JoDD
Journal on Developmental Disabilities
Le journal sur les handicaps du développement

Volume 18, Number 3, 2012

BRIEF REPORT: Reliability of the York Measure of Quality of IBI (YMQI)

Authors

Shauna Whiteford,
Ksusha Blacklock,
Adrienne Perry

Department of Psychology,
York University,
Toronto, ON

Correspondence

perry@yorku.ca

Keywords

Intensive Behavioural Intervention,
inter-observer agreement,
observational measure,
intervention quality

© Ontario Association on Developmental Disabilities

Abstract

The York Measure of Quality of Intensive Behavioural Intervention (YMQI) is an observational measure designed to assess the quality of Intensive Behavioural Intervention (IBI) provided to children with autism, based on videos of 1-to-1 teaching sessions. This study examined various measures of reliability of the YMQI, including scale reliability and inter-observer agreement, as well as the potential of observer drift over time. Inter-observer agreement was very high, and reliability was consistent over time. Internal consistency was moderately good, while specific item-total correlations were variable.

While Intensive Behavioural Intervention (IBI) is widely believed to be the most effective treatment method for children with autism, outcomes are known to be highly variable across children. In a large study of the Ontario IBI program, roughly one-quarter of the sample achieved good outcomes, half showed more modest improvement, but another 25% did not improve (Perry et al., 2008). Several factors have been identified as contributing to this heterogeneity, including child factors such as age and IQ (Perry et al., 2011), family factors (e.g., Shine & Perry, 2010), as well as factors related to the treatment itself. Treatment variables could include the quantity and quality of intervention. While quantity may be measured by the number of hours of IBI per week ("intensity") and/or duration of treatment, quality has been much more difficult to measure, in part due to the lack of a valid and reliable way of measuring the quality of IBI.

A research team at York University set out to tackle this challenge and develop an empirically validated measure to evaluate the quality of videotaped IBI sessions. Perry and her colleagues developed the pilot version of the York Measure of Quality of IBI (YMQI), based on a review of the IBI literature, various training manuals and unpublished rating scales, a survey of parents and professionals (Perry, Prichard, & Penn, 2006), as well as their own clinical experience. The reliability (Prichard, 2005) and validity (Penn, 2005) of the pilot version were promising (Penn, Prichard, & Perry, 2007), but revisions were made in an attempt to improve the psychometric properties.

The current version of the YMQI (Perry, Flanagan, & Prichard, 2008) includes ratings on a 5-point scale (1 to 3 with half-points) of 31 individual items within nine broad categories (see Table 1), scored from two 5-minute video segments. The YMQI manual (Perry et al., 2008) describes the scoring procedure and includes a self-guided training DVD. However, the reliability of coders trained using the DVD has not yet been evaluated. This paper examines the reliability of the current version of the YMQI.

Table 1. Categories and Individual Items of the YMQI (Perry, Flanagan, & Prichard, 2008)

A. Discriminative Stimuli
1. Attending during S^Ds
2. Varying S^Ds

B. Reinforcement
3. Rapid reinforcer delivery
4. Motivational reinforcers
5. Varying reinforcers
6. Relation of reinforcers to the task
7. Sincere/motivating verbal reinforcers
8. Differential reinforcement

C. Prompting
9. Effectiveness of prompts
10. Fading and augmenting of prompts
11. Lack of prompting errors
12. Follow through
13. Implementation of error correction

D. Organization
14. Clear plan and teaching goals
15. Accessible materials

E. Pacing
16. Length of inter-trial intervals
17. Suitable pace for the child
18. Intensive teaching

F. Teaching Level
19. Suitable task difficulty
20. Evidence of skill acquisition

G. Instructional Control
21. On-task following requests
22. Maintenance of the child's focus

H. Generalization
23. Varying teaching materials
24. Mixing tasks
25. Teaching away from the table
26. Teaching embedded in naturalistic activities
27. Response generalization
28. Flexible teaching

I. Problem behaviour
29. Result of problem behaviour
30. Reinforcement of appropriate behaviour
31. Use of prevention strategies

Method

The study received ethics approval at York University.

Videotapes of children in IBI sessions, collected for another study (Dunn Geier, Freeman, & Perry, in progress), were scored by one of three undergraduate-level coders. The coders were trained using the DVD with one additional in-person training session with the project coordinator (the third author). The extra coaching improved reliability during training beyond that obtained based on the self-guided DVD alone (Blacklock, Perry, & Whiteford, 2011). Coders did not begin coding until they had passed a written test and achieved over 80% on three videos (this took several months). Two additional coders did not meet this standard and thus did not continue. Once coders were reliable, it took them approximately one hour to code the two 5-minute segments from each video.

Reliability calculations were based on 33 randomly chosen videos (25% of those that had been coded for the other study, prior to the beginning of the current study). These 33 videos were re-evaluated by a fourth coder (the first author), trained in the same way as the original coders. Inter-observer Agreement (IOA) was calculated, using percentage agreement, for each of the 33 videos, as well as for the 31 individual items of the YMQI.

Results and Discussion

Inter-observer agreement was calculated for each item based on agreement within one half point, across all 31 items. IOA ranged from 74% to 97% [*M* = 88.95%, *SD* = 5.93], for the 33 videos, which is very high. IOA for the individual items was also very high, as shown in Table 2, with more than half (*n* = 17) above 90%. These current results are an improvement over the previous study in which Prichard (2005) reported inter-rater reliability ranging from 71% to

Table 2. IOA for Individual Items

Item	*IOA (%)*	*Item*	*IOA (%)*
1	97	17	99
2	76	18	93
3	85	19	99
4	84	20	87
5	81	21	94
6	81	22	100
7	100	23	74
8	93	24	93
9	87	25	91
10	82	26	93
11	81	27	93
12	97	28	97
13	82	29	78
14	91	30	93
15	82	31	87
16	90		

83%, with a mean of 77%. Thus, revisions made to the YMQI coding definitions and manual (and perhaps the training DVD) seem to have resulted in a higher level of IOA. However, it should be reiterated that this calculation is based, not on exact agreement, but agreement within one half-point on the YMQI scale.

As videos were coded by the three original coders over a period of 9 months, and up to a year since training began, observer drift was a potential concern. Therefore correlations between IOA and time since training (of the original coder) were also computed. As shown in Figure 1, IOA was found to be consistent over time, and showed no correlation with time since training (*r* = -.014), indicating that these coders were not drifting from the manual and coding rules, even over a prolonged period of time. This is encouraging, especially for studies taking place over a long time.

Internal consistency, as measured by Cronbach's alpha, was moderately high (α = .77), and item-total correlations for the 31 individual items ranged from poor to good (*r* = -.13 to .55) (Table 3). It may be that there are two or more factors in the YMQI with separate items loading

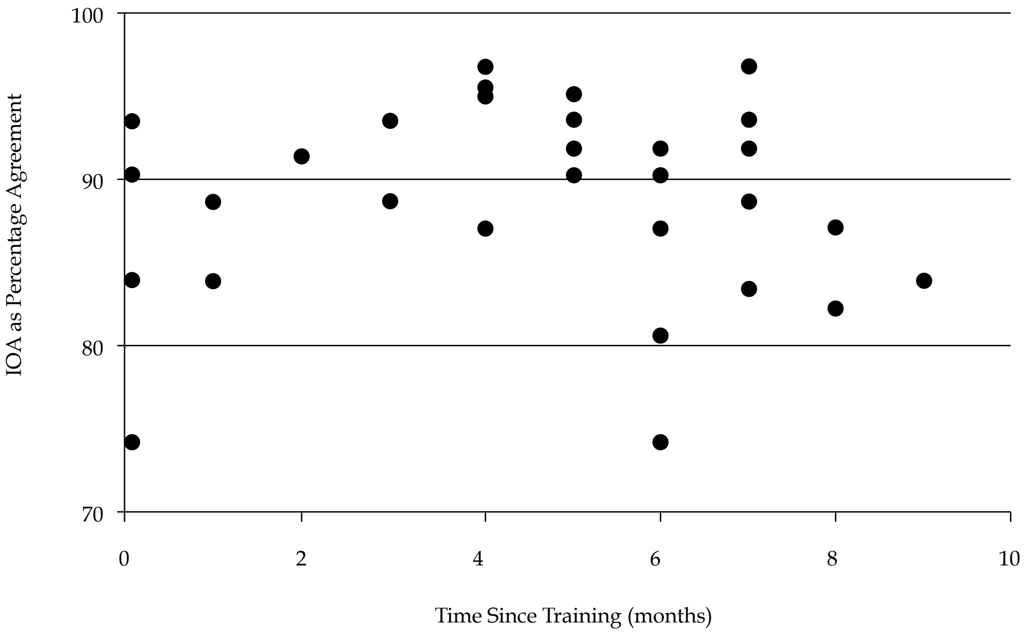

Figure 1. Correlation between IOA and time since training (r = -.014)

Table 3. Item-Total Correlation Ranges

r	*Total # of Items*	*Items*
-.20 to 0	4 (13%)	4, 5, 27, 28
0 to .20	5 (17%)	6, 23, 25, 26, 30
.20 to .40	9 (30%)	1, 2, 7, 8, 10, 11, 17, 22, 31
.40 to .60	12 (40%)	3, 9, 12, 13, 14, 15, 16, 18, 19, 20, 21, 24
	30[a]	

[a] *one item omitted because of insufficient data*

onto them, a possibility to investigate in future research. However, none of the items would significantly improve internal consistency if removed, based on alpha with item deleted calculations (not shown). Interestingly, this alpha level was somewhat lower than in Prichard's (2005) study, where she found a higher coefficient (α = .89). This may be a function of the particular sample of videos and the types of IBI depicted perhaps, which may have been more variable than the previous sample of videos.

Study limitations include possible issues of generalizability. The coders used in this study may not be representative of other people who may try to use the training DVD. Not all coders who started with the project attained this level and we found that one extra in-person training session was necessary to get coders over 80% initially (Blacklock et al., 2011). It is important to consider that these coders had access to a project coordinator as well as one of the YMQI's developers. The second coder for reliability was always the same person (the first author), which may have impacted upon the results, rather than having a reliability videos assigned randomly to any of the group of coders. Furthermore, it is also hard to evaluate whether the videos used here are a representative sample of IBI as delivered in other contexts, in terms of quality or difficulty of scoring. Finally, the number of videos was modest (*n* = 33). It would be beneficial for additional research to be done to replicate these findings in different (hopefully larger) samples, with different coders and different types of IBI, and with different children.

Nevertheless, the study is useful in showing that the quality of IBI sessions can be assessed reliably using the YMQI. This also means that the YMQI can be used in studies like the one currently underway in Ontario (Dunn Geier et al., in progress), in which YMQI scores are being used as a predictor of outcome, along with other variables such as quantity of IBI and child characteristics. This will allow us to assess how much children's outcomes are accounted for by quality of treatment, something that was not possible before the development of the YMQI.

Key Messages From This Article

People with disabilities: It is important to know if the services for children with autism are good quality.

Professionals: Reliable measurement of the quality of intervention services is essential. The YMQI provides a reliable measure of the quality of Intensive Behavioural Intervention for children with autism.

Policymakers: Reliable measurement of service quality is essential.

Acknowledgements

The authors are very grateful for the dedicated, careful work of the YMQI coders: Azin Taheri, Teresa Lee and Milena Stelmaszak.

This paper is based on the Honours Thesis of the first author, and it was presented in poster form at the OADD Research Special Interest Group conference in April 2011, in St. Catharines, ON.

For more information about the YMQI or to order a copy, please contact: perrylab@yorku.ca.

References

Blacklock, K., Perry, A., & Whiteford, S. (2011, May). *Current research on the York Measure of Quality of Intensive Behavioural Intervention*. Presentation at the Association for Behavior Analysis International, Denver, CO.

Dunn Geier, J., Freeman, N. L., Perry, A., (in progress). *Prospective controlled study of IBI in community settings.* Grant funded by the Provincial Centre of Excellence for Child and Youth Mental Health at CHEO.

Penn, H. E. (2005). *The validity of the York measure of quality of IBI.* Unpublished Master's thesis, York University, Toronto, ON.

Penn, H. E., Prichard, E. A., & Perry, A. (2007). The reliability and validity of a pilot version of the York Measure of Quality of Intensive Behavioural Intervention. *Journal on Developmental Disabilities, 13(3)*, 149–166.

Perry, A., Cummings, A., Dunn Geier, J., Freeman, N. L., Hughes, S., LaRose, L., et al. (2008). Effectiveness of Intensive Behavioral Intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 2,* 621–642.

Perry, A., Cummings, A., Dunn Geier, J., Freeman, N. L., Hughes, S., Managhan, T., et al. (2011). Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 5,* 592–603.

Perry, A., Flanagan, H. E., & Prichard, E. A. (2008). *Background and development of the YMQI.* Unpublished Manual. York University, Toronto, ON.

Perry, A., Prichard, E. A., & Penn, H. (2006). Indicators of quality teaching in intensive behavioral intervention: A survey of parents and professionals. *Behavioral Interventions, 21*, 1–12.

Prichard, E. A. (2005). *The reliability of the York measure of quality of intensive behavioural intervention.* Unpublished Master's thesis, York University, Toronto, ON.

Shine, R., & Perry, A. (2010). Brief Report: The relationship between parental stress and intervention outcome of children with autism. *Journal on Developmental Disabilities, 16*(2), 64–66.