Developmental Disabilities Bulletin, 2008, Vol. 36, No. 1 & 2, pp. 67-80

Assessment of the reliability and validity of the Discrete-Trials Teaching Evaluation Form

Danielle A. Babel, Garry L. Martin, Daniela Fazzio, Lindsay Arnal, and Kendra Thomson
University of Manitoba
St. Amant Research Centre

Discrete-trials teaching (DTT) is a frequently used method for implementing Applied Behavior Analysis treatment with children with autism. Fazzio, Arnal, and Martin (2007) developed a 21-component checklist, the Discrete-Trials Teaching Evaluation Form (DTTEF), for assessing instructors conducting DTT. In Phase 1 of this research, three experts on DTT rated all 21 components of the DTTEF as very important, demonstrating its high face validity. In Phase 2, the DTTEF had high interobserver reliability for live scoring of trainees' DTT performances, and it differentiated between the DTT performances of trainees before and after receiving instruction in applying DTT. In Phase 3, the DTTEF evaluations of the DTT performances of trainees in Phase 2 compared favorably to ratings of video clips of those performances by DTT experts, demonstrating high concurrent validity.

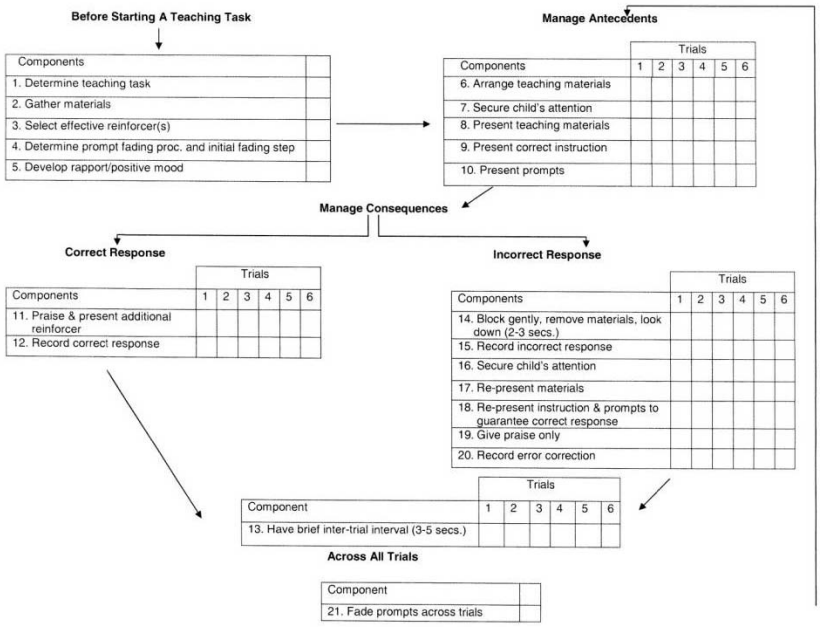
Intensive behavioral intervention based on Applied Behaviour Analysis (ABA) is considered to be the treatment of choice for children with autism (NYSDOH, 1999; Tews, 2007). A commonly used method for implementing ABA training sessions is known as discrete-trials teaching (DTT). A discrete trial involves the presentation of an antecedent by an instructor, followed by a response by the learner, and followed by the delivery of a consequence contingent upon the learner's response. Discrete trials are repeated many times in fairly rapid succession during a teaching session (Smith, 2001). In ABA training programs for children with autism in North America, there is a great need for training procedures to teach DTT to instructors and parents who conduct the training sessions. In addition, there is a need for an evaluation system for reliably evaluating the accuracy and consistency with which instructors and parents apply DTT. To meet this latter need, Fazzio, Arnal, and

Martin (2007) developed the Discrete-Trials Teaching Evaluation Form (DTTEF). The purpose of this research was to assess the reliability and validity of the DTTEF.

The DTTEF was developed in two stages. First, the authors of the DTTEF observed a large number of training sessions conducted by staff in the St. Amant ABA Preschool Program for Children with Autism. That program is a government-funded program in Manitoba that was initiated in 1999, and that currently funds intensive ABA treatment for a total of 58 children with autism from the ages of 2-6. Training staff include tutors (individuals with a high school diploma or a BA degree in progress and appropriate training), senior tutors, ABA consultants (individuals with an MA or PhD degree in psychology with ABA specialization), and a clinical coordinator (an individual with a PhD in psychology and who is a Board Certified Behavior Analyst). In the program, each child receives 36 hours of one-one instruction per week, including 31 hours with tutors and a senior tutor, and 5 hours with a parent. Curricula for children are individualized and include cooperation training, visual matching, imitative behavior, receptive and expressive language, abstract concepts, play skills, self-help skills, socialization, school readiness, and classroom preparation. DTT is the main vehicle for implementing teaching sessions. On the basis of observations of teaching sessions in the St. Amant program, Fazzio et al. (2007) developed a 19-component checklist and scoring manual for evaluating DTT (see Figure 1, all components except 4 and 5).

The second stage in developing the DTTEF was to review published research that investigated a variety of strategies for teaching staff and parents to implement DTT (Arco, 1997; Crockett, Fleming, Doepke, & Stevens, 2007; Dib & Sturmey, 2007; Downs, Downs, Johansen, & Fossum, 2007; Gilligan, Luiselli, & Pace, 2007; Koegel, Glahn, and Nieminen, 1978; Koegel, Russo, & Rincover, 1977; Lafasakis & Sturmey, 2007; Leblanc, Ricciardi, & Luiselli, 2005; Ryan & Hemmes, 2005; and Sarokoff & Sturmey, 2004). An important variable of such studies is the number and variety of DTT components that were taught to trainees and measured by direct observation. With one exception, the number of DTT

Figure 1.
The 21 components of the Discrete Trials Teaching Evaluation Rating Form



components in the above studies ranged from 3 to 14. Downs et al. reported using a 30-component checklist to rate instructor performance but only mentioned a few of the DTT components. After reviewing the above studies, Fazzio et al. (2007) added Components 4 and 5 to the DTTEF to produce the 21-component DTTEF shown in Figure 1. As can be seen in Figure 1, the DTTEF includes components to be performed before a DTT session, and components to be performed before and after a child's response during each teaching trial throughout a DTT session. The components of the DTTEF that were used by previous researchers are summarized in Table 1 (larger print copy available from authors).

We evaluated the reliability and validity of the DTTEF in three phases. First, three experts on DTT were recruited to assess the face validity of each of the 21 components of the DTTEF. Second, the DTTEF was used to score live sessions of seven university students attempting to apply DTT

Table 1
DTT Items Included in Previous Research as Compared to the DTT Items Included in the DTEF

DTEF	Koegel et al. (1977)	Koegel et al. (1978)	Arco (1997)	Sarakoff & Sturmey (2004)	Leblanc et al. (2005)	Ryan & Hemmes (2005)	Crockett et al. (2007)	Dib & Sturmey (2007)	Gilligan et al. (2007)	Lafasakis & Sturmey (2007)	Downs et al. (2007)
1. Determine teaching task	✔	✔		✔			✔			✔	✔
2. Gather materials					✔	✔		✔			✔
3. Select effective reinforcer(s)	✔	✔		✔		✔	✔			✔	
4. Determine prompt fading procedure & initial fading step						✔		✔			
5. Develop rapport/ positive mood					✔						
6. Arrange teaching materials					✔	✔					
7. Secure child's attention	✔	✔	✔	✔	✔	✔	✔	✔		✔	
8. Present teaching materials	✔	✔	✔		✔		✔				✔
9. Present correct instruction	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	
10. Present prompts	✔	✔	✔		✔	✔	✔	✔	✔		✔
Correct Response											
11. Praise & present additional reinforcer	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	
12. Record correct response				✔		✔	✔		✔	✔	
13. Have brief inter-trial interval (3-5 sec)	✔	✔		✔	✔	✔	✔		✔	✔	
Incorrect Response											
14. Block gently, remove materials, look down (2- 3 sec)	Extinction	Extinction	Verbal punishment				Verbal punishment				
15. Record incorrect response				✔					✔	✔	
16. Secure child's attention											
17. Re-present materials											
18. Re-present instruction & prompts to guarantee correct response				✔ Error correction	✔	✔ Error correction			✔ Error correction	✔	
19. Give praise only											
20. Record error correction				✔		✔	✔		✔	✔	
13. Have brief inter-trial interval (3-5 sec)	✔	✔		✔	✔	✔	✔		✔	✔	
21. Fade prompts across trials											

to a confederate role-playing a child with autism, before and after the students received instruction on DTT. This allowed us to evaluate the inter-observer reliability between two observers who used the DTTEF to score an instructor's DTT performance live. It also allowed us to use the DTTEF to compare the pre- and post-training DTT performances of the university students. While the university students were attempting to apply DTT to the confederate, we videotaped them and prepared clips of pre- and post-training performances of those students. Then, in Phase 3, we asked the three experts on DTT to watch the tapes and to rate the DTT performances of the seven students. This allowed us to assess the concurrent validity of the DTTEF by comparing our DTTEF assessments

of the students' live DTT performances to the ratings of the experts who watched video tapes of those performances.

Method

Participants and Settings

For Phase 1, to evaluate the face validity of the components of the DTTEF, and Phase 3, to evaluate the convergent validity of the DTTEF, we recruited 3 experts on DTT. The experts were female ABA consultants in the St. Amant ABA Preschool Program for Children with Autism. An important part of the job of the consultants was to supervise the DTT sessions conducted by the tutors, senior tutors, and parents. In Phase 1 the experts conducted face validity assessments (described later) in their respective offices, and in Phase 3 they rated the performances of students conducting DTT while watching a videotape of student performances in a research room at St. Amant, a community and residential resource centre for persons with developmental disabilities.

In Phase 2, in which the DTTEF was used to score live sessions of university students attempting to apply DTT to a confederate role-playing a child with autism, seven female university students were recruited from a second-year undergraduate psychology course taught at the University of Manitoba. Training and DTTEF assessments of the students were conducted in a quiet assessment room at St. Amant equipped with a table, two chairs, and a camera for videotaping.

Materials

In Phase 1, to assess the face validity of the DTTEF (as described later), the three DTT experts were given a list of the components of the DTTEF as shown in Figure 1. Also in Phase 1, the first and fourth authors studied the 11-page DTTEF scoring manual (Fazzio et al., 2007, available from the second author upon request). In Phase 2, the DTTEF score form shown in Figure 1 (also available on request) was used to score live sessions of seven university students who attempted, individually, to apply DTT to a confederate role-playing a child with autism, before and

after receiving instruction on DTT. During the "before-training" sessions, the students were provided with three 1-page abbreviated instructions (described in detail in Arnal et al., 2007) for teaching three tasks (described later) to children with autism, plus a data sheet for each task. The students used picture flashcards to teach the tasks, and edibles and small toys as reinforcers. All of these materials were also available to the students for the sessions that they conducted after receiving instruction on DTT. In addition, during the "after-training" sessions, the students were allowed to use Figure 1 as a prompt sheet. The training received by the students is described in Thiessen et al. (in press).

Procedure

Phase 1: Face validity. The three DTT experts were given the DTTEF score form shown in Figure 1 and a questionnaire to complete individually. The questionnaire asked the experts to rate each of the 21 items on the DTTEF using a seven-point scale where 1 = "not important," 4 = "somewhat important," and 7 = "very important." The experts were also asked to indicate if they believed that there were any items that should be added to the DTTEF.

Phase 2a: Inter-observer reliability (IOR). In Phase 2a, we assessed the IOR between two trained observers for live scoring using the DTTEF. First, the first and fourth authors studied the DTTEF scoring manual and then used the DTTEF score form (shown in Figure 1) to practice scoring a video, showing an instructor applying DTT to teach a confederate role-playing a child with autism, until they achieved 90 % IOR (computed as described below). Then the seven university students, during "before-training," were asked, individually, to attempt to teach three tasks to a confederate role-playing a child with autism, based on a one-page instruction sheet per task (as described in Arnal et al., 2007). One task involved 12 trials of teaching the confederate to point to named pictures. A second task involved 12 trials of teaching the confederate to perform a visual match-to-sample task. The third task involved 12 trials of teaching the confederate to imitate simple motor actions. While a student was teaching a task to the confederate, the two observers independently used

the DTTEF to score her performance live. Approximately one to two weeks after being scored on the three tasks, each participant was again scored on the three tasks as a part of the "before training" assessment. After all students had been scored on all three tasks twice, the students were then administered a self-instructional training package designed to produce mastery of DTT (as described in Thiessen et al., in press). After passing mastery tests on DTT, the students were once again asked to teach the three tasks to a confederate role-playing a child with autism, as they had done previously, except they were allowed to use the DTTEF score form (shown in Figure 1) as a guide. While they were doing so, the two observers once again used the DTTEF to independently score, live, the DTT performance of each student. In all teaching sessions, the performances of the students were videotaped for use in Phase 3, as described later.

An IOR score for a DTT session conducted by a university student was determined by comparing the DTTEF evaluations of that student that were recorded by the two observers. An agreement was defined as the two observers marking the same cell on the DTTEF score form with the same symbol (+ = correct, - = incorrect, and / = not applicable). A disagreement was defined as the two observers having a different symbol in the same cell. An IOR score was computed for a session by dividing the number of agreements by the number of agreements plus disagreements, and then multiplying by 100% (Martin & Pear, 2007).

Phase 2b: DTTEF assessments of participants before and after receiving training. The DTTEF scores obtained by the seven students while teaching the three tasks to a confederate before DTT training, and their DTTEF scores while teaching the three tasks after DTT training, were compared using a Wilcoxon signed-rank test.

Phase 3: Concurrent validity. In Phase 3, we evaluated the concurrent validity of the DTTEF by comparing DTTEF scores of the seven students in Phase 2 to a rating of their DTT performances by the three DTT experts. First, from the videotapes of the university students attempting to teach three tasks to a confederate role-playing a child with autism, the first author randomly selected the "before-training" and "after-training"

sessions of one task for each of the seven participants. She then prepared a videotape, to be viewed by the ABA consultants, that consisted of one 4-minute clip from before-DTT training and one 4-minute clip from after-DTT training of each of the seven trainees from Phase 2 described previously, 14 clips in total. Each clip was taken from the first four minutes of the observed session. Each of the 3 DTT experts then independently viewed all 14 clips. The order in which the experts viewed the before-DTT training versus the after-DTT training clips, and the order in which the seven participants were seen, were randomly selected. The experts were told that the clips showed university students in a study of DTT. However, they were not told which clips were before-DTT training and which were after-DTT training. The experts were asked to use their clinical experience to provide an overall rating of the DTT performance observed on each 4 minute clip, using a seven-point scale with 1 = "poor quality, comparable to DTT instructors prior to receiving training," 4 = "average quality, comparable to DTT instructors who have received limited training in discrete-trials teaching and have minimal experience," and 7 = "high quality, comparable to well-trained DTT instructors.. The mean rating of each clip of each student was then compared to the mean of the live DTTEF scores of that session with that student obtained in Phase 2.

Results

For the face validity evaluation in Phase 1, the three DTT experts evaluated each of the 21 components of the DTTEF with an average of six or higher on a seven-point scale. The experts did not report any components to be missing from the DTTEF. However, one expert suggested that there should be more information on shaping, and fading of prompts.

In Phase 2, IOR's between the two observers were assessed after they used the DTTEF to score the live performances of students who conducted DTT to teach a confederate role-playing a child with autism. An IOR of at least 90% was achieved for all but two sessions (42 out of 44 sessions), and the IORs for those two sessions were 88% and 89%.

In Phase 2 we also used a Wilcoxon signed-rank test to assess the DTTEF for distinguishing the performance of the students before and after receiving DTT training. The before and after scores differed significantly (W = -26, Z = -2.197, n = 7, p = .031, 2-tailed exact).

In Phase 3 we evaluated the concurrent validity of the DTTEF by comparing DTTEF scores of university students conducting DTT while teaching a confederate role-playing a child with autism, before and after the students received DTT training, to the before and after ratings of the DTT performance of those students by the DTT experts. The average before and after measures according to the DTTEF and according to the experts' ratings are shown in Table 2. As can be seen in Table 2, according to DTTEF evaluations, six of the seven participants showed considerable improvement after DTT training, and one participant showed a small decline in performance. According to the experts' ratings, the same six of the seven participants showed considerable improvement following DTT training, and one (the one that declined slightly according to the DTTEF) showed no change.

Table 2
Mean performance of participants as assessed by the Discrete-Trials Teaching Evaluation Form (DTTEF), and as rated by DTT experts (on a 1 to 7 scale with 7 as high), before and after receiving training to conduct discrete-trials teaching.

Participants	Mean % Correct DTEF Scores Before	Mean % Correct DTEF Scores After	Mean Expert Ratings Before	Mean Expert Ratings After
1	61	90	2.7	4.3
2	54	88	1	5.3
3	54	78	1	4
4	41	86	1	3
5	49	97	3.3	5.7
6	44	42	1.7	1.7
7	55	78	1.9	4.2
Mean	51	80	1.8	4.0

Discussion

In Phase 1, the face validity evaluation revealed that the 21 items on the DTTEF score form were rated at a mean of 6 or 7, on a seven-point scale, by the three DTT experts. This indicates that the components of the DTTEF are considered to be important to the assessment of DTT performance.

In Phase 2, the IOR evaluation showed very high reliability between the two observers for live scoring of DTT trainees using the DTTEF. Also in Phase 2, the evaluation showed that the DTTEF can be used to detect differences in DTT performance of trained and un-trained individuals.

In Phase 3 there was a high level of agreement between the DTTEF scores of trainees and their ratings by the DTT experts for detecting post-training improvement (or lack of it) for all seven trainees. This high level of concurrent validity occurred in spite of complaints from two of the experts that the videotapes were difficult to rate without knowing more about the prompt-fading procedures that were supposed to be used by the videotaped trainees.

One limitation of this study is that there were only three DTT experts who participated in Phases 1 and 3, and they were all from one location (St. Amant). Phases 1 and 3 should be replicated with additional DTT experts from a variety of ABA training programs for children with autism.

Another limitation of the study is that during Phase 2, to assess the reliability of the DTTEF for live scoring and the use of the DTTEF for detecting differences in performance of individuals before and after training, the data were collected while studying university students conducting DTT sessions with a confederate role-playing a child with autism. In subsequent research, we have used the DTTEF for assessing the DTT performance of university students conducting training sessions with children with autism, and in those studies we had very high IOR and social validity (Fazzio, Martin, Arnal, & Yu, in press; Thiessen et al.,

in press). Nevertheless, the reliability and validity assessments of the DTTEF should be replicated while scoring instructors and parents conducting DTT with children with autism.

Although intensive ABA intervention is widely considered to be the treatment of choice for children with autism, treatment outcomes remain highly variable with some children showing remarkable improvement while others improve minimally (Tews, 2007). In part to account for variability in outcome, and in part because millions of dollars are being spent on public programs to fund intensive ABA treatment of children with autism, reviewers of the outcome literature have urged the field to develop measures to assess quality of treatment (e.g., NRC, 2001; Schreibman, 2000; Wolery & Garfinkle, 2002). One approach to this problem is to identify key teaching characteristics, such as "adapts well to unexpected situations" or "creates opportunities for child-directed learning" that apply somewhat broadly to high quality behavioral intervention programs for children with autism (Perry, Prichard, & Penn, 2006). Another approach is to develop quality assessment systems for specific components of intensive behavioral intervention programs, such as the development of the DTTEF to assess the quality of one-on-one DTT sessions. We believe that these approaches are compatible and are both needed.

In summary, our research thus far demonstrates that the DTTEF has high face validity, high IOR for live scoring of persons conducting DTT sessions, high concurrent validity, and it can be used to distinguish between the performance of university students before and after they receive DTT training. Considering that a large number of instructors (parents, educators, and tutors) are needed to provide DTT in ABA early intervention programs for children with autism, the development of a reliable method for evaluating the quality of the DTT performances of such instructors is an important pursuit.

References

Arco, L. (1997). Improving program outcome with process-based performance feedback. *Journal of Organizational Behavior Management, 17(1)*, 37-64.

Arnal, L., Fazzio, D., Martin, G. L, Yu, C. T., Kielback, L., & Starke, M. (2007). Instructing university students to conduct discrete-trials teaching with confederates simulating children with autism. *Developmental Disabilities Bulletin*, *35*, 131-147.

Crockett, J. L., Fleming, R. K., Doepke, K. J., & Stevens, J. S. (2007). Parent training: Acquisition and generalization of discrete-trials teaching skills with parents of children with autism. *Research in Developmental Disabilities, 28*, 23-36.

Dib, N., & Sturmey, P. (2007). Reducing student stereotypy by improving teachers' implementation of discrete-trial teaching. *Journal of Applied Behavior Analysis, 40*, 339-343.

Downs, A., Downs, R. C., Johansen, M., & Fossum, M. (2007). Using discrete trial teaching within a public preschool program to facilitate skill development in students with developmental disabilities. *Education and Treatment of Children, 30*(3), 1-27.

Fazzio, D., Arnal, L., & Martin, G. (2007). *Discrete-trials teaching evaluation form (DTEF): Scoring manual*. Unpublished manuscript.

Fazzio, D., Martin, G. L., Arnal, L., & Yu, D. (in press). Instructing University Students to Conduct Discrete-Trials Teaching with Children with Autism. *Research in Autism Spectrum Disorders.*

Gilligan, K. T., Luiselli, J. K., & Pace, G. M. (2007). Training paraprofessional staff to implement discrete-trial instruction: Evaluation of a practical performance feedback intervention. *The Behavior Therapist, 30*, 63-66.

Koegel, R. L., Glahn, T. J., & Nieminen, G. S. (1978). Generalization of parent-training results. *Journal of Applied Behaviour Analysis, 11,* 95-109.

Koegel, R. L., Russo, D. C., & Rincover, A. (1977). Assessing and training teachers in the generalized use of behavior modification with autistic children. *Journal of Applied Behavior Analysis, 10,* 197-205.

Lafasakis, M., & Sturmey, P. (2007). Training parent implementation of discrete-trial teaching: Effects on generalization of parent teaching and child correct responding. *Journal of Applied Behavior Analysis, 40*, 685-689.

Leblanc, M. P., Ricciardi, J. N., & Luiselli, J. K. (2005). Improving discrete-trial instruction by paraprofessional staff through an abbreviated performance feedback intervention. *Education and Treatment of Children, 28*(1), 76-82.

Lovaas, O. I. (1982, August). *An overall evaluation of the young autism project*. Paper presented at the annual meeting of American Psychological Association, Washington, DC.

Martin, G., & Pear, J. (2007). *Behaviour modification: What is it and how to do it* (8th ed.). New Jersey: Prentice Hall.

National Research Council. (2001). *Educating Children with Autism.* Committee on Education and Interventions for Children with Autism: Division of Behavioral and Social Sciences and Education. Washington, D. C: National Academy Press.

New York State Department of Health. (1999). *Clinical Practice Guideline: The Guideline Technical Report. Autism/Pervasive Developmental Disorders, Assessment and Intervention for Young Children (Age 0-3 Years).* Author.

Perry, A., Prichard, E. A., & Penn, H. E. (2006). Indicators of quality teaching in intensive behavioral intervention: A survey of parents and professionals. *Behavioral Interventions, 21*, 85-96

Ryan, C. S., & Hemmes, N. S. (2005). Post-training discrete-trial teaching performance by instructors of young children with autism in early intensive behavioral intervention. *Behaviour Analyst Today, 1,* 1.

Sarakoff, R. A., & Sturmey, P. (2004). The effects of behavioral skills training on staff implementation of discrete-trail teaching. *Journal of Applied Behavior Analysis*, *37,* 535-538.

Schreibman, L. (2000). Intensive behavioral/ psycho-educational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders, 30*, 373-378.

Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Developmental Disabilities, 16*(2), 86-92.

Tews, L. (2007). Early intervention for children with autism: Methodologies critique. *Developmental Disabilities Bulletin, 35*(1/2), 148-168.

Thiessen, C., Fazzio, D., Arnal, L., Martin, G. L., Yu, D. C. T., & Kielback, L. (in press). Evaluation of a Self-Instructional Manual for Conducting Discrete-Trials Teaching with Children with Autism. *Behavior Modification*.

Wolery, M., & Garfinkle, A. N. (2002). Measures in intervention research with young children who have autism. *Journal of Autism and Developmental Disorders, 32*, 463-478.

Author Note

This research was supported by the St. Amant Research Centre and the Psychology Department of the University of Manitoba. For further information regarding this study, write to Garry L. Martin, University of Manitoba, 129 St. Paul's College, 430 Dysart Road, Winnipeg, MB, Canada, R3T 2M6, gmartin@cc.umanitoba.ca