Maintaining Reliability in Ontario’s Outcome Initiative (CAFAS): Rater Drift and Training Approaches

Melanie Barwick, Ph.D., C.Psych1,2  
Christine Omrin, M.Ed. 1  
Denice Basnett, M.A1  
1The Hospital for Sick Children  
Toronto, Ontario, Canada  
2Departments of Psychiatry and Public Health Sciences, University of Toronto

The 17th Annual Research Conference: A System of Care for Children’s Mental Health: Expanding the Research Base, Tampa, March 2, 2004

ONTARIO, CANADA


Functional outcomes measured by the Children and Adolescent Functional Assessment Scale (Hodges, 2000a).

The Hospital for Sick Children provides training and implementation support.

STAGES OF IMPLEMENTATION

1999       2000       2001       2002      2003       2004      2005

Mandate  
Reliability  
Training  
Reliability, Software, & Train-the-Trainer; Support Strategies  
Reliability, Software, Train-the-Trainer & Communities of Practice  
Reliability, Software, Train-the-Trainer & Communities of Practice; Policy

OBJECTIVE

(A) To assess rater drift on the CAFAS and (B) the reliability of a train-the-train method for training interrater reliability for new practitioners.

RATIONALE

Ontario’s CAFAS outcome data will contribute to clinical, organizational, and system delivery decisions. Assessing the reliability of the data set and developing a process for maintaining the legitimacy of the database over time is critical.

METHOD

A) Interrater Reliability & Rater Drift  
Rates: 315 raters selected from 1,000 practitioners who had achieved initial interrater reliability and had also completed a booster interrater reliability exercise one year later. Stratified sampling was used to select 35 raters from each of the 9 regions.

B) Train-the-Trainer Method  
Interrater reliabilities for 140 raters trained by an in-house trainer were compared to interrater reliabilities for 315 raters trained by the implementation team.
Interater Reliability: the consistency with which the same information is rated by different raters.

Rater Drift: lack of consistency within a single rater over time.

Criterion vs. Intraclass correlation: Rater reliabilities were compared to a criterion (gold standard) set by the CAFAS author rather than the sample of raters. High interrater reliability between raters may not be indicative of valid ratings.

---

Interater Reliability

Rater Drift

Criterion vs. Intraclass correlation

---

For Ontario:
Rater drift is evident on four subscales, statistically, yet coefficients remain adequate across all scores one year following initial training. Our 315 raters will be asked to complete 10 booster vignettes in two years. If reliabilities remain adequate after a two-year gap, all practitioners will re-establish reliabilities every two years.

For Behavioral Health Services:
Other systems of care may benefit from Ontario’s training methods and rater drift data in planning their own training and reliability policies.