (TOC), designed to facilitate description of behaviors that have been documented to affect test performance. On the TOC, test-taking behaviors are grouped into two areas: 1) Characteristics, and 2) Specific behaviors. The former are considered more intrinsic and refer to child characteristics that most likely extend across situations. The latter are related to test taking behaviors that are more specific to the period of observation (although they also may reflect more persistent behavioral tendencies). Both will have an impact on the young child's test performance. While some behaviors will facilitate test-taking, the emphasis here is on those that potentially have a negative impact, or prognostic implications. Therefore, items denoted with an asterisk are considered extremes of behavior and should be used by examiners as "red flags." Underscoring test behaviors promotes identification of influences on test scores that otherwise might not be apparent had only quantitative test data been used. Items are not necessarily mutually exclusive, and examiners are encouraged to record additional descriptors in the areas allotted for such. Many of the behaviors are on a continuum, while others are categorical.

Characteristics

1. Motor Skills

Observation of motor skills bridges the realms of developmental and cognitive assessment. However, the interrelatedness of these realms is great in early childhood, and some may argue that it is impossible to separate the two concepts.

Gross motor. Gross motor skills are an important area of function in early childhood, as motor abilities often provide an early window to central nervous system function (Aylward, 1997). Neurologic soft signs (motor, sensory, or integrative functions--not localized brain dysfunction) are often associated with increased risk for compromised intellectual abilities, or with learning problems (particularly reading) in children with normal IQs (Breslau, Chicoat, & Johnson, 2000). Motor abnormalities tend to persist during later childhood (Fouder-Hughes & Cooke, 2003). Gross motor abilities also include visually guided ballistic arm movements (fast movements from one limb posture to another—e.g., bouncing, catching, throwing objects). Therefore, while observation of gross motor skills is not as critical in the assessment of older children, it is important when working with those of early childhood age. Examiners should particularly note the presence of *poor/clumsy gross motor abilities*.

Fine Motor. Fine motor function involves multiple components: visual motor control, visual perception, paper and pencil coordination, visual motor integration, eye-hand coordination, fine motor dexterity/modulation, spatial organization, perceptual planning, and fine motor speed. Copying, perceptual matching, spatial processing, finger tapping, pegboard performance, visual memory, spatial organization and visual-sequential memory are also involved. Paper and pencil skills, fine motor dexterity and fine-motor speed are particularly important at this age, and incidental observation of the child's skills in this area may provide information regarding potential areas of future academic difficulty. Examiners should note the presence of *poor or awkward fine motor/visual-motor integrative abilities.*

Oral Motor. *Drooling and/or poor tongue control* may be indicative of oral-motor apraxia, mild neuromotor disabilities, or poor expressive language and articulation skills. Children with oral motor problems are at increased risk for communication disorders, which in turn increase the possibility of behavioral issues (e.g., 20% of children with oppositional-defiant disorder have a comorbid communication disorder).

Handedness. By early childhood, most children have a hand preference, although some will still be ambidextrous. Handedness is not absolute, with many children still using both hands for various tasks, but one typically is used more often. Handedness emerges from age two-years onward, and should be established by the upper end of the Early SB5 age range. A lack of dominance or established preference is a non-specific indicator, but may be associated with neurodevelopmental immaturity.

2. Activity Level

Activity level should be compared to what is typical for the child's developmental level. That is to say, activity in a 2 ½-year old may differ markedly from that of a 5-year old, simply by virtue of the child's age. *Hyperactive/excessively restless* behavior would be most disruptive to the testing process, and most likely would also be related to inattentiveness and distractibility. Activity level is best considered to be on a continuum, and scoring for this item reflects such. However, a *hypoactive/slow moving* activity type can also be problematic, particularly in terms of responsiveness to test materials and to demands in early classroom environments. The desired activity level is one that is *well-modulated.*

3. Attention/distractibility

As is the case of activity level, attention/distractibility should be gauged based on the child's developmental age. While the first two options would not interfere with test performance, the third option, "often inattentive or distractible; frequently requires redirection," may have a negative effect. Obviously, the child whose testing performance is punctuated by poor focus and an inability to maintain interest would experience difficulty in completing tasks during the session. Extremes in this realm may be indicative of a possible attention deficit hyperactivity disorder. It is helpful to differentiate whether the child simply cannot attend or maintain focus, versus whether the problem is found in an inability to *sustain* focus over time (the latter potentially being more of a developmental issue).

4. Impulsivity

Age-appropriate impulsivity (second scoring option) refers to behaviors that sometimes require intervention by the examiner to slow the child down so that all test stimuli could be processed before a response is made. This level of impulsivity should have a minimal impact on test results. However, if the child is highly impulsive, it is quite likely that problems will continue despite examiner interventions, with a negative influence on testing results. Extremes in this area may reflect inhibitory problems that could be precursors of an attention deficit hyperactivity disorder, executive dysfunction, or poor behavioral regulation.

5. Language

Articulation. Articulation refers to production of speech sounds. It is not unusual for children to display some developmental misarticulations (e.g., "r", "l"). By age two, a child should be 50% intelligible, and by age 3-years 75% intelligible. A stranger should understand the four year-old child 100% of the time (with leeway for mild developmental articulation errors). Poor articulation might be indicative of oral-motor problems or hearing deficits. The type of articulation errors (e.g., leaving off initial sounds, dropping ending consonants, high frequency errors [th, sh]) is important to document if possible.

Receptive language. In general, a young child's receptive language skills are more advanced than are expressive abilities. Problems following directions, answering questions posed by the examiner, or understanding concepts explained by the examiner during testing may be indicative of hearing problems, inattention, or verbal processing issues. If the problem is severe, difficulties may extend across verbal and non-verbal test-taking realms, because of an inability to comprehend directions. Noteworthy is the fact that receptive language problems underscored on this item are reflective of more complex functions than simple receptive vocabulary abilities.

Expressive language. Expressive language refers to the child's communicative skills, both spontaneously and "on-demand." This is distinct from articulation per se, and includes word usage, sentence complexity, and success in conveying a message. It is helpful to delineate whether problems in this area involve circumlocutions due to word retrieval problems, a paucity of verbalizations, or disrupted grammatical structures.

Specific Behaviors

Consistency in Performance

Very inconsistent behavior may reflect the child's typical performance style, or it may describe a haphazard test-taking approach. Documentation of patterns of inconsistency is often helpful, particularly in terms of what types of tasks inconsistency is most apparent. Inconsistent performance can be attributable to attention problems, low frustration threshold, impulsivity, or high activity levels. Highly inconsistent behavior detracts from the validity of test data, particularly with respect to ceiling scores.

Mood

While some negativity and withdrawal is not unusual in young children, this usually can be handled as the child becomes more relaxed and comfortable during testing. However, if the child demonstrates *predominantly negative mood with tantrums and/or crying*, the negative impact on testing is enormous. Such behavior may reflect temperamental issues, behavioral problems, oppositionality, or more serious affective issues; it will also serve to compromise the reliability and validity of the resulting test data.

Frustration Tolerance

A low frustration tolerance may reflect the child's sensitivity to known deficits, reaction to being forced to perform in areas that are more difficult, impulsivity/ lack of inhibitory capacity, aggressiveness, poor emotional regulation, or an excessive need to be successful. While some frustration is manageable, more extreme manifestations may cause some children to subsequently refuse to proceed, shut-down, or demonstrate other

forms of non-compliance. It is often helpful to identify any patterns or commonalities associated with frustration-related behaviors.

Change in Mental Set

This item refers to cognitive flexibility, often associated with executive function. Children with problems in transitioning may have difficulty moving from one test task to another, or they may continue to attempt a non-successful solution numerous times. Perseverative tendencies definitely interfere with test performance, and could provide early clues of executive dysfunction, cognitive-adaptive disabilities, or poor self-monitoring skills.

Motivation

Although most children are intrinsically motivated as they approach testing, some are not. Many young children simply do not understand the ramifications of testing, and are prone to respond to immediate urges and conditions. As a result, they find the testing activity aversive, and therefore do not put forth sufficient effort. Poor motivation could be due to causes such as illness, oppositional behavior, or passive-aggressive tendencies. The child who expresses disinterest from initiation of testing may differ from one who loses motivation when the tasks become more difficult, and this should be noted on the TOC form.

Fear of Failure

This item provides insight into areas of function that may be difficult for the young child. Reluctance to engage in specific tasks (e.g., verbal) differs from a more general, blanket refusal. The examiner determines the presence of fear of failure through inference, which necessarily requires qualitative judgment about the child's reasons for not attempting all tasks. Although inferential in nature, the examiner can seek to corroborate any hypotheses related to fear of failure through parental report. Fear of failure may be obvious when a child begins to refuse items that he or she apparently perceives as difficult, or is reluctant to initiate activity on tasks that may seem overwhelming.

Degree of Cooperativeness/Refusals

It is not unusual for young children to refuse to cooperate during testing. Once again, the number of refusals is critical, as numerous refusals will invalidate testing, or lead to underestimation of the child's abilities. While prorating may be useful in the case of infrequent refusals, it becomes problematic when the number increases. Refusals may be due to oppositional behavior, control issues, or the child's perception that the task is too difficult.

Anxiety

Excessive fearfulness and shyness will also have a major impact on the child's test-taking behaviors. Many children need a warm-up period before they feel comfortable interacting with the examiner. However, once they adjust, their agreeableness to testing and their performance improves. Some children manifest levels of anxiety that are so

high as to cause withdrawal, a reluctance to speak, need to have a parent present or a need to repeatedly check in with the parent during testing.

Need for Redirection

Some children cannot stay on task, despite numerous attempts by the examiner. This behavior differs from the child who needs occasional redirection, as most young children will require some reminders. Constant need for redirection may reflect attention problems, overactivity, task avoidance, poor goal-directedness, or executive dysfunction. If such problems are noted in a one-on-one encounter, there is the strong likelihood that they will be intensified in group settings.

Parental Behaviors

Testing of young children is unique in that parents are often present during the testing session. Parental behaviors will have a trickle-down effect on the child's own behaviors, the effects varying from positive, to neutral, to detrimental. Parental presence can offer security and reassurance for the child; conversely, some parents unwittingly reinforce problematic behavior by laughing, not setting limits, providing tacit approval for misbehavior, or indicating to the examiner that the child simply does not want to participate in testing. Some parents need to be reminded not to give verbal cues to the child, or assist the child in handling test items. Examiners need to be sensitive to the parents' concerns during testing, and be reassuring. At the same time, examiners must maintain standardized testing procedures.

Representativeness of Test Behaviors

If the parent is present for testing, it is helpful to obtain an impression of the level of performance displayed during the session in comparison to the child's "typical" behavior. The key concept is whether the current performance is below expectations or the child's usual level of ability as determined by the parent, because this may produce an underestimation of abilities.

Present Evaluation and Future Directions

At present, the TOC is a structured but nevertheless qualitative method for identifying problematic areas of examinee behavior that warrant attention from the examiner and that may affect the validity of the basic results of testing. In the seven-step system proposed by Roid (2005) for interpretation of the results of the Early SB5, it falls into step 7, qualitative analysis.

Future research involving the Test Observation Checklist should include factor analytic investigation of the TOC items, documentation of associations among items and IQ subtest and summary scores, description of relationships among TOC scores and other relevant demographic information, and longitudinal studies of behavioral consistency and prediction to later intelligence scores. In addition, the comparative utility of a brief measure such as the TOC compared to that of more extensive test observation measures should be investigated.

References

Aylward, G.P. (2002). Cognitive and neuropsychological outcomes: more than IQ scores.

Mental Retardation and Developmental Disabilities Research Reviews, 8, 234-240.

Aylward, G. P. (1997). *Infant and Early Childhood Neuropsychology*. New York: Plenum.

Bishop, D., & Butterworth, G.E. (1979). A longitudinal study using the WPPSI and WISC-R with an English sample. *British Journal of Educational Psychology, 49*, 156-168.

Breslau, N., Chilcoat, H.D., and Johnson, E.O. (2000). Neurologic soft signs and low birth weight: Their association and neuropsychiatric implications. *Biological Psychiatry, 47*, 71-79.

Foulder-Hughes, L.A., & Cooke, R.W.I. (2003). Motor, cognitive and behavioral disorders in children born very preterm. *Developmental Medicine and Child Neurology. 45*, 97-103

Landry, S. Chapieski, L., Fletcher, J., & Denson, S. (1988). Three year outcomes for low birth weight infants: Differential effects of early medical complications. *Journal of Pediatric Psychology, 13*, 317-327.

Langkamp, D.L., & Brazy, J.E. (1999). Risk for later school problems in preterm children who do not cooperate for preschool developmental testing. *Journal of Pediatrics, 135*, 756-760.

Mantynen, H., Poikkeus, A.M., Ahonen, T., Aro, T., & Korkman, M. (2001). Clinical significance of test refusal among young children. *Child Neuropsychology, 7*, 241-250.

Ounsted, M., Cockburn, J., & Moar, V.A. (1983). Developmental assessment at four years: are there any differences between children who do, or do not, cooperate? *Archives of Disease in Childhood, 58*, 286-289.

Roid, G. H. (2005). *Stanford-Binet Intelligence Scales for Early Childhood, Fifth Edition*. Itasca, IL: Riverside Publishing.

Roth, K., Eisenberg, N., & Sell, E. (1984). The relation of preterm and full-term infants temperament to test-taking behaviors and developmental status. *Infant Behavior and Development, 7*, 495-505.

Wocaldo, C., & Rieger, I. (2000). Very preterm children who do not cooperate with assessments at three years of age: Skill differences at five years. *Journal of Developmental and Behavioral Pediatrics, 21*, 107-113.