

PROCEDURE FOR ESTABLISHING INTER-RATER RELIABILITY WITH THE *SPECIALINK EARLY CHILDHOOD INCLUSION QUALITY SCALE*

1. If possible, attend a full day of training in the use of the inclusion scales.
2. Buy or borrow a copy of the “How to Measure Inclusion Quality” DVD (produced by SpecialLink). Then, study and practice, using all the 17 items in the inclusion scales.
3. Schedule an appointment to spend a full day at an early childhood centre that includes at least two children with special needs in its preschool classroom. Explain to the director/supervisor that you will need to spend most of the morning observing the preschool classroom, approximately 1 hour to interview her, 1/2 hour to interview the lead ECE, 10 minutes to speak with another classroom staff and 10 minutes to speak with a support staff (such as a secretary, cook or maintenance person). You will also want to spend 10 minutes with the parent of a child with special needs, in person or on the telephone.
4. Visit the centre at scheduled time with another inclusion quality observer. Both of you will observe the same classroom and sit in on the same interviews. Interviewing can be done by either observer, with the other asking for clarification.
5. Observers will rate all the indicators in the inclusion scales *independently* of each other.
6. After completing all the indicators, the observers should complete their scoring of the scales. They should schedule at least one hour, outside of the observed centre and on the same day as the observations, to do this in close proximity to each other.
7. Each observer should then fill out her/his column of an inter-rater score sheet. If both observers agree on an item, the figure 100% should be written into the "Exact" score column. If they do not agree, the figure 50% should be written into the "within 1" column, if the scores of the two observers do not differ by more than “1.” When all 17 items are scored, the scores should be totaled. If observers agree on all 17 items, the total score is **1,700**. When that number is divided by 17, the inter-rater score is 100%, the highest possible level of inter-rater reliability. If their score is 1445, the inter-rater score is 85%, a sufficiently high level of agreement to qualify for inter-rater reliability.
8. If the inter-rater score is below 85%, observers will each look at all indicators in the items where they have disagreed by no more than 1 point to see whether they have made errors in reading the question (such as seeing an “or” as an “and”). This should be done when they are together. If the observed scores are more than one point apart, the process cannot continue for that item.
9. If this review yields an inter-rater score of 85%, they qualify for inter-rater reliability.
10. If this review does not yield an inter-rater score of 85%, observers will discuss interpretations of indicators in items where they have disagreed by no more than 1 point. Their own discussion may help them find agreement on outstanding items. Rescoring of items is acceptable if they agree that they interpreted an item differently and agree now on a single interpretation.
11. It is still possible that this process will not yield satisfactory inter-reliability scores. If that is the case, the observers should choose another centre and repeat the process. Try not to feel frustrated or angry if you have to go out and observe again. Some of the most esteemed researchers and consultants in our field had to observe two or three times before achieving inter-rater reliability.
12. If you are using this tool on a regular basis, it is important that you test yourself for inter-rater reliability at least after every five observations. Otherwise, you may lose your objectivity because of the types of centres that you observe. For example, if you are observing struggling centres in high risk situations, you may find yourself giving higher scores when in any doubt. On the other hand, if you are observing centres in well-supported areas, you may find yourself being tougher in the scoring. If you then get together with a colleague and observe for inter-rater again, you may find that your scoring has fallen victim to “statistical drift” — a condition you must counter with frequent inter-rater checks.

SpecialLink Early Childhood Inclusion Quality Scale Inter-rater Reliability Score Sheet

Date: _____

Room: _____

Centre: _____

Teachers: _____

Observers: _____

Practices Inter-Rater Subscale

	Observers		% Agreement		Comments
			Exact	Within 1	
1. <i>Physical Environment and Special Needs</i>					
2. <i>Equipment and Materials</i>					
3. <i>Director and Inclusion</i>					
4. <i>Staff Support</i>					
5. <i>Staff Training</i>					
6. <i>Therapies: PT; OT; S&L; Behavioural</i>					
7. <i>Individual Program Plans (IPPs)</i>					
8. <i>Parents of Children with Special Needs</i>					
9. <i>Involvement of Typical Children</i>					
10. <i>Board of Directors and Similar Units</i>					
11. <i>Preparing for Transition to School</i>					

Principles Inter-Rater Subscale

	Observers		% Agreement		Comments
			Exact	Within 1	
1. <i>Zero Reject</i>					
2. <i>Natural Proportions</i>					
3. <i>Same Hours/Days of Attendance Available</i>					
4. <i>Full Participation</i>					
5. <i>Maximum Feasible Parent Participation</i>					
6. <i>Leadership, Pro-active Strategies and Advocacy</i>					

Add % agreement totals for all 17 items.

Total Exact	+	Total Within 1

 $\div 17 =$ % (Your inter-rater score)

Reviewing Inter-Rater Scoring

7. Each observer should then fill out her/his column of an inter-rater score sheet. If both observers agree on an item, the figure 100% should be written into the "Exact" score column. If they do not agree, the figure 50% should be written into the "within 1" column, if the scores of the two observers do not differ by more than "1." When all 17 items are scored, the scores should be totaled. If observers agree on all 17 items, the total score is **1,700**. When that number is divided by 17, the inter-rater score is 100%, the highest possible level of inter-rater reliability. If their score is 1445, the inter-rater score is 85%, a sufficiently high level of agreement to qualify for inter-rater reliability.
8. If the inter-rater score is below 85%, observers will each look at all indicators in the items where they have disagreed by no more than 1 point to see whether they have made errors in reading the question (such as seeing an "or" as an "and"). This should be done when they are together. If the observed scores are more than one point apart, the process cannot continue for that item.
9. If this review yields an inter-rater score of 85%, they qualify for inter-rater reliability.
10. If this review does not yield an inter-rater score of 85%, observers will discuss interpretations of indicators in items where they have disagreed by no more than 1 point. Their own discussion may help them find agreement on outstanding items. Rescoring of items is acceptable if they agree that they interpreted an item differently and agree now on a single interpretation.

JOINING THE *SPECIALINK EARLY CHILDHOOD INCLUSION QUALITY* RESEARCH NETWORK

Professor Donna S. Lero, Jarislowsky Chair in Families and Work, Centre for Families, Work and Well-Being at the University of Guelph, has recently completed a 58-page report entitled *Assessing Inclusion Quality in Early Learning and Child Care in Canada with the SpecialLink Child Care Inclusion Practices Profile and Principles Scale* (2009), for the Canadian Council on Learning. Analyses confirmed the reliability and utility of the *Practices* and *Principles* subscales for assessing inclusion quality in centres that already include children with special needs and inclusion capacity in centres preparing for inclusion. The report is available at <http://www.ccl-cca.ca/ccl>. Journal articles developed from that study will be available. See <http://www.specialinkcanada.org> for updates on publications.

Analyses of the *SpecialLink Principles and Practices* scores were based on data obtained in over 550 classrooms, in a purposive, voluntary sample of 216 child care centres in eight provinces. With permission from centre directors and agency consultants, observers who were trained at 2-day workshops that included direct observations of classrooms, as well as in inter-rater reliability, provided the data. All centres were identified by code numbers only, and the research procedures were approved by the University of Guelph's Research Ethics Review Board.

According to Dr. Lero, "This report provides strong evidence for the utility and reliability of both *SpecialLink* inclusion measures when used together as measures of inclusion quality. In addition, the validity of both measures is supported. Scores on the *SpecialLink Inclusion Principles Scale* assess the extent to which early learning programs have consciously adopted principles that reflect a strong commitment to include all children in the community, to ensure their full participation in the program, and to support their parents as full partners. Scores on the *SpecialLink Inclusion Practices Profile* assess the extent to which staff focus on individualized approaches to support children's unique needs, and the physical and human resources available in the centre to support their efforts. Both measures predict directors' ratings of how well they feel their centre is doing in including children with special needs."

The SpecialLink staff and Dr. Lero hope that many users of the *Inclusion Scale* will continue to collect data for ongoing research, while they are doing their observations. We can offer email consultation regarding issues that arise during your observations (info@specialinkcanada.org) as well as the opportunity for you to be part of a project that will improve the *Scale* and help provide the community with a better instrument for assessing inclusion quality.

If you are interested in participating in the SpecialLink Research Network, please contact us at info@specialinkcanada.org