

The Perils of Confusing Performance Measurement with Program Evaluation

Cornelia Taylor, Kathleen Hebbeler, Donna Spiker
The Center for IDEA Early Childhood Data Systems
Christina Kasprzak
The Early Childhood Technical Assistance Center

A group of researchers recently published a [paper](#) critiquing the child outcomes performance indicator for Part C and Part B 619. They also presented some of their thoughts in a recent webinar sponsored by the Association of University Centers on Disabilities (AUCD). The researchers' critique is based on several faulty assumptions and consequently unfairly discredits the system for measuring child outcomes and the use of the data. Let's look at our concerns with their critique.

First, the authors have confused performance measurement with program evaluation. Their primary argument is that the child outcomes measurement requirement produces misleading information because it is based on a flawed evaluation design. The researchers' critique wrongly assumes that the child outcomes indicator is designed as an evaluation. The child outcomes measurement is not a program evaluation; it is one performance indicator embedded within a larger performance measurement system that is required by the [Individuals with Disabilities Education Act](#) (IDEA). States report on a number of performance indicators that address compliance with federal regulations and program results. As such, these indicators yield information that supports program improvement and ongoing monitoring of program performance. Performance measurement systems are common in both the public (for example, [Maternal and Child Health](#)) and the private sector (for example, the [Pew framework for home visiting](#)). The Office of Special Education Programs (OSEP) implemented the child outcomes indicator in response to the [Government Performance and Results Act](#) which requires all federal agencies report on results being achieved by their programs. OSEP also uses the child outcomes indicator data to monitor states on results achieved, consistent with the strong emphasis in IDEA to improve results for children with disabilities.

The Government Accounting Office has produced a succinct [summary](#) that highlights some of the differences between the performance measurement and program evaluation. *Performance measurement* refers to ongoing monitoring and reporting of program accomplishments. Performance measures may address program activities, services and products, or results. The OSEP child outcomes indicator is a performance measure that addresses results. Examples of other results performance measures are teen pregnancy rates, percentage of babies born at low birth weight, 3rd grade reading scores, and high school graduation rates. In contrast, *program evaluations* are periodic or one time studies usually conducted by experts external to the program and involve a more in depth look at a program's performance. *Impact evaluations* are a particular type of program evaluation that determine the effect of a program by comparing the outcomes of program participation to what would have happened had the program not been provided.

A major difference between measuring outcomes in a performance measure system versus a program evaluation is that a well-designed impact evaluation is able to conclude unequivocally that the results observed were caused by the program. Performance measures cannot rule out alternative explanations for the results observed. Nevertheless, performance measurement data can be used for a variety of

Performance Measurement Compared to Program Evaluation

Feature	Performance Measurement	Program Evaluation
Data collected on a regular basis, e.g., annually	Yes	No
Usually conducted by experts to answer a specific question at a single point in time	No	Yes
Provides information about a program's performance relative to targets or goals	Yes	Possibly
Provides ongoing information for program improvement	Yes	No
Can conclude unequivocally that the results observed were caused by the program	No	Yes, if well designed impact evaluation
Typically quite costly	No	Yes

purposes including accountability, monitoring performance, and program improvement. Data on performance measures such as the Part C and Part B Section 619 child outcomes indicator can be used to track performance compared to a target or to compare results from one year to the next within programs or states. They can be used to identify state or local programs that could benefit from additional support to achieve better results. Comparing outcomes across states or programs should be done with an awareness that they might serve different population which could contribute to different outcomes. The solution to this is not to conclude that results data are useless or misleading but rather to interpret the results alongside other critical pieces of information such as the performance of children at entry to the program or the nature of the services received. Two of OSEP's technical assistance centers, the Center for IDEA Early Childhood Data Systems ([DaSy](#)) and the Early Childhood Technical Assistance Center ([ECTA](#)), have developed a variety of [resources](#) to support states in analyzing child outcomes data including looking at outcomes for subgroups to further understand what is contributing to the results observed. Just like tracking 3rd grade reading scores or the percentage of infants who are low birth weight, there is tremendous value in knowing how young children with disabilities are doing across programs and year after year.

Second, the authors incorrectly maintain that children who did not receive Part C services would show the same results on the child outcomes indicator as children who did.

The researchers' claim that the results states are reporting to OSEP would be achieved even if no services had been provided rests on a flawed analysis of the [ECLS-B](#) data, a longitudinal study of children born in 2001. For their analysis, the authors identify a group of 24 months olds in the data set who they label as "Part C eligible children who did not receive Part C services." These children

- Received a low score on a shortened version of the Bayley Scales of Infant Development (27 items) administered at 9 months of age by a field data collector; and
- Were reported by a parent when the child was 24 months old as not having received services to help with the child's special needs.

Few would argue that the determination of eligibility for Part C could be replicated by a 27-item assessment administered by someone unfamiliar with infants and toddlers with disabilities. Furthermore, data from the [National Early Intervention Longitudinal Study](#) show that very few children are identified as eligible for Part C based on developmental delay at 9 months of age. The first problem with the analysis is assuming all of these children would have been Part C eligible. The second problem is that it is impossible in this data set to reliably identify which children did and did not receive Part C services. Parents were asked a series of questions about services in general; they were not asked about Part C services. As we and others who have worked with national data collections have learned, parents are not good reporters of program participation for a variety of reasons. The only way to confirm participation in Part C services is to verify program participation which the study did not do. Given that children who received Part C services cannot be identified in the ECLS-B data, no one should be making conclusions about Part C participation based on this data set.

The authors also argue that a measurement phenomenon called “regression to the mean” explains why Part C and Part B 619 children showed improved performance after program participation. In essence this argument says that improvements seen in the functioning of the children are not real changes but are actually due to measurement error. One can acknowledge the reality of errors in assessment results but to maintain that measurement error is the sole or even a major explanation for the progress shown by children in Part C and Part B 619 programs is absurd.

Moving Forward

State Part C and 619 programs are required by IDEA to report on multiple performance indicators including child outcomes as part of a larger performance measurement system. The child outcomes indicator was developed with extensive stakeholder input in order to maximize its utility to local programs, state agencies, and the federal government. The process of building the infrastructure needed to collect and use child outcomes data has been complex which is why states have been working on it for over ten years. State agencies continue to identify and implement strategies for improving the data collection and use of the data. We know that the data collection processes are not perfect and more work needs to be undertaken to address data quality and other concerns. Building a national system for measuring the outcomes for young children with disabilities receiving IDEA services is a long-term undertaking that requires ongoing effort to make the process better. Disparaging the performance indicator and the data reported by states based on incorrect assumptions and flawed analyses is not productive. Instead, the field needs to collectively engage in ongoing dialogue around critical issues of data quality, data analysis, and appropriate use of the data based on an informed understanding of what the child outcomes indicator is and is not. Part C and Part B 619 state agencies and OSEP are on the forefront of collecting and using early childhood outcomes data to improve programs – which is exactly what performance measurement is intended to do.

Posted February 27, 2018

The contents were developed under grants from the U.S. Department of Education, #H326P120002 and #H373Z120002. However, those contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal Government. Project Officers: Meredith Miceli, Richelle Davis, and Julia Martin Eile