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

Melanie Barwick, Ph.D., C.Psych[1,2].
Christine Omrin, M.Ed. [1]
Denice Basnett, M.A[1]

[1]The Hospital for Sick Children
Toronto, Ontario, Canada

[2]Departments 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

Measurement initiative mandated in 2000.

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

Mandate	Reliability Training	Reliability Training	Reliability Software, Train-the-trainer; Support Strategies	Software, train-the-trainer & communities of practice	Software, train-the-trainer & communities of practice; policy
1999	2000	2001	2002	2003	2004	2005

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.

CAFAS RATER RELIABILITY TRAINING

Day 1	Day 2
Context / Initiative	10 reliability vignettes
CAFAS background	Evaluation of reliability
Scoring rules	Supplemental vignettes
6 demonstration vignettes	

METHOD

A) Interrater Reliability & Rater Drift

Raters: 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.

RATER CHARACTERISTICS	Range	Analysis x Region
Clinical Experience	9.5 years – 18.8 years	F = 4.40, p< .0001
Job Description		nd
Clinician	79.4% - 91.2%	
Clinician Manager	8.6% - 20.6%	
Executive Director	0 - 5.7%	
Level of Education		X = 74.24, p<.001
MA, MEd,MSW (37%)	17% - 63%	
Child & Youth Worker (30%)	8.3% - 50%	
BA.BSW,BSc (22%)	11.8% - 37.5%	
Social Service Worker (7%)	2.9% - 11.8%	
PhD,EdD (3%)	0 – 11.8%	
MD (1%)	0 – 8.6%	

Interrater 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.

RESULTS FOR RATER DRIFT	Reliability (Pearson Correlation)			
CAFAS Scores	Initial Reliability (m r) (N=315)	1-Year Reliability (m r) (N=315)	Paired t	Direction of Drift
School	.954	.973	-3.542***	
Home	.978	.972	2.026*	
Community	.973	.961	2.534*	
Behavior Towards Others	.954	.878	7.356***	
Moods & Emotions	.914	.952	-6.407***	
Self-Harm	.995	.970	6.758***	
Substance Use	.986	.996	-3.848***	
Thinking	.985	.964	6.002***	
CAFAS Total	.990	.989	1.898	
Material Needs[1]	No data	.961	-	No data
Social Support[1]	.848	.828	-1.324	

RESULTS FOR TRAINING METHOD	Reliability (Pearson) Correlation		
CAFAS Scores	Trained In-House Trainer (m r) (N=140)	Trained by CAFAS in Ontario (m r) (N=315)	t
School	.940	.954	1.683
Home	.971	.978	1.972*
Community	.976	.973	-.467
Behavior Towards Others	.950	.954	.541
Moods & Emotions	.914	.914	.097
Self-Harm	.993	.995	.494
Substance Use	.988	.986	-.368
Thinking	.983	.985	.444
CAFAS Total	.989	.990	.744
Material Needs	-	-	-
Social Support	.839[1]	.847[2]	.598

IMPLICATIONS

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.

www.cafasinontario.ca