GNETS STUDENT PERFORMANCE OUTCOMES, 2015-16

Executive Summary

This report contains results from the first full year beyond a three-year pilot effort to document student performance outcomes systematically across all GNETS programs. The same data collection and analyses procedures developed in the pilot years were used during the current 2015-16 school year. The *Developmental Teaching Objectives and Rating Form-Revised*, (DTORF-R, 2007) was the performance outcome measure, providing a composite score of student achievement in four developmental domains: *behavior, communication, socialization,* and *cognition.* Analyses of student performance outcomes used the DTORF-R assessment, analysis, and reporting procedures to obtain answers to these questions:

1. Did GNETS students make gains in DTORF-R competencies in each program and in each age group by the end of the current school year?
2. Did students make gains in center based and in school based locations?
3. How did students' scores change in comparison to age peers?
4. Did GNETS students achieve predicted goals for final outcome scores?
5. How could the system be sustained with integrity each year?

Repeated ratings with paired t-tests ($\partial = 0.05$), were used to analyze changes in DTORF-R scores from initial to final assessment for 2,069 GNETS students in 23 programs, in three age groups from pre-k/elementary to high school, at 46 center based (CB) and 69 school based (SB) locations, and in comparison to age peers. Findings indicate that mean DTORF-R scores increased significantly overall ($p < .0.0001$) and for each age group. At both CB and SB locations, GNETS students made significant group gains in each age group. However, mean scores were significantly lower in CB than in SB locations for both initial and final scores in all age groups. This suggests that students in CB locations may have different instructional needs. Slower developmental trajectories (rate of achievement) might be a major factor for educators and parents to consider. The findings also raise a question of need for clear descriptions of CB and SB goals and services, and student characteristics appropriate for CB or SB services.

In comparison to age peer scores, the pre-k/elementary group had the least delay (80.94%), while largest gap occurred for the high school group (69.72%). By the end of school year all three age groups had narrowed the differences significantly between their scores and those of their age peers.

Maintaining accuracy and a sustainable accountability system requires an ongoing system of professional development for using the DTORF-R to accommodate annual staff turnover among administrators and teachers. Opportunities for classroom teachers and paraprofessionals are essential. A classroom focus on teaching to DTORF-R competencies with Developmental Therapy-Teaching (DTT) instructional practices has resulted in improved student outcome scores. The DTORF-R guides instruction in social-emotional development and responsible behavior when used with DTT practices for both typically developing students and those with special needs across multi-tiered support systems (MTSS) with the same relevance for students in general and special education.

Background

In 2011-2012 a statewide work group of GNETS stakeholders was formed to develop a strategic plan for an accountability system appropriate to the needs of GNETS students with severe social-emotional and behavioral disabilities. The mission was to determine ways to collect annual student performance data from each of the 24 GNETS programs to measure student progress, guide instructional improvement, and document program effectiveness. To address these goals, the group identified ten standards unique to the needs of students in GNETS.

- Provide initial assessment, instructional applications, and outcome predictions for the IEP.
- Address specific needs of students with severe social, emotional, or behavioral challenges.
- Focus on content about positive social, emotional and behavioral development.
- Include desired competencies for each age group, Pre-K to grade 12.
- Measure observable competencies for growth, not deficits.
- Partner with other interventions, PBIS, social skills programs, and academic curricula.
- Provide metrics for analyzing student annual performance outcomes.
- Compare student outcomes to those of typically developing age peers.
- Facilitate program continuity and accountability across multi-tiered placements
- Convey clear meanings for parents, professionals, and paraprofessionals.

The outcome measure that met these standards was the DTORF-R (*Developmental Teaching Objectives and Rating Form-Revised*, 2007).[1] As a criterion referenced growth model, the DTORF-R defines key social, emotional, and behavioral competencies needed by children from birth through the teen years. Additionally, its sequential structure of competencies and system for rating and scoring provides comparisons between GNETS students and typically developing age peers. Used for three decades by teachers of students with EBD, the DTORF-R provides a metric with EBD content validity, field validity, and inter-rater reliability.[i] Attachment 1 contains a summary chart of DTORF-R content.

The DTORF-R Assessment and Accountability project began in 2012-13. The idea was to gradually phase in the system as teaching teams learned to use it accurately. In the first year, repeated assessments included 482 pre-k/elementary age students. The second year, 958 middle school age students were added. This group included some high school age students whose teaching teams had an option to include them. By the third year 2,082 students were included: 994 high school age, 605 middle school age, and 483 elementary age. They received GNETS services in 46 center based and 69 school based locations**.**

[1] Buros Center for Testing, *2016 Tests in Print, Mental Measurements Yearbook*. Available: www.marketplace.unl.edu/buros

Current Student Performance Outcome Study, 2015-16

Procedures

Seeking answers to the same five questions about student performance outcomes as during the pilot years, the same procedures for data collection, file reviews, and analysis were followed. GNETS teaching teams repeated DTORF-R assessments of 2,384 students in 23 programs at entry, mid-year, and end of the school year. There were 36 center based and 45 school based GNETS locations. The first assessments were treated as Initial (Time 1) measures, so it is possible that a student may have been in GNETS in previous years. After each assessment, results were used by teaching teams to adjust individual student programs when needed.

DTORF-R total scores were reported on the *DTORF-R Group Performance Record* online in CSV format or in print with Excel. ID designations, without identifying names, were pre-assigned by GNETS for each program (A – X). Each program then assigned IDs for their service locations, teachers, and by four numerals for individual student IDs. No student characteristic data except birthdates were collected. The programs submitted the total DTORF-R scores electronically to the Department of Education using the assigned code without identifying individuals, teachers, or locations by name. File reviews, conversion to SAS, analyses, and graphs were prepared by Developmental Therapy Institute staff and Dr. Kim Love, quantitative consultant with K. R. Love QCC.

Removing Invalid Data

DTORF-R scores of each student were reviewed initially for invalid data using the same protocol as in previous years. There were 315 files (13%) removed from the original pool of 2,384 files for one of four reasons: Age greater than 20.8 years or 250 months (8 files, 0.34%); less than 2 months or more than 9 months between ratings (35 files, 1.47%); a single rating or other missing data (78 files, 3.27%); or extreme gains or losses of questionable validity (194 files, 8.14%).[2] The final data set included 2,069 students retained as valid measures for the analyses.

Analyses

Frequency counts and paired t-tests ($\partial = 0.05$) were used to analyze changes in mean DTORF-R scores from initial to final assessment for 23 programs, three age groups, program locations, and comparison to age peers. To correct minor calculation errors, months between assessments, changes in scores between ratings, and student age at Time 3 were recalculated. If a Time 3 score was missing and a Time 2 score was recorded on or before March 1, the Time 2 score became a Time 3 score.

Three age groups were then formed: 449 students in the pre-k/elementary age group, ages < 121 months (M = 99.72, SD = 14.49); 518 students in the middle school age group, ages 121 through 156 months, (M = 138.96, SD = 10.33); and 1,062 in the high school age group, ages 157 through 250 months (Mean = 188.82, SD = 21.06).

---

[2] Extreme gains or losses outside the parameters of the DTORF-R structure: Gains/losses > 12 in 2 to 4 months (122 days); gains/losses > 20 in 4 to 5 months (123-183 days); or gains/losses > 28 in 6 to 9 months (183 days).

Ratings from Program Q (40 students) were excluded from overall analyses and then analyzed separately because the initial DTORF-R ratings were from the previous spring, completed four to five months earlier, before summer break. One of these locations also rated 33 students in the fall as the final measure. This atypical rating schedule made its data invalid for inclusion in GNETS group analyses.

Results

The results provide a highly positive picture of how GNETS students are progressing in achieving competencies needed by all students for success in school and in life. Significant gains were achieved by students in both school based and in center based programs. Findings addressing the study questions are described below in detail.

Outcome Scores Overall and by Age Groups (Table 1)
Results from analysis of group means and paired t-tests are shown in Table 1. Scores increased significantly overall and for each age group from initial to final ratings. The average time from initial to final ratings was 6.15 months (SD = 1.70).

Table 1. Changes in DTORF-R Scores Overall and by Age Group (Program Q excluded)

Age Group	N Obs.	Mean Months T1-T3 (SD)	Mean Initial Scores	SD	Mean Final Scores	SD	Mean Change (SD)	*t*-value	*p-value*
PreK/ Elementary	449	6.15 (1.71)	80.84	20.51	86.68	21.06	6.53 (8.03)	17.22	<0.0001
Middle School	518	6.18 (1.73)	100.14	22.46	104.83	23.50	4.93 (9.70)	11.56	<0.0001
High School	1,062	6.11 (1.65)	113.47	23.29	118.05	23.65	4.74 (10.06)	15.34	<0.0001
Overall	2,029	6.15 (1.70)	102.92	25.81	108.10	25.92	5.18 (9.57)	24.37	<0.0001

Progress Within Programs (Tables A-W at end of narrative)
Similar analyses were conducted to examine changes over time for each program. Results indicate that 19 of the 22 programs (Q excluded) had significant gains in mean scores overall. Seven of these programs had significant gains in all three age groups; five programs had significant progress in pre-k/elementary and high school groups, but not for middle school groups. Three programs had no significant changes in mean scores (Programs I, O, and W).

Not surprisingly, in Program Q , because of invalid rating schedules before summer break, mean DTORF-R scores decreased significantly from initial to final ratings as a group and within each age group. This is the only program where the mean change was significantly negative.

Center Based and School Based Comparisons (Table 2)

For this analysis, 1,901 student files contained center based (CB) or school based (SB)

information. This location information was not available from several programs, including all files from Programs Q and W. Results of t-test comparisons indicate mean scores increased significantly for both groups, the 1,216 students in CB ($p<0.0001$) and 685 students in SB locations ($p<0.0001$), and in each age group. Mean scores for all age groups were significantly lower in CB than in SB locations for both initial and final scores, shown below in Table 2.

Table 2. Comparison of Mean Scores at Center Based (CB) and School Based (SB) Locations (Programs Q and W excluded)

Age Group	Time	CB N	SB N	CB Mean Score (SD)	SB Mean Score (SD)	Difference (CB - SB)	t Statistic	P Value
PreK/Elementary	T1	262	170	77.67 (20.53)	85.44 (19.69)	-7.78	-3.90	0.0001
	T3	262	170	85.44 (21.00)	90.60 (19.44)	-5.16	-2.57	0.0105
	Change	262	170	7.77 (8.44)	5.16 (6.98)	2.61	3.49	0.0005
	t Stat			14.90	9.63			
	P Val			<0.0001	<0.0001			
Middle School	T1	318	165	97.37 (20.14)	104.36 (22.71)	-6.99	-3.46	0.0006
	T3	318	165	101.80 (20.53)	109.95 (23.60)	-8.15	-3.76	0.0002
	Change	318	165	4.43 (9.17)	5.58 (10.34)	-1.15	-1.25	0.2101
	t Stat			8.61	6.94			
	P Val			<0.0001	<0.0001			
High School	T1	636	350	110.81 (21.11)	119.22 (22.39)	-8.41	-5.86	<0.0001
	T3	636	350	116.37 (22.18)	123.46 (22.08)	-7.09	-4.81	<0.0001
	Change	636	350	5.56 (10.22)	4.24 (8.59)	1.32	2.15	0.0315
	t Stat			13.71	9.23			
	P Val			<0.0001	<0.0001			
Total	T1	1216	685	100.15 (24.49)	107.26 (25.86)	-7.10	-5.95	<0.0001
	T3	1216	685	105.89 (24.78)	112.05 (25.65)	-6.16	-5.13	<0.0001
	Change	1216	685	5.74 (9.66)	4.79 (8.70)	0.95	2.19	0.0023
	t Stat			20.73	14.42			
	P Val			<0.0001	<0.0001			

Difference scores showing changes from initial to final mean scores for CB and SB locations revealed that pre-k/elementary and high school groups in CB locations had significantly greater score increases than those in SB locations. There was no difference in change between the CB and SB locations for the middle school group.

Comparison to Age Peer Scores (Table 3)

For comparison of GNETS student progress to their typically developing age peers, extrapolated scores were calculated using each student's age in months at initial rating and the rate of mastery achieved by typically developing students over the same time from initial to final rating (*DTORF-R User's Manual*, 2012, pp. 17, 22-23).

These comparisons to an equivalent age peer DTORF-R score are reported as a

percentage of a student's actual score. A low percentage indicates a large developmental lag, while a high percentage suggests a student is approaching competency levels of a typically developing age peer. When students' scores approach age peer equivalent scores (> 90% of age peer score) this metric offers support to staff decisions to transition a student into an inclusive program.

At the beginning of the school year (T1), the largest gap between GNETS students' scores and their age peers occurred for the high school group with a corresponding age peer equivalent ratio of 69.72%. There was less delay for the middle school group achieving 75.07% of age peer scores, and even less delay for pre-k/elementary with 80.94%. By the end of school year (T3), all three age groups had narrowed the gap between their scores and those of their age peers.

Table 3. Age Peer Comparisons (Program Q excluded)

Age Group	N	Average T1 % Age Peer (SD)	Average T3 % Age Peer (SD)	% Age Peer Change (SD)	t Statistic	P Value
PreK/Elementary	449	80.94% (19.99%)	83.20% (19.35%)	2.26% (7.81%)	6.13	<0.0001
Middle School	518	75.07% (17.47%)	75.88% (17.55%)	0.82% (7.02%)	2.65	0.0084
High School	1062	69.72% (14.11%)	71.72% (14.16%)	2.00% (6.15%)	10.62	<0.0001
Total	2029	73.57% (17.05%)	75.32% (16.93%)	1.76% (6.79%)	11.65	<0.0001

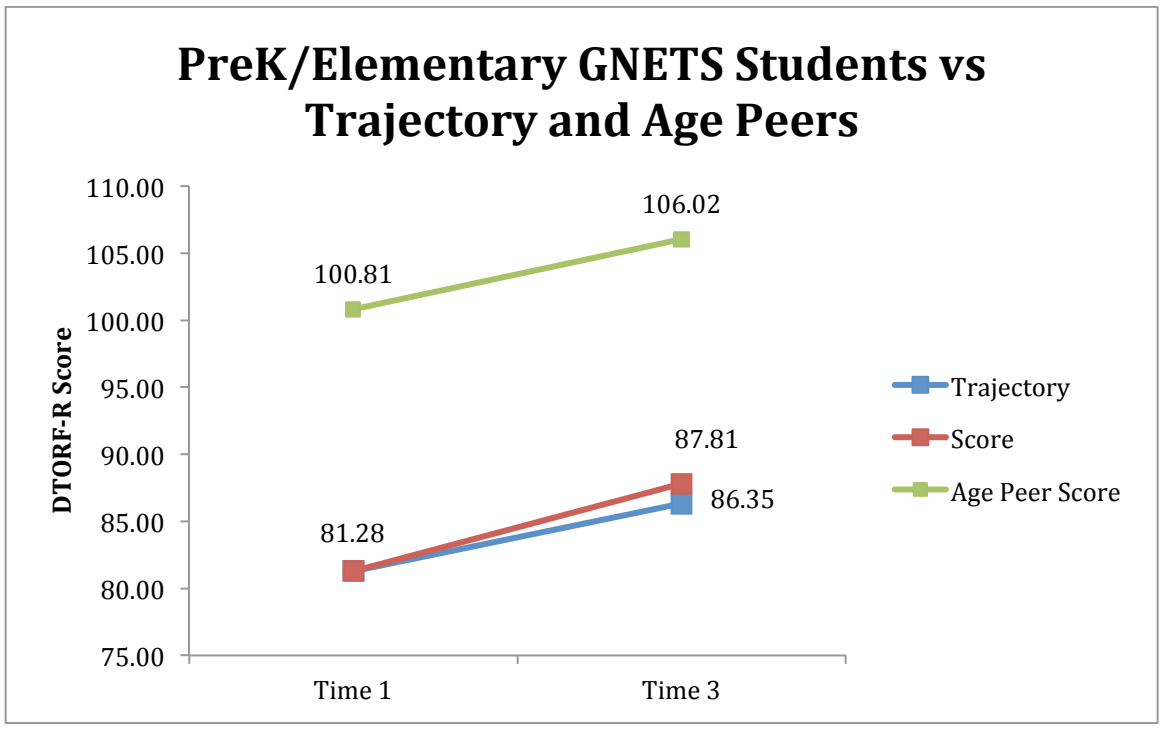
Growth Comparisons Between Student Progress, Past Development, and Typical Age Peer Progress (Table 4 and Figures 1-4; Program Q excluded)

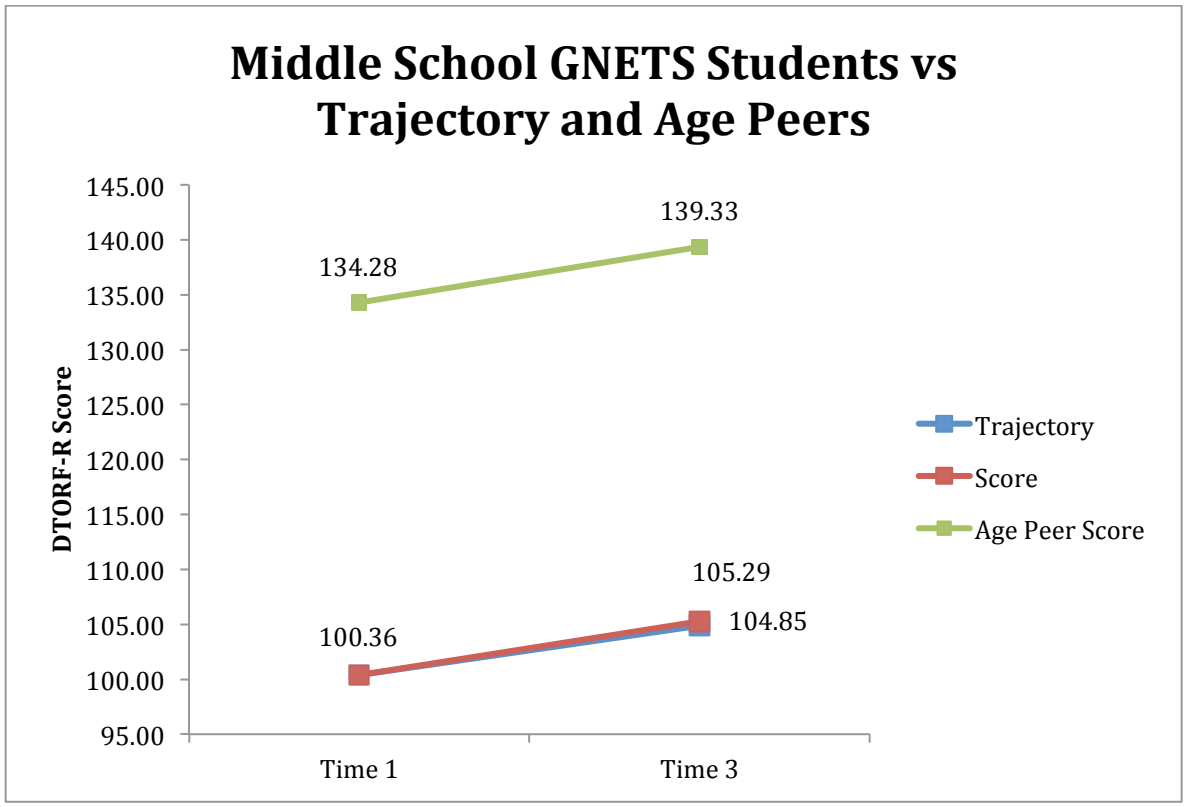
To obtain a comparison between past development and actual achievement during the year, a separate pre-intervention rate of item mastery was calculated for each student at the time of the initial rating. Based on the sequential structure of the DTORF-R, this rate was then projected to continue on a trajectory of scores adding items at that same rate each month from initial to final rating, assuming no intervention effect. These trajectories were used as proxies for growth based on rate of past development.

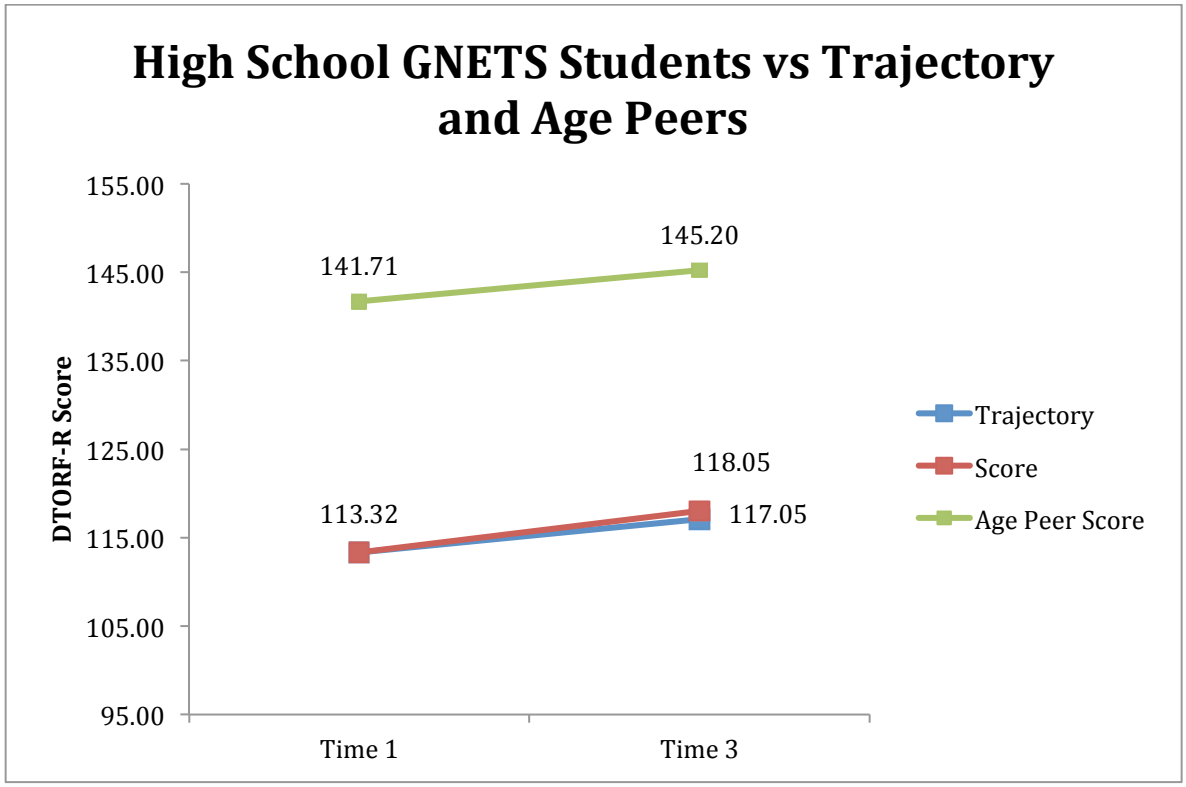
Table 4 contains mean ages, beginning and ending trajectory scores for past development, actual scores, and corresponding age peer scores. These means were then used to portray growth lines for comparison in Figures 1-4.

Table 4. Mean Scores for Students' Age, Trajectory Score, Actual Score, and Age Peer Equivalent Scores (Program Q excluded)

Age Group	N	Variable	Mean T1 (SD)	Mean T3 (SD)
PreK/Elementary	449	Age	99.72 (14.49)	105.88 (14.62)
		Trajectory Score	81.28 (21.71)	86.35 (21.71)
		Actual Score	81.28 (21.71)	87.81 (20.54)
		Actual Peer Score	100.81 (12.29)	106.02 (12.45)
Middle School	518	Age	138.96 (10.33)	145.11 (10.72)
		Trajectory Score	-	104.85 (23.52)
		Actual Score	100.36 (22.45)	105.29 (23.37)
		Actual Peer Score	134.28 (8.85)	139.33 (8.93)
High School	1062	Age	188.82 (21.06)	194.94 (20.96)
		Trajectory Score	-	117.05 (23.95)
		Actual Score	113.32 (23.20)	118.05 (23.71)
		Actual Peer Score	162.63 (7.12)	164.63 (6.23)
Total	2029	Age	156.37 (40.56)	162.50 (40.54)
		Trajectory Score	-	107.14 (26.34)
		Actual Score	102.92 (25.81)	108.10 (25.92)
		Actual Peer Score	141.71 (26.34)	145.20 (24.92)




Middle School GNETS Students vs Trajectory and Age Peers
DTORF-R Score
145.00
140.00
135.00
130.00
125.00
120.00
115.00
110.00
105.00
100.00
95.00
139.33
134.28
105.29
104.85
100.36
Time 1
Time 3
Trajectory
Score
Age Peer Score


High School GNETS Students vs Trajectory and Age Peers
DTORF-R Score
155.00
145.00
135.00
125.00
115.00
105.00
95.00
145.20
141.71
118.05
117.05
113.32
Time 1
Time 3
Trajectory
Score
Age Peer Score

Figures 1-4. Means for actual DTORF-R scores (red) compared to past development trajectory scores (blue) and age peer scores (green) during the same time period (6.15 months).

Achievement of Predicted Goals

Using a complex calculation procedure at the time of initial ratings, teaching teams were asked to predict a goal for each student's actual outcome score six months before final ratings. A frequency count of predicted goals achieved revealed that 56% of the students achieved their predicted goals and another 19% made progress toward the predicted goals. These findings are problematic. Currently the process involves three options for predicting a DTORF-R outcome score: (1) Achieved age peer equivalent score, (2) Achieved maximum predicted score, or (3) achieved score in a predicted range. These options are extremely difficult to achieve, as they indicate that a student will close the existing gaps between their own scores and those of corresponding age peers. For non-statistical individuals, the formulas are difficult to understand, and the inclination in to predicted the optimistic maximum score possible, not a realistic option. Consideration should be made to redesign or eliminate this function on the DTORF-R and E-DTORF formats.

Discussion

Student Progress in GNETS Programs

In 2015-16, 23 GNETS programs submitted initial-to-final DTORF-R outcome data for students initially identified with serious social, emotional, or behavioral disabilities. During the six months between ratings, mean scores for 2,029 students increased significantly in each age group. Additionally, outcome data from each program were analyzed separately. Mean scores increased significantly in 19 of the 22 programs (Program Q was excluded as described above). In seven of these programs, significant gains occurred in all three age groups.

Equally important is the finding that mean scores in 3 of the 22 programs showed no significant gains. It is possible that their programs did not focus on teaching and achieving the milestone competencies measured by the DTORF-R. It is also possible that teaching teams doing those ratings were unfamiliar with the competencies and the developmental progression of competencies from birth to age 17 years represented sequentially on the DTORF-R. This variability between programs emphasizes that the DTORF-R assessment and accountability process is being used effectively by some programs and not by others. Informal observations of these programs suggest that programs able to use rating information and translate it into developmentally enhanced classroom practices have better outcomes. Clearly, professional development opportunities for each program could improve GNETS student outcomes overall and within each program.

Results from the unusual rating schedule for students in Program Q, excluded from the overall analyses, illustrate the importance of using a fresh rating during the early fall each year. Ratings completed before summer break begins cannot be a reliable measure of how much is retained in the fall. Additionally, the failure of one location to complete a final rating in the spring of 2016 for Program Q contributed to the significantly negative change in mean DTORF-R scores for that program. However, it is not conclusive that services for students in Program Q were either ineffective or effective. The use of an old initial assessment with a summer break in between does not provide a fresh baseline to focus current instruction.

CB and SB Locations

Both initial and final mean scores for the 36 CB locations were significantly lower than the scores for the 45 SB locations. While each group made significant progress, the consistent difference in scores raises a question about the criteria used for initial placements in SB or CB programs. It may be helpful for educators participating in the referral and intake process for GNETS to define student characteristics appropriate for referral and assignment to CB or SB locations. The current study did not receive information about characteristics of students other than birthdates. While protecting student identity, this limitation leaves the question of student characteristics open for future studies

The findings also raise a question about a need for clear descriptions of CB and SB services. For example, SB may indicate full inclusion; or it may indicate pull out or self-contained classes in a regular school. Similarly, in some CB locations, students may have a schedule that included partial participation in their local schools' general education programs. It is clear that greater specificity would be helpful in determining which locations are CB and SB. Such clarity would give greater meaning to the metrics surrounding progress of students in both locations.

Developmental Progression of GNETS Students and Age Peers

Age is a specific characteristic that has mutual relevance for GNETS students and their typically developing age peers on the DTORF-R. GNETS students in the pre-k/elementary, middle school, and high school age groups made significant gains in life-skill competencies included in the DTORF-R, outlined in Attachment 1. These competencies were the same developmental milestone competencies of their age peers, but achieved at slower rates. Comparison of initial and final mean scores clearly portrays this growth trend. Both GNETS and age peers are following a universal developmental progression of key milestone competencies. However, while GNETS students' were achieving competencies, their corresponding age peer equivalent scores also were increasing in competencies during the same time. This resulted in GNETS students as a group still behind age peers by the end of the school year. This same developmental progression is seen in the mean scores for students served in CB and SB locations. These finding suggests that slower trajectories (rate of achievement) might be a major factor for educators and parents to consider as they set IEP goals at the time of entry into the program.

Integrity of DTORF-R Assessment and Accountability

These results provide a highly positive picture of how GNETS students are progressing in achieving DTORF-R competencies needed by all students for success in school and in life. These achievements occurred in both school based and in center based programs and in relation to typically developing age peers.

The process used to identify and remove invalid data provides confidence that the resulting metrics are accurate. From the original pool of 2,384 student files the process resulted in 8.14% excluded and 2,069 valid files (86.79% of the files submitted). This exclusion rate was less than the previous year (9.56%), when the high school age group was phased in for the first

time.

Maintaining accuracy and a sustainable accountability system with the DTORF-R requires an ongoing system of professional development to accommodate staff turnover each year. An open online introductory course was available to introduce new staff to using the DTORF-R with accuracy. Five regional workshops open to all programs were repeated twice during the year and offered follow-up guidelines for administrators and coordinators. Of the 24 programs, 17 sent at least one representative. To maintain a valid database over multiple years, on-going professional development must be available annually to new teachers, program administrators, and coordinators. Without administrative understanding of the DTORF-R and its classroom applications, the system becomes just another test to use and forget.

The most critical factor in student outcome at any location is the proficiency (or lack) in classroom teaching teams. Professional development opportunities for classroom teachers and paraprofessionals is essential. This may have contributed to the lack of gains in several programs this year. It may also explain the number of files excluded as extreme outliers. A classroom focus on teaching missing DTORF-R competencies with Developmental Therapy-Teaching (DTT) instructional practices has been shown in past studies to result in highly improved student outcome scores. When used with DTT, the DTORF-R guides needed instruction in social-emotional development and responsible behavior for both typically developing students and those with special needs across an array of placement options and locations. As educators become more experienced with DTT and the DTORF-R, applications become even more evident across multi-tiered support systems (MTSS) with the same relevance for students in general and special education.

Major Developmental Strands	**KEY MILESTONES**
Attachment	Aware of Others Interactions With Adults
Autonomy & Mastery	Interactions With Peers Play & Friendship Self Esteem
Social Uniformity	Conform to Others Use Values in Social Interactions
Social Self	Assert Self in Group Recognize Social Meaning Reciprocate Relationships (Empathy & Understanding)

"THINKING" Domain: **COGNITION** The Cognitive System	
Major Developmental Strands	**KEY MILESTONES**
Sensory-Motor & Memory	Attend & Recognize Respond & Imitate Visual Discrimination (Match & Sort)

Body Coordination	Fine and Large Muscle Skills Games & Dexterity
Pre schematic Concept Formation	Label Use of Objects Count Understand Opposites & Differences Categorize & Sequence
Formal Operations	Read, Write & Spell Number, Time, & Mathematics Logical Reasoning Human Behavior Personal Problem Solving & Daily Living