

Evaluation Report

Davis Joint Unified School District
GATE Identification Procedures

David Jelinek, Ph.D. Principal Investigator*
College of Education California State University, Sacramento
6000 J St. Sacramento, CA 95819-6079
djelinek@csus.edu (916) 278-4491

April 21, 2005

**Thank you to Dr. Li-Ling Sun and her graduate student, Jessica Martinez, for volunteering their expertise in the creation of this report. Recognition also goes to the GATE director, Deanne Quinn, and to Kathleen Whalen and Jackie Linn for their work on the evaluation committee that includes interviews of Davis parents, teachers and administrators.*

Table of Contents

Introduction	1
A Framework for Evaluation	3
Theories and Issues Related to GATE Identification	5
<i>The Nature of Giftedness and Talented</i>	5
<i>Definitions of Giftedness</i>	5
<i>Identifying Gifted Children</i>	6
Davis Identification Processes	7
<i>Overview of Davis Identification Process</i>	7
<i>Review of Literature Related to Davis Identification Process</i>	7
Davis Evaluation Questions	9
1. <i>How valid and reliable is the current testing process?</i>	9
2. <i>Is OLSAT the best measure?</i>	9
3. <i>What non-verbal measures correlate best with the OLSAT?</i>	9
4. <i>What are the strengths and weaknesses of the qualifying criteria?</i>	11
5. <i>How can equity for all students be assured?</i>	13
6. <i>Is 3rd grade the best year to assess students?</i>	16
Recommendations	18
References	19
Appendix	
A: GATE Identification Instruments	22

Introduction

The primary intent of this evaluation is to analyze the Davis GATE Universal Testing Process, review the Davis GATE identification process in light of current literature and common practice, and propose recommendations for improvements or changes. During the 2003-2004 academic year, the Davis School District began the GATE Universal Testing Process for all of its 3rd grade students. In an effort to assure that the universal testing process is addressed fairly and equitably, the School Board is seeking answers to the following questions:

1. How valid and reliable is the current testing process?
2. What are the strengths and weaknesses of the qualifying criteria?
3. Is OLSAT (*Otis Lennon School Abilities Test*) the best measure?
4. What non-verbal measures correlate best with the OLSAT?
5. How can equity for all students be assured?
6. Is 3rd grade the best year to assess students?

Background

Before addressing the research and evaluation steps taken to address these questions, it will be helpful to provide some background.

1999-2000 Academic Year

In discussions during the writing of the GATE Master Plan, the GATE Advisory Committee recommended lowering the OLSAT-7 qualifying score from a Total (T) = 95% to (T) = 92%; or Verbal (V) = 97% to (V) = 95%; or Non-Verbal (NV) = 97% to (NV) = 95%. The primary rationale for this change was to identify underrepresented students who had been qualifying primarily with risk factors.

2002-2003 Academic Year

The district conducted a pilot to universally test third grade students at Valley Oak, North Davis, and Pioneer elementary schools for GATE eligibility. Subsequently, the Board voted to expand the universal testing of third graders district-wide.

2003-2004 Academic Year

The first universal testing process occurred in November, 2003, with re-screening in March and April, 2004. Key findings were reported to Board at its May 6, 2004 meeting, and are summarized below.

- 35% (228 out of 657 tested) of the total 3rd Grade population qualified as GATE
- The number of underserved students who became GATE qualified improved significantly through re-screening. The population of African American students increased from 17% to 25%; the population of Hispanic students increased from 2.5% to 16%.

- The use of *discernible blocks* (risk factors) favorably improved qualification rates for the bilingual population

The administration recommended that:

1. The District increase its priority focus on differentiation of instruction for all classes
2. The District reassess the criteria and methods by which GATE qualification occurs.

From May, 2004 to September, 2004; staff, the GATE Advisory Committee, and regional experts in GATE education reassessed the criteria and methods for GATE qualifications. Data collected and analyzed included:

- Criteria Used by Comparable Districts for Identifying Students as Intellectually Gifted
- Comparison of Tests of Intellectual Ability
- Flow Chart of Process for Identification of 3rd Grade Students
- Comparison of 2004 third Grade GATE Identified Students and STAR Results
- Master Plan Committee Meeting of August 30, 2004 (including a consideration of alternative assessments)
- Summary of 3rd Grade students qualifying for GATE
- Qualifying Population based on OLSAT Testing Only, Re-screening Only, and Non-Standard District GATE Process
- Universal Testing Results Students to be Re-screened
- GATE Qualifications by Ethnic Group

These data were presented to the Board on September 16, 2004, with the following recommendations:

1. Change the initial OLSAT qualifying score from 92% to 95% Total and maintain a 95% on either the Verbal or Non-Verbal score;
2. Maintain the Search and Serve process;
3. Restrict the use of the private assessment results to be used only for consideration for re-screening, not for qualifying as GATE, Intellectual Ability; and
4. Continue to administer the full Universal Testing process, including the Search and Serve component, for all third grade students.

On October 21, 2004, the staff and GATE Advisory Committee presented the current assessment process and various options to the Board. The Board was asked to decide the qualifying scores for GATE identification for 3rd grade students involved with the 2004-2005 testing and screening process.

The Board requested a more in-depth consideration of the identification process; i.e., an objective study that addresses the issues raised relative to current research, plus recommendations for fair and equitable alternatives (or additions) to the existing process.

A Framework for Evaluation

Literature related to the evaluation of gifted and talented programs indicates that well-defined gifted programs consist of the following clearly articulated components: a definition, a philosophy, identification procedures and criteria, program goals and objectives, student goals and objectives, a differentiated curriculum, in-service for school personnel and parents, position descriptions of personnel, a budget, and a system for program evaluation (Callahan, 2004). Figure 1 provides a graphic model of the components of a comprehensive GATE program evaluation. However, since no single component of evaluation operates independent of the other components, it is helpful to consider a more dynamic model that depicts the interconnectedness of the program components. Figure 2 provides that model.

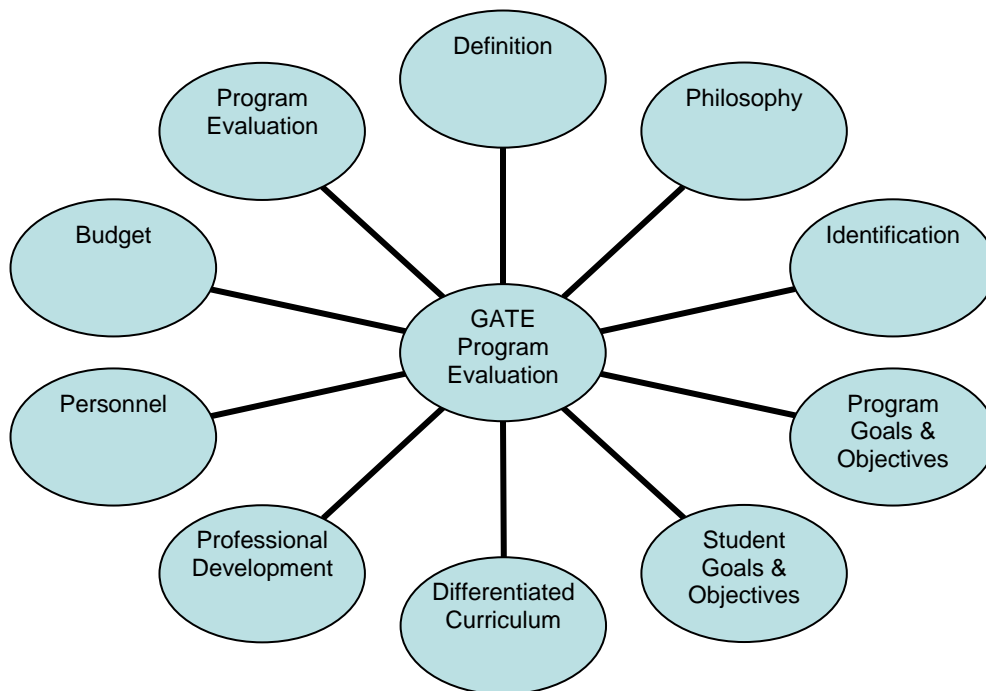


Figure 1. Model 1 of GATE Program Evaluation

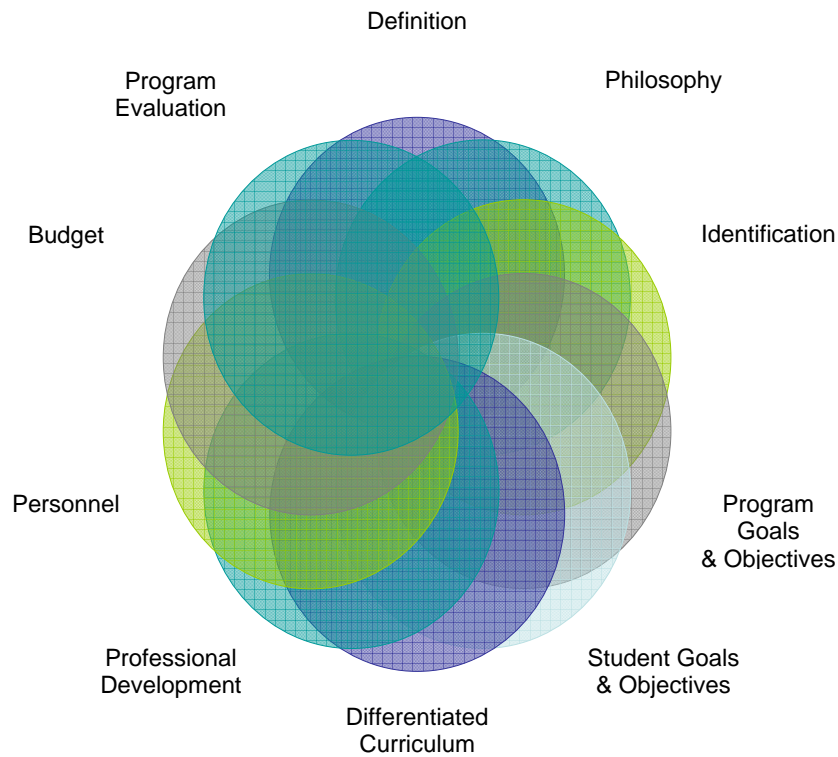


Figure 2. Model 2 of GATE Program Evaluation Representing Interconnected Components

With this interconnected characteristic of program components in mind, this report will focus primarily on *identification procedures and criteria*, drawing upon aspects of the other components when necessary to provide a well-rounded perspective.

This report is organized under four sections:

1. Theories and issues related to Gifted and Talented Identification
2. Davis Identification Processes
3. Davis Evaluation Questions
4. Recommendations

Theories and Issues Related to Gifted & Talented Identification

The Nature of Giftedness and Talented

If there is one area of agreement in Gifted and Talented literature it is that giftedness has many definitions. Consensus on which definition to use is important in terms of the Davis School District's identification process because, simply put – one needs to be able to answer *identification for what?* Current research generally agrees that giftedness involves more than just high IQ, that the definition should include non-cognitive components as well as cognitive, in a continuing search "...for those elusive things that are left over after everything explainable has been explained..." (Callahan, 2004). While intelligence tests and other cognitive ability assessments provide critical information about the dimension of a student's verbal and analytic potential, they fall short of telling us everything we need to know about *who* should be identified. "Fair identification systems use a variety of multiple assessment measures that respect diversity, accommodate students who develop at different rates, and identify potential as well as demonstrated talent. New constructs of giftedness have proven particularly important in the identification of underrepresented populations because of "...increased attention to cultural and contextual variability, use of more varied and authentic assessments, performance-based identification, and identification opportunities through rich and varied learning opportunities" (Callahan, 2004).

Definitions of Giftedness

Existing definitions of giftedness vary based on (1) the degree to which the intelligence ability is the core of the definition, (2) the way intelligence is conceptualized, and (3) the way cognitive domains of performance are combined with other non-intellective factors (Sternberg, 2004). The federal definition states that gifted students "...exhibit high performance capability in **intellectual, creative** and/ or **artistic** areas, possess an unusual **leadership** capacity or excel in specific **academic** fields." (US Department of Education, 1993). Renzulli (Renzulli, 2004) suggests there are two kinds of giftedness: "Schoolhouse giftedness (assessed by IQ tests)" and "creative producers" who possess three clusters of traits (creativity, well above average general ability and task commitment/motivation). Tannenbaum (Tannenbaum, 1983) provides a psychosocial definition characterized by three traits: (1) giftedness in children denotes their potential for becoming critically acclaimed performance or exemplary producers of ideas in various fields; producing knowledge, not consuming knowledge is the indicator of giftedness; and Giftedness is a constellation of factors beyond intellectual ability (such as environmental influences, chance/luck, social aptitude, self concept, motivation). Sternberg (Sternberg & Clinkenbeard, 1995) defines giftedness in terms of the Triarchic model of intelligence, which states that giftedness may manifest in any one of three domains: analytical, creative and practical giftedness; and that the most gifted are seen as those who are both strong in all three domains and able to balance these three domains. Developmentalist definitions view the IQ as a narrow way to identify giftedness and suggest that giftedness is domain specific and manifests itself in the process of development (Feldman, 1993).

Numerous tests of intelligence exist, emerging from the work of Binet to Terman and Wechsler. New breeds of theory-based IQ tests include the Woodcock-Johnson Psycho-ed battery (1989), the Kauffman adolescent and adult IQ tests (1993), and the Cognitive Assessment System (1997). Different tests provide varying degrees of information on how students attempt to solve problems, signs of mistaken beliefs, gaps in knowledge, selection of incorrect strategies, and cognitive deficiencies. The way in which intelligence is conceived depends in a major way on the function that

intelligence is viewed as serving (Sternberg, 1994, p 263). Definitions of giftedness that do not grapple with this question of function will lack clarity in the identification process.

The OLSAT test could provide far more information to the Davis program if item response analysis was conducted. Currently Davis receives a general report from the OLSAT publisher that provides each student's national percentile ranking in the overall test and subtests. Item response analysis, however, would help predict how gifted a student is in various domains. Further, these items could be clustered which would provide information about what types of questions are problematic for certain populations of students, which are not, and what are the implications. The means to address this process should be relatively straightforward since all 3rd grade students already take the test, thus it would be a matter of requesting a more detailed report from the test publisher. The down side is that this kind of detailed report is relatively expensive, thus it may not be feasible unless additional funding is secured.

Identifying Gifted Children

The identification of gifted children is dependent upon whether or not the definition incorporates a unitary or multidimensional perspective. The unitary model essentially views the gifted child in terms of higher than average intelligence. The multidimensional model posits that more than one characteristic should be considered when identifying gifted children; e.g., intelligence and creativity. To this end tests of creativity designed to assess divergent thinking (while IQ tests assess convergent thinking) are used in many instances of identifying giftedness. Although creativity has been embraced by the field, the use of creativity tests remains widely criticized because reliability and validity are difficult to determine. (It is interesting to note, however, that while valid and reliable measures of creativity remain dubious, increased creativity often results from well-implemented instruction for the gifted and talented.)

It is important that criteria for giftedness be clarified, as this informs which tools should be used for identification. IQ tests remain the dominant tool used in the screening and placement of students in school programs for the gifted, even though they are not considered ideal. Innovative alternative assessments are preferred by many because of their tendency to identify more underserved populations of gifted students, but they face an uphill battle because of the lack of published, standardized reliability and validity data. Furthermore, alternative assessments require a considerable investment of time and resources which may not be readily available.

Still, consideration of alternative assessments is worthy of further consideration, especially in light of the fifth evaluation question posed by the Davis School Board: *How can equity for all students be assured?* See page 18 for a review of alternative identification processes.

Davis Identification Processes

Overview of Davis Identification Process

The primary instrument Davis uses to assess students is the *Otis Lennon School Abilities Test (OLSAT)*; students in the 95th percentile for the Total Score and 95th percentile for Verbal or Non Verbal qualify. A GATE Screening Committee also considers other factors to determine if additional assessments are warranted; including: parent evaluation, teacher evaluation, most recent standardized scores, at least three representative pieces of class work. The reassessment instruments most often used are the *Wechsler Intelligence Scale for Children (WISC-IV) Verbal*, *TONI-3*, *Structure of the Intellect (SOI)*, *Slosson*, and in some cases the entire *WISC-IV*. A Search and Serve Process for underserved students includes a review of STAR data, analysis of teacher and parent comments, analysis of *Test of Nonverbal Intelligence (TONI)* results for students with lower non-verbal scores, and/or an analysis of the *WISC-IV* results for students with lower verbal scores.

Review of Literature Related to Davis Identification Process

(Note that the graphic in Appendix A provides a visual to guide the navigation through this section.)

Presently there is no literature evaluating the OLSAT-7. However, the publisher has conducted extensive validity and reliability tests of its OLSAT tests with large populations and different demographics. Thus, validity and reliability have been well established. In terms of the WISC-IV, the one article addressing the WISC-IV only provides descriptions of the new revisions. According to Burns and O’Leary (2004), the modifications to the WISC-IV include better assessment of working memory and improved scoring representing cognitive functioning. Currently, there is no literature comparing the WISC-IV with similar tests.

Despite the lack of literature evaluating the OLSAT-7 and the WISC-IV, there is research examining and comparing the OLSAT-6, WISC-III, and WISC-R. With the identification of gifted students, research indicates there are scoring inconsistencies between the OLSAT-6 and WISC-III (Beal, 1995). When testing eight and nine year olds, Beal (1995) found the WISC-III to identify students as gifted more than the OLSAT-6. Similarly, when testing third through fifth grade students referred for learning disabilities, results showed the WISC-III IQ scores to be considerably higher than the OLSAT-6 scores (Guilmette, Kennedy, & Queally, 2001). However, despite the scoring differences, Guilmette et al. (2001) found the OLSAT-6 and WISC-III to be significantly correlated. In addition, researchers have compared the OLSAT and the WISC-R by testing Hispanic students in third through sixth grade referred for placement in a gifted program (Ortiz & Volloff, 1987). Researchers found the WISC-R to identify significantly more gifted students than the OLSAT. Thus, research suggests that the WISC-III and WISC-R reveal higher scores and identify a higher number of gifted students than the OLSAT.

The Structure of Intellect (SOI) assesses creative processes, particularly creative problem solving. The SOI is based on the structure of intellect model, which suggests creativity involves divergent thinking and multiple processes (Mumford, 2001; Richards, 2001). After administering the SOI to sixth grade students, researchers concluded that the assessment reflected multiple dimensions of the students’ creative thinking (Michael & Bachelor, 1992). Additionally, researchers have examined the SOI’s retest effects. After administering the SOI to second through sixth grade students, researchers found the measure to be susceptible to significant increase in scoring after retest

(LeGagnoux, Michael, Hocevar, & Maxwell, 1990). Because of positive gain scores, researchers recommended the SOI not be used for critical decisions when previous testing has occurred. Therefore, research indicates the SOI may be a useful assessment in measuring multiple creative processes; however, test administrators should be conscientious of previous testing.

The Slosson Intelligence Test (SIT) measures children's mental ability, especially emphasizing children's verbal skill (Clarke & Scagliotti, 1989). Researchers have examined multiple versions of the SIT, discovering similar findings for each version. After testing teenage potential gifted students, Clark and Scagliotti (1989) found significant correlations between the SIT and WISC-R Verbal Scale. Similarly, significant correlations were found between the SIT-Revised and the WISC-R for children between the ages of 6 and 16 (Campbell & Ashmore, 1995). Additionally, researchers compared the SIT Full-Range and the WISC-III by testing students referred for special education (Bell, Rucker, & Finch, 2002). Researchers found the SIT Full-Range to be associated with verbal abilities, math abilities, and overall intelligence. Results also indicated that the SIT Full-Range correlated more with the WISC-III than achievement tests. Furthermore, after review, researchers supported the use of the SIT Full-Range as a screening instrument (Bell et al., 2002). Thus, the SIT appears to be significantly correlated with the WISC-R and WISC-III for children between the ages of 6 and 16. Additionally, researchers consider the SIT a useful measurement, especially for assessing verbal ability.

The TONI is a nonverbal assessment of cognitive abilities (D'Amato, Lidiak, & Lassiter, 1994). Researchers comparing the TONI with the WISC-R found varying correlations. Whorton and Morgan (1990) administered the TONI and WISC-R to Native American and Caucasian students between the ages of 5 and 16. Results indicated that the two assessments were moderately correlated. However, after testing African-American and Caucasian students of similar ages, researchers found the TONI to be a good predictor of the WISC-R (Vance, Hankins, & Reynolds, 1988). Other researchers have found contrasting correlations. D'Amato, Lidiak, and Lassiter (1994) administered the tests to students referred for classroom problems between the ages of 6 and 15. Results showed a low correlation between the TONI and WISC-R. Additionally, researchers found the correlation between the two tests to vary depending on students' ethnicity. Kowall, Watson, and Madak (1990) compared the TONI and WISC-R by testing non-indigenous children and Native Canadian children. With the non-indigenous children, there was a high correlation between the two tests. In contrast, with the Native Canadian children, a comparison of the tests showed a low correlation. Therefore, researchers have found varying correlations between the TONI and WISC-R with different populations.

Davis Evaluation Questions

Responses to the evaluation questions posed by the Davis School Board are presented below. The brevity of responses for the first two questions in no way implies less importance; rather, these questions have been addressed more fully in the previous sections of this report.

1. *How valid and reliable is the current testing process?*

The individual tests used in the Davis identification process meet the statistical criteria of validity and reliability inasmuch as the test developers have conducted the necessary measures with large populations of students and have normed these measures for various groups and subgroups. Thus, in terms of specific tests (e.g., OLSAT), the instruments are valid and reliable. The Davis question, however, asks whether the *current testing process* is valid and reliable which, statistically speaking, cannot be answered since validity and reliability imply complex statistical measures that require an extensive database well beyond that possessed by Davis. However, as the foregoing review of literature suggests, there are strengths and weaknesses in each of the tests used in the Davis identification process. An identification procedure that relies on OLSAT or any of the tests exclusively would be problematic, but the criteria and tests for the Davis search and serve process greatly strengthens the effectiveness of the testing process. This is further substantiated by data gathered by GATE staff and advisory committee members indicating a higher percentage of underserved students were identified.

2. *Is OLSAT the best measure?*

Certainly an argument can be made for OLSAT in that it measures both verbal and non verbal cognitive skills, it is supported by a large database and has been normed to subgroups similar to those Davis wishes to identify. And while there is evidence that suggests the WISC, Slosson or Structure of Intellect (SOI) might provide better assessments, one has to weigh this against the extra time and expense required to administer these tests. In light of the Davis district's universal testing policy, it would be prohibitive to administer these three tests without additional financial and personnel resources. It is neither practical nor sound to administer these tests on a large scale basis, therefore the OLSAT is the best measure.

3. *What non-verbal measures correlate best with the OLSAT?*

As the literature suggests, researchers have found contrasting correlations between the TONI and the OLSAT, or between the TONI and the other cognitive tests, for that matter. Theoretically, the TONI measures intelligence, aptitude, abstract reasoning and problem solving in a language-free medium, making it preferable to the OLSAT for EL students and many students with special needs. Interestingly, 32 of the 64 questions in the OLSAT-7 consist of non-verbal questions, and 17 of those 32 questions are identical to TONI. Currently the publisher does not provide Davis with enough specificity of test results to ascertain how well students perform on the 17 TONI questions relative to the other test questions, but this would be helpful information to request as it would allow the district to calculate correlations between TONI scores and OLSAT scores.

More specificity of test results would also serve another important function: item and cluster analysis. Currently the publisher provides scores that indicate percentile ratings which then inform the decision-making process for GATE identification. A more detailed report of which students answered

which questions or clusters of questions correctly or incorrectly, however, could serve as predictive indicators for gifted students.

Similar to the TONI, the Raven's Progressive Matrices (RPM) are nonverbal assessments measuring intelligence. There are two levels of the RPM intended for children – the Standard Progressive Matrices (SPM) and the Colored Progressive Matrices (CPM). The SPM, the average level, is designed to measure the intelligence of children and adults from 6 to 65 years of age (Sattler, 1988). The CPM, the easiest level, targets children between 5 and 11 years old, as well as the elderly and people with disabilities. The RPM contains arrangements of figural problems progressing from easy to difficult. Additionally, the CPM is printed in several colors to attract and hold the attention of examinees (Sattler, 1988).

The RPM's manual indicates that correlations between the SPM and other intelligence tests are generally high. However, reports present a wide range of correlation coefficients, suggesting inconsistencies between studies' findings. For English-speaking children, correlations between the SPM and the WISC range from .54 to .86 (Raven, Raven, and Court, 1998). Additionally, the manual reports that correlations between the SPM and achievement measures are lower than intelligence tests. Similarly, reports indicate a broad range of correlations for the CPM and other intelligence tests. According to Matthews (1988), correlations between the CPM and the WISC range between .22 and .83. Thus, reports on correlations between the RPM levels and other intelligence tests show high variability.

Concerning reliability, the manual suggests that the SPM's scores are significantly reliable across multiple cultural and socio-economic groups. In the United States, correlations between item difficulties found independently for various ethnic groups (i.e. Caucasian, African-American, Hispanic, Asian, and Navajo) ranged from .97 to 1.00, suggesting the reliability for the SPM and CPM is high.

Because of its design to decrease item dependence on acquired information and cultural content, researchers have considered the RPM to be a culturally reduced intelligence test (Jensen, 1980). Furthermore, considering gifted populations, researchers have compared the SPM with other nonverbal measures of intelligence, such as the Culture-Fair Intelligence Test and the Naglieri Nonverbal Abilities Test. Researchers administered the three tests to a sample of mostly African-American students between third and eighth grade (Stephens, Kiger, Karnes, & Whorton, 1999). Results showed that the SPM identified the largest number of students compared to the other two tests. However, the Culture-Fair Intelligence Test and the Naglieri Nonverbal Abilities Test identified students not found by the SPM. Therefore, researchers recommended that one test is not sufficient for identifying culturally diverse gifted students (Stephens et al., 1999). The Davis Search and Serve identification process addresses this by re-testing students using instruments other than OLSAT when initial scores fall below the GATE cutoff scores.

4. *What are the strengths and weaknesses of the qualifying criteria?*

The following research-based indicators of GATE identification processes will help frame the responses to this question (Callahan, 2004).

Identification criteria should be specifically related to the definition

Based on *The Nature of Giftedness and Talented* discussion earlier in this report, it is evident that the process of identification involves explicitly stating what it is the school district values and why. The GATE evaluation committee's most recent efforts to address this through parent, teacher and administrator interviews should be commended as this intention to define giftedness for the Davis School District points to an obvious strength in the identification process. The Davis GATE master plan tends to focus on cognitive components in terms of its definition and identification criteria. While this approach is certainly in line with general practice, it could be strengthened by focusing on non-cognitive components both in its definition and identification procedures.

Performance indicators need to be reliable, valid measures of the defined areas of giftedness

Research suggests that programs need to consider *excellence* independently of *rarity* –we seek out rarity because of our inability to serve all students who may truly have impressive potentials (e.g., 35% of third grade students qualifying for GATE). Norm-based measurements (e.g., national percentile rankings) equate excellence and rarity so we should think in terms of criterion-based terms to answer this question. The Davis identification process includes consideration of criterion-based indicators by using results from the California standardized tests such as the CAT-6 and standards-based tests. An example of another kind of criterion-based indicator might be identifying specific OLSAT questions that demonstrate a strong predictive quality of giftedness. Used as predictive indicators, these questions then serve as criterion-based terms.

Multiple criteria should be used

As outlined in the Davis GATE identification process, the district employs multiple criteria in its identification process, beginning before the universal testing process occurs and extending through the screening and rescreening processes. This all suggests an obvious strength. One aspect that could strengthen this process even more would be the addition of alternative assessments. The Structure of Intellect (SOI), for example, would identify creative processes and problem solving skills that would be difficult to identify under the current system. As stated earlier, however, cost and time factors make this impractical for a universal testing process. Thus, in terms of its universal testing policy, the Davis GATE program should be commended for its attention to the use of multiple criteria.

Cutoffs should be reasonable in light of relevant research and the amount of error found in each performance indicator

The district has experimented with a range of OLSAT cutoff scores – all within a reasonable interpretation of relevant research and practices by districts with similar demographics. This resulted, however, to the identification of 35% of the third grade population qualifying for GATE in 2003-2004, a challenging large number under even the best of circumstances (e.g., unlimited resources). A far smaller percentage was identified during the current academic year by raising the cutoff scores. Unfortunately, the practice of changing cutoff scores until an ideal number of students are identified

does little to address the larger questions of what defines giftedness and who will be best served by a program for the gifted and talented. This is a multi-faceted problem requiring multiple sources of data to address; i.e., the use of multiple assessments are necessary. An example might be administering the OLSAT and the SOI, or the SOI and WISC (Note that these are not actual recommended test combinations. They are merely provided as examples of multiple assessment.

Separate scores should be converted to a common scale and weighted appropriately when composite scores have been computed

Scores from dissimilar instruments should not be combined, therefore the inherent challenge in determining cutoff scores of multiple assessments lies with how to combine those scores from dissimilar instruments. Davis does not combine scores, but they do replace one score with another when the rescreening process is triggered and it is determined that another test should be administered, and its score replaces one or more OLSAT subscores. For purposes of rescreening this is not necessarily problematic, but in the event Davis elects to administer more than one assessment, it becomes important to avoid combining scores into composite scores. The recommendation is to convert the scores of dissimilar assessments into t-scores and to weight them according to the value they will hold for the overall cutoff score.

The process should allow for an appeals procedure (due processes should be followed).

The Search and Serve process articulates a well defined process of identification and placement, and the Davis GATE Master Plan provides numerous avenues to assure an equitable appeals procedure. This is a strength provided all stakeholders are *aware* of the process, but as revealed in the evaluation committee's interviews of stakeholders, few of them are currently aware. A significant challenge Davis faces is related to the *ranking* of students who meet identification criteria but are put on waiting lists because of space limitations. This challenge is exacerbated by a lack of clarity in what determines a child's rank. OLSAT scores, for example, determine some children's ranking, but when rescreening occurs and alternate test scores replace sub-test scores, this presumably changes the rank. Further, some students are tested privately which has raised questions of inequity with some parents, and expectations of assured program placement by others.

Entire process should reflect the stated program philosophy

Strengths and weaknesses in this final category come down to what the Davis GATE program defines as *giftedness*. The section on *definitions of giftedness* earlier in this report bears out how multi-faceted the definition process is, and the section on *identifying gifted children* underscores the importance of agreement on that definition before determining identification criteria. The Davis GATE Master Plan suggests that a coherent program philosophy and commitment to a *process* of continued program clarification; it is to be commended for that commitment.

In the beginning of this report it was stated that program components can not be considered independent of one another, therefore it would behoove Davis GATE stakeholders to revisit it's program in light of the ten interconnected components depicted in Figure 2. Most notably it appears that *differentiated curriculum* and *professional development* deserve strong consideration. The evaluation committee's interview results suggest that parents see little evidence of differentiated instruction occurring, while teachers cite varying examples of differentiation strategies; but it is apparent that this is more than a communication gap. A concerted effort to address differentiation for gifted students in regular classrooms appears to be a strong need.

Teachers also expressed a need for more time to collaborate, observe, practice and debrief differentiation strategies. A strength in this regard is the district's launch of a three-year professional development program focusing on differentiated instruction. To echo the evaluation committee's recommendations to the board, this is a commendable effort. Its strength or weakness, however, depends upon program delivery and teacher participation. As the administration readily admits, old paradigms of professional development (*talking heads*) don't work. The teacher interviews also bear this out – they want to collaborate, practice and debrief. Thus it will behoove the administration to incorporate innovative professional development approaches that expel the notion of “one more mandate” with a “this, too, shall pass” attitude.

A promising model of professional development practice to assure that differentiation really happens is *Lesson Study*. At its core *Lesson Study* is a collegial process to improve teaching and learning. “*Learning at its best,*” in the words of Alfie Kohn, “...is a result of sharing information and ideas, challenging someone else's interpretation and having to rethink your own, and working on problems in a climate of mutual support.” In the Lesson Study process, groups of teachers meet regularly over a period of time ranging from several months to a year to work on the design, implementation, testing, and improvement of “research lessons” (i.e., “differentiation strategies” for gifted students in the regular classroom). More information on the Lesson Study process can be found at <http://www.csus.edu/indiv/j/jelinekd>. Follow the link to *Lesson Study*.

5. *How can equity for all students be assured?*

Multiple assessments that include alternative identification processes will increase the probability that equity for a greater number of students will be obtained. The strongest argument to use alternative identification processes is to increase the participation and performance of ethnically diverse populations in programs for the gifted. Chinn & Hughes (1987) found that the reason children of low SES and of different cultures are **less likely** to be identified as gifted is because of a lack of effective assessment tools. Smith, LeRose and Clasen (1991) found that unidentified/ unserved /unplaced gifted ethnic students are more likely to drop out of school and less likely to go to college.

A major task is to identify those minority children with the **potential** for academic excellence because, due to the lack of sufficient and appropriate experiences, their demonstrated academic performance may not reflect what they might do under improved/ enriched educational circumstances (Mills & Tissot 1995). Borland & Wright (1994) used both traditional and non-traditional assessments (nonverbal) that resulted in an increased identification of underrepresented gifted students. Tyler-Wood & Carri (1993) administered a battery of nine cognitive tasks (non-verbal) to assess mild learning problems and concluded that some of these tasks can be used to effectively identify high performing students. Scott, Deuel, Jean-Francois, Urbano (1992) used the same battery of cognitive tasks to test minority kindergarteners in a “game-like” manner and found that a larger number of underrepresented children currently in regular education classes were identified as gifted.

Plucker, et al (1996) investigated Multiple Intelligence- based approaches applied to the process of identifying potential talents. Five components are included in this study—talent identification, student instruction, parent outreach, mentoring and program evaluation. The results of Plucker's study suggest that Gardner's Multiple Intelligence theory can be translated into **reliable** assessments but that creating **valid** assessments is quite difficult. Administrators tend to exhibit a bias toward linguistic and logical-mathematical intelligence which results in little construct validity outside of these two measures (linguistic and logical-math). Plucker suggested that intensive staff training is necessary and that performance assessments used for high-stakes purposes (identifying potential

talented students) need to be reliable, valid, appropriately normed and equally fair to students regardless of gender and ethnicity.

DISCOVER, the acronym for Discovering Intellectual Strengths and Capabilities through Observation while allowing for Varied Ethnic Responses, is an assessment based on Gardner's Theory of Multiple Intelligences (Sarouphim, 1999). DISCOVER is also performance-based and specifically intended to identify gifted students from diverse cultural backgrounds. The identification method involves five activities, which takes place in the classroom. For each of the activities there are different tasks that increase in complexity during the assessment process, involving both convergent and divergent thinking. Additionally, the tasks are designed differently for grade levels from kindergarten to high school (Sarouphim, 1999).

The classroom teacher administers the first two activities: a math worksheet and a writing task (Kornhaber, 1999). Trained observers oversee the three remaining activities on a separate day, which occur in small groups and take approximately three hours. During the activities, children are permitted to work with peers. The first observed activity assesses spatial ability and involves the manipulation of geometric shapes. The second activity, measuring logical-mathematical and spatial abilities, includes problems requiring the use of tangram pieces. Finally, the third observed activity involves storytelling tasks. During these three tasks, observers provide directions and observe children's problem-solving development, solutions, and interactions with peers using checklists and written notes. After the assessment in the classroom, observers meet to complete another checklist covering multiple intelligences, discussing and determining children's scores (Kornhaber, 1999).

Additionally, researchers have examined whether DISCOVER increases the identification of underrepresented students. According to Kornhaber (1999), identification of Navajo Indian children in Arizona schools has been low. After using DISCOVER, identification rates rose between 10 and 30%. After administering DISCOVER to Navajo Indians and Mexican American students in kindergarten, second, fourth and fifth grades, researchers found a high percentage of students to be identified as gifted (Sarouphim, 2001). Furthermore, research indicated that DISCOVER increased identification of Hispanic and Native American high school students (Sarouphim, 2002). Thus, DISCOVER appears to be a useful measure regarding the identification of underrepresented students.

Similar to DISCOVER, the Problem Solving Assessment (PSA) is linked to Gardner's Theory of Multiple Intelligence (Kornhaber, 1999). The PSA has a pre-assessment phase that occurs weeks before the actual assessment. The pre-assessment phase entails the classroom teacher conducting three lessons involving tasks based on linguistic, logical-mathematical, and spatial abilities. During the lessons, the classroom teacher and gifted teacher record observations of the children on a checklist. The actual assessment involves nine activities, with two activities administered by a teacher. The activities include a story writing task and the Matrix Analogies Test. Like DISCOVER, the seven other activities involve children participating in tasks with observers noting their problem solving, solutions, and interactions with others. Observers also meet later to discuss children's performances and designate scores (Kornhaber, 1999).

Researchers have also compared the PSA's identification of gifted students with traditional measures. In the public schools of Charlotte-Mecklenburg, North Carolina 40% of the school population is African American; however, traditionally only 8 to 12% of African American students have been identified as gifted (Kornhaber, 1999). After using the PSA, the percentage of African American students increased to 18%. Furthermore, researchers compared the PSA and the Matrix Analogies Test-Short Form (MAT-SF) during the identification of second grade students (Reid, Udall,

Romanoff, & Algozzine, 1999). With traditional testing, 7% of African American students have been identified. Using the MAT-SF, 8% of African American students were recommended for placement in a gifted program. However, the PSA identified 32% of the African American students as gifted. Therefore, researchers have found the PSA to increase the discovery of African American gifted students.

Like the previous described assessments, the Gifted Model Program considers the Theory of Multiple Intelligence in its identification process. The Gifted Model Program, used and developed in Montgomery County, Maryland, includes objective assessments, such as the Raven's Progressive Matrices, and subjective measures, such as the Renzulli/Hartman teacher checklist (Kornhaber, 1999). Additionally, the Gifted Model Program involves a Multiple Intelligence Checklist used by teachers in an enriched curriculum with hands-on approaches and multiple intelligence centers. The checklist includes 7 to 11 observable behaviors representing multiple intelligences. Concerning the identification of underrepresented students, the Gifted Model Program increased the identification of students after five years by approximately 50%; yet, the proportion of underrepresented students in the gifted program did not rise. However, the lack of increase with underrepresented students may be due to schools already having a proportional representation of minority students before the program's implementation (Kornhaber, 1999).

Project STAR is another performance-based assessment; however, it does not consider the Theory of Multiple Intelligence in its design (VanTassel-Baska, Johnson, & Avery, 2002). The assessment involves a series of verbal and nonverbal tasks administered by teachers to small groups of students in second through fifth grade. Project STAR also incorporates the concept of dynamic assessment by pre-teaching activities to students before the actual assessment process. The tasks in Project STAR emphasize problem solving instead of prior learning. Moreover, the tasks are advanced and open-ended, allowing for multiple answers and approaches to problems. Tangible manipulatives are also included to offer students a concrete approach. Additionally, during the process of solving problems, students are required to articulate their thinking processes. Verbal tasks include problem solving, reasoning, writing, analogies, verbal relationships, and letter puzzles. Nonverbal tasks entail such areas as arithmetic, logic, spatial reasoning, patterns, and geometry. Furthermore, regarding the identification of underrepresented students, researchers found Project STAR to increase the identification of minority and economically disadvantaged students (VanTassel-Baska et al., 2002).

An additional performance-based alternative approach is dynamic assessment. This method is a diagnostic process that considers context and students' learning abilities (Kirschenbaum, 1998). Dynamic assessment involves a test-intervene-retest format that considers students' progress in learning to solve problems. To use the assessment, instructors teach students how to perform tasks, later measuring their learning ability while solving similar problems. The goal of the assessment is to assist students in developing cognitive skills corresponding with their true intellect, not to raise IQ scores. Additionally, Bolig and Day (as cited in Kirschenbaum, 1998) offer various reasons why dynamic assessment may be beneficial for identifying gifted students, such as focusing more on learning than knowledge, determining differences in learning abilities, and seeking understanding of students' problem solving methods.

An example of a dynamic assessment is Renzulli's Triad/Revolving Door Identification Model (RDIM). The purpose of the RDIM is to offer students the chance to demonstrate gifted behaviors through creative outcomes (MacRae & Lupart, 1991). Gifted behaviors include above average general or specific abilities, elevated task commitment, and high levels of creativity. Additionally, attached to the identification process is an enrichment program that allows students to develop gifted behaviors.

Hence, the RDIM approach does not consider giftedness a stable concept, enabling students to enter and exit different levels of the enrichment program. Furthermore, according to Renzulli (as cited in MacRae & Lupart, 1991), the RDIM increases the identification of gifted students, including disadvantaged and bilingual children, because it provides multiple criteria and flexibility in its selection process.

Researchers have critiqued performance assessments for their lack of objectivity and standardization (MacRae & Lupart, 1991). Further issues include the implementation of highly time-consuming measures and the needed involvement of multiple people with training and experience. However, despite these limitations, researchers have also described multiple strengths of performance assessments, including an increase in the identification of underrepresented students. According to VanTassel-Baska et al. (2002), performance assessments are a useful identification method because their approach corresponds with gifted curricula, such as emphasis on higher-level cognition, open-ended problem solving, and articulation of thinking processes. Additionally, performance assessments stress measurement of specific types of abilities in multiple areas instead of general reasoning, which may increase underrepresented students chances of being identified. This method also offers an alternative way of examining students' abilities in contexts (VanTassel-Baska et al., 2002).

6. *Is 3rd grade the best year to assess students?*

Answering this question depends upon the district's definition and philosophy of giftedness. A definition that views giftedness primarily in terms of intelligence quotients and other cognitive abilities often argues that giftedness is best identified in the earlier grades – perhaps as young as kindergarten or before. An increasing body of evidence suggests that early identification (the earlier the better) may help in the identification of minority students with the potential for academic excellence which might not otherwise be realized due to their educational circumstances. Scott, et al (1996) conducted a study to see if minority children with a potential for high academic achievement could be identified by assessing cognitive abilities and using a sample of high IQ already identified as gifted as a criterion reference group. Results indicted that the most promising tasks are open-ended items which encourage fluency. In contrast to frequent reports of test bias, the most successful tasks are verbal, which use familiar concepts to all children (for a discussion on the use of measures based on events/concepts common to both majority and minority students, see Clasen, Middleton and Connell, 1994)

Project Synergy, led by Borland, Schnur & Wright (2000) at Columbia University, set goals to develop nontraditional ways to identify young, disadvantaged, potentially gifted students. The identification process involved gathering data on all kindergarten students through observation of behaviors in structured and free-play activities, teacher and parent referrals, draw-a-person tests, and curriculum-based assessments that yield a portfolio for each student. All data were used to identify 40% of the kindergarteners who then were placed in the candidate pool. Students in this pool then received individual assessments (Peabody Picture Test, the Test of Early Math Abilities, the Test of Early Reading Ability, dynamic assessment, a child interview, an individualized curriculum based assessment). The authors suggest that the potential for gifted exists in roughly equal proportions in all groups and that social forces are undermining its development in some groups. Three phases are needed: (1) site-specific identification focusing on potential giftedness; (2) transitional services designed to help students move from being potentially gifted to being manifestly gifted; and (3) placement of students when they are ready, in classes for gifted children. In the final analysis the authors of Project Synergy emphasize that external funding is necessary because of the extensive time commitment collecting multiple sources of data over time.

Time and finance arguments aside, the position on “ how early should we identify a child with special gifts” is still divided. The equity/developmentalist argument holds that early identification will exclude a larger percentage of students of diversity who develop at varying rates. Horowitz (1987) proposed a developmental model because, due to a variety of combinatorial circumstances, children who give no early hint of giftedness often later show extremely gifted performance. Sometimes we treat later discoveries of gifted as delayed discoveries, assuming the individual’s gifts have just been overlooked or not fostered in early years. However, if development is a continuous process of possibility, it is not unreasonable to suppose that a set of combinatorial circumstances that results in gifted can emerge at various points in the course of development.

A further developmentalist argument posits that children before the age of seven or eight have not necessarily developed the cognitive, mental or emotional maturity to exhibit characteristics of giftedness. Noted child psychologist David Elkind (2001) warns against the dangers of “pushing” children too fast to perform beyond their normal levels of development. The issue here is not that small percentage of children who will exhibit traits of giftedness at a very young age, but rather those children who will be prematurely pushed in order to qualify for a gifted program. The premise is that a large percentage of children need significantly more time to demonstrate the potential for giftedness. This is particularly true if that definition incorporates multiple indicators such as excellence, rarity, demonstrability, productivity, and creativity.

The field is also divided on the question of third grade or later as the ideal time to identify gifted students. One camp ascertains that the onset of puberty is the ideal time because, in Piagetian terms, this is the time of transition from the *Concrete Operational* to *Formal Operational* stage when experiences are understood in terms of concrete or abstract symbols that can be logically manipulated. Benbow & Minor (1990) investigated the construct of intelligence by contrasting tests of specific abilities among 13-year-olds and suggested that two distinct forms of giftedness seem to exist –verbal and nonverbal. Their evolution appeared to follow different developmental paths that may not manifest until age twelve or thirteen.

It is fair to say, though, that only a small percentage would argue for the extremes (nursery school or middle school), and the majority select second, third or fourth grade as the age of identification. This is also in line with post-Piagetian theorists who argue that that development is continuous rather than stage-like, and that children are able to exhibit characteristics of reasoning and problem-solving much earlier than age 11 or 12 (Piaget’s Formal Operational stage). Thus, Davis is well within this norm of usual practice. Furthermore, the parent and teacher interviews suggest that the majority are comfortable with identification in third grade, with the beginning of self-contained GATE classes in fourth grade. Pragmatically this also makes sense because assessments that may require small group administration prior to third grade can be administered to a whole class beginning with third grade.

Recommendations

The following recommendations are derived from this evaluation report of the Davis GATE identification process and considerations of alternative identification processes, differentiated curriculum and instruction, professional development, and program options.

- 1) Pursue external grants designed for programs employing innovative methods for identifying and serving under-served populations (e.g., Javits).
- 2) Administer a screening effectively so it does not detract from the limited time and resources of a gifted program (i.e., the identification process must match the definition of the Davis gifted program).
- 3) Address equity for all students through the administration of multiple assessments that includes alternative identification processes.
- 4) Results from the test should be used for purposes beyond identification (e.g., identification should have an impact on programming).
- 5) When using “cut off” scores from more than one instrument, the scores should be T-score and weighted; i.e., do not sum up scores from identification instruments to obtain a single composite index of “giftedness”. Scores from dissimilar tests should not be combined. The ideal way to combine scores which are similar is to standardize them by conversion to T-score and then weight as appropriate.
- 6) Once students are identified, a process of differentiation should begin in which their interest, abilities, aptitudes are examined and appropriately differentiated learning activities are prescribed.
- 7) The screening process should provide in-depth information to be used to produce programs for **both** students in a special program and those who remain in the regular class.
- 8) Consider additional program options such as magnet and/or enrichment programs, which could serve as viable alternatives for some students.
- 9) Collect and analyze data for program improvement and validation of the identification process, not solely for accountability.
- 10) Implement a teacher mentoring process such as Lesson Study for differentiated instruction in the regular classroom .
- 11) Formulate action plans in response to recommendations.

References

- Beal, A. L. (1995). A comparison of WISC-III and OLSAT-6 for the identification of gifted students. *Canadian Journal of School Psychology, 11*(2), 120-129.
- Bell, N. L., Rucker, M., & Finch, A. J. (2002). Concurrent validity of the Slosson Full-Range Intelligence Test: Comparison with the Wechsler Intelligence Scale for Children–Third Edition and the Woodcock Johnson Tests of Achievement – Revised. *Psychology in the Schools, 39*(1), 31-38.
- Benbow, C. P., & Minor, L. L. (1990). Cognitive profiles of verbally and mathematically precocious students: Implications for identification of the gifted. *Gifted Child Quarterly, 34*, 21-26.
- Borland, J.H., & Wright, L. (1994). Identifying young potentially gifted, economically disadvantaged students. *Gifted Child Quarterly, 38*(4), 164-171.
- Borland, J. H., Schurr, R., & Wright, L. (2000). Economically disadvantaged students in a school for the academically gifted: A postpositivist inquiry into individual and family adjustment. *Gifted Child Quarterly, 44* (13-32).
- Burns, T. G., & O’Leary, S. D. (2004). Wechsler Intelligence Scale for Children-IV: Test review. *Applied Neuropsychology, 11*(4), 233-236.
- Callahan, C. M. (Ed.). (2004). *Program evaluation in gifted education*. Thousand Oaks, CA: Corwin Press.
- Campbell, C. A., & Ashmore, R. J. (1995). The Slosson Intelligence Test–Revised (SIT-R). *Measurement and Evaluation in Counseling and Development, 28*(2), 116-118.
- Clarke, R., & Scagliotti, J. (1989). Are the Slosson Intelligence Test and the Wechsler Intelligence Scale for Children – Revised interchangeable for identifying gifted students? *Psychology: A Journal of Human Behavior, 26*(2-3), 33-38.
- Clasen, D. R.; Middleton, J. A.; Connell, T. J. (1994). Assessing artistic and problem-solving performance in minority and non-minority students using a nontraditional multidimensional approach. *Gifted Child Quarterly, 38* (1), 27-38.
- D’Amato, R. C., Lidiak, S. E., & Lassiter, K. S. (1994). Comparing verbal and nonverbal intellectual functioning with the TONI and WISC-R. *Perceptual and Motor Skills, 78*(2), 701-702.
- Elkind, D. (2001). *The hurried child: Growing up too fast too soon*. (3rd ed.). Boston: Perseus Publishing.
- Feldman, D. H. (1993). Child prodigies: A distinctive form of giftedness. *Gifted Child Quarterly, 37*(4), 188-193.
- Friedman, R. C., & Lee, S. W. (1996). Differentiating instruction for high-achieving/gifted children in regular classrooms: A field test of three gifted-education models. *Journal for the Education of the Gifted, 19*(4), 405-436.
- Garner, B., Haugland, S., Spezia, M., & Boissoneau, R. (1990). Integrated learning: A viable alternative to gifted programs. *Early Child Development and Care, 63*, 107-118.
- Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic Books.
- Guilmette, T. J., Kennedy, M. L., & Queally, P. T. (2001). A comparison of the WISC-III and the Otis-Lennon School Ability Test with students referred for learning disabilities. *Journal of Psychoeducational Assessment, 19*(3), 239-244.
- Hertzog, N. B. (2004). Open-ended activities: Differentiation through learner responses. In C. A. Tomlinson (Ed.), *Differentiation for gifted and talented students* (pp. 77-103). Thousand Oaks, CA: Corwin Press.
- Horowitz, F. D. (1987). A developmental view of giftedness. *Gifted Child Quarterly, 31*, (4), 165-168.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.
- Kirschenbaum, R. J. (1998). Dynamic assessment and its use with underserved gifted and talented populations. *Gifted Child Quarterly, 42*(3), 140-147.
- Kornhaber, M. (1999). Enhancing equity in gifted education: A framework for examining assessments drawing on the theory of multiple intelligences. *High Ability Studies, 10*(2), 143-161.
- Kowall, M. A., Watson, G. M., & Madak, P. R. (1990). Concurrent validity of the Test of Nonverbal Intelligence with referred suburban and Canadian native children. *Journal of Clinical Psychology, 46*(5), 632-636.

- LeGagnoux, G., Michael, W. B., Hocevar, D., & Maxwell, V. (1990). Retest effects on standardized structure-of-intellect ability measures for a sample of elementary school children. *Educational and Psychological Measurement, 50*(3), 475-492.
- MacRae, L., & Lupart, J. L. (1991). Identifying gifted and talented students. *Roepers Review, 14*(2), 53-58.
- Maker, C. J., Rogers, J. A., Nielson, A. B., & Bauerle, P. R. (1996). Multiple intelligences, problem solving, and diversity in the general classroom. *Journal for the Education of the Gifted, 19*(4), 437-460.
- Matthews, D. J. (1998). Raven's matrices in the identification of giftedness. *Roepers Review, 10*(3), 159-162.
- Michael, W. B., & Bachelor, P. A. (1992). First-order and higher-order creative ability factors in structure-of-intellect measures administered to sixth-grade children. *Educational and Psychological Measurement, 52*(2), 261-273.
- Mills, C.J., & Tissot, S.L. (1995). Identifying academic potential in students from under-represented populations: Is using the Ravens Progressive Matrices a good idea?
- Morelock, M. J., & Morrison, K. (1999). Differentiating 'developmentally appropriate': The Multidimensional Curriculum Model for young gifted children. *Roepers Review, 21*(3), 195-201.
- Mulhern, J. D. (2003). The gifted child in the regular classroom. *Roepers Review, 25*(3), 112-115.
- Mumford, M. (2001). Something old, something new: Revisiting Guilford's conception of creative problem solving. *Creativity Research Journal, 13*(3-4), 267-276.
- Olenchak, F. R., & Renzulli, J. S. (2004). The effectiveness of the Schoolwide Enrichment Model on selected aspects of elementary school change. In C. A. Tomlinson (Ed.), *Differentiation for gifted and talented students* (pp. 17-37). Thousand Oaks, CA: Corwin Press.
- Ortiz, V., & Volloff, W. (1987). Identification of gifted and accelerated Hispanic students. *Journal for the Education of the Gifted, 11*(1), 45-55.
- Plucker, J. A., Callahan, C. M., & Tomchin, E. M. (1996). Wherefore art thou, multiple intelligences? Alternative assessments for identifying talent in ethnically diverse and economically disadvantaged students. *Gifted Child Quarterly, 40*, 81-92.
- Raven, J. (2000). *Raven manual research supplement 3: American norms*. Oxford: Oxford Psychologists Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 3 Standard Progressive Matrices*. Oxford: Oxford Psychologists Press.
- Reid, C., Udall, A., Romanoff, B., & Algozzine, B. (1999). Comparison of traditional and problem solving assessment criteria. *Gifted Child Quarterly, 43*(4), 252-264.
- Reis, S. M., & Westberg, K. L. (2004). The impact of staff development on teachers' ability to modify curriculum for gifted and talented students. In C. A. Tomlinson (Ed.), *Differentiation for gifted and talented students* (pp. 39-57). Thousand Oaks, CA: Corwin Press.
- Reis, S. M., Westberg, K. L., Kulikowich, J. M., & Purcell, J. H. (2004). Curriculum compacting and achievement test scores: What does the research say? In C. A. Tomlinson (Ed.), *Differentiation for gifted and talented students* (pp. 105-116). Thousand Oaks, CA: Corwin Press.
- Renzulli, J. (Ed.). (2004). *Identification of students for gifted and talented programs*. Thousand Oaks, California: Corwin Press.
- Richards, R. (2001). Millennium as opportunity: Chaos, creativity, and Guilford's structure of intellect model. *Creativity Research Journal, 13*(3-4), 249-265.
- Sarouphim, K. M. (1999). DISCOVER: A promising alternative assessment for the identification of gifted minorities. *Gifted Child Quarterly, 43*(4), 244-251.
- Sarouphim, K. M. (2001). DISCOVER: Concurrent validity, gender differences, and identification of minority students. *Gifted Child Quarterly, 45*(2), 130-137.
- Sarouphim, K. M. (2002). DISCOVER in high school. *Journal of Secondary Gifted Education, 14*(1), 30-38.
- Sattler, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Jerome M. Sattler, Publisher.
- Scott, M. S., Perou, R., Urbano, R., Hogan, A., & Gold, S. (1992). The identification of giftedness: A comparison of White, Hispanic and Black families. *Gifted Child Quarterly, 36* (3), 131 - 139.

- Scott, M. S., Deuel, L. S., Perou, R., Scott, M., Claussen, A. H., & Sanchez, M. (1996). Evaluating Potential Test Components for a New Cognitive Screening Test: A preliminary study. *Diagnostique*, 22, 21-38.
- Shaunessy, E., Karnes, F. A., & Cobb, Y. (2004). Assessing potentially gifted students from lower socioeconomic status with nonverbal measures of intelligence. *Perceptual and Motor Skills*, 98(3), 1129-1138.
- Stein, G., & Poole, P. (1997). Meeting the interests and needs of gifted children: A strategy for teaching and learning. *Early Child Development and Care*, 130, 13-19.
- Stephens, K., Kiger, L., Karnes, F. A., & Whorton, J. E. (1999). Use of nonverbal measures of intelligence in identification of culturally diverse gifted students in rural areas. *Perceptual and Motor Skills*, 88(3), 793-796.
- Sternberg, R. (Ed.). (2004). *Definitions and conceptions of giftedness*. Thousand Oaks, California: Corwin Press.
- Sternberg, R., & Zhang, L. (1995). What do we mean by giftedness? A pentagonal implicit theory. *Gifted Child Quarterly*, 39(2), 88-94.
- Sternberg, R. J. (Ed.). (1994). *Thinking and problem solving* (Second ed.). San Diego: Academic Press.
- Sternberg, R. J., & Clinkenbeard, P. R. (1995). The triarchic model applied to identifying, teaching, and assessing gifted children. *Roeper Review*, 17(4), 255-261.
- Tannenbaum, A. J. (1983). *Gifted children: Psychological and educational perspectives*. New York: Macmillan.
- Tyler-Wood, T., & Carri, L. (1993). Verbal measures of cognitive ability: The gifted low SES students' albatross. *Roeper Review*, 16 (2), 102-105.
- Vance, H. B., Hankins, N., & Reynolds, F. (1988). Prediction of Wechsler Intelligence Scale for Children-Revised Full Scale IQ from the Quick Test of Intelligence and the Test of Nonverbal Intelligence for a referred sample of children and youth. *Journal of Clinical Psychology*, 44(5), 793-794.
- VanTassel-Baska, J., Johnson, D., & Avery, L. D. (2002). Using performance tasks in the identification of economically disadvantaged and minority gifted learners: Findings from Project STAR. *Gifted Child Quarterly*, 46(2), 110-123.
- Whorton, J. E., & Morgan, R. L. (1990). Comparison of the Test of Nonverbal Intelligence and Wechsler Intelligence for Children – Revised in rural Native American and White children. *Perceptual and Motor Skills*, 70(1), 12-14.

Appendix A: GATE Identification Instruments
(See page 7 for a narrative discussion related to this outline)

Wechsler Intelligence Scale for Children (WISC) & Otis Lennon School Abilities Test (OLSAT)

- New WISC IV → better working memory assessment
- No literature on OLSAT 7 yet
- WISCIII > OLSAT 6 in identifying # of gifted students'
- WISCIII scores > OLSAT when testing 3-5 graders for disabilities
- $R(\text{WISC III, OLSAT}) = +$
- $\text{WISC-R} > \text{OLSAT}^{\wedge}$ in identifying more Hispanic 3-6 graders

WISCIII, WISC-R reveal higher scores and identify HIGHER # of gifted students than OLSAT

Structure of Intellect (SOI)

Due to the significant re-test effect

SOI should not be used for critical decisions when previous testing has occurred

Slosson Intelligence Test (SIT)

- $r(\text{WISC-R, SIT}) = +$
- $r(\text{WISC-R, SIT revised}) = +$

Researchers support the use of SIT full range as a screening instrument (especially for verbal ability)

Test of Nonverbal Intelligence (TONI)

- $r(\text{TONI, WISC-R}) =$ varying correlation
- In native American vs. Caucasian American comparison $r =$ moderate
- In African American vs. Caucasian American comparison $r =$ sig +
- In referring classroom problems, $r =$ low

Correlations between TONI and WISC-R depend on ethnicity—native American populations show lower r in TONI, WISC correlation than the Caucasian population

Though not used in Davis, the Raven and Alternative Assessment provide valuable comparisons:

Raven (SPM, CPM)

- $r(\text{SPM, other IQ}) =$ high but inconsistent
- $r(\text{SPM, achievement tests}) =$ lower than IQ r
- $r(\text{CPM, WISC}) = .22-.83$
- High variability, high reliability but lack national norm
- $\text{SPM} >$ Cultural-Free IQ Test, Neglieri Nonverbal Test in identifying largest # of students
- But the later 2 tests identifying students not found by SPM

Researchers suggest that one test is not sufficient for identifying culturally diverse gifted students

Alternative Assessment: (DISCOVER, PSA, STAR)

- Strength: increase identification of underrepresented students, stress measures of specific types of abilities
- Limitations: lack of objectivity, standardization, implementation of highly time-and resources consuming