4-1-2011

Validity and reliability of two abbreviated versions of the Gross Motor Function Measure.

Laura K Brunton

Doreen J Bartlett

Citation of this paper:

Brunton, L. K., & Bartlett, D. J. (2011). Validity and reliability of two abbreviated versions of the Gross Motor Function Measure. Physical therapy, 91(4), 577-588.

Validity and Reliability of Two Abbreviated Versions of the Gross Motor Function Measure

Laura K. Brunton[1], BKin, MSc, PhD Student, Health and Rehabilitation Sciences Program, Faculty of Health Sciences, The University of Western Ontario

Doreen J. Bartlett, PT, PhD, Associate Professor, School of Physical Therapy, Faculty of Health Sciences, The University of Western Ontario

[1] Address Correspondence to: Laura Brunton
1588 Elborn College, 1201 Western Road
The University of Western Ontario, London, Ontario, Canada N6G 1H1
Email: lbrunto3@uwo.ca

Validity and Reliability of Two Abbreviated Versions of the Gross Motor Function Measure

Abstract

Aim: The "gold standard" to measure gross motor functioning for children with cerebral palsy (CP) is the Gross Motor Function Measure (GMFM-66). The purpose of this study was to estimate the validity and reliability of two abbreviated versions (item set (GMFM-66 IS) and basal and ceiling (GMFM-66 B&C) approaches) of the GMFM-66.

Methods: Twenty-six children with CP aged 2 to 6 years across all GMFCS levels participated. At session one, both abbreviated versions were administered by two independent raters, followed by the full GMFM-66. In the subsequent session, only the abbreviated versions were administered, by the same raters. Concurrent validity, comparability between versions and test-retest reliability were determined using intraclass correlation coefficients (ICC 2,1).

Results: Both versions demonstrated high levels of validity with ICCs reflecting associations with the GMFM-66 of 0.99 (95% CIs ranging from 0.972-0.997). Both versions were also shown to be highly reliable with ICCs greater than 0.98 (95% CIs ranging from 0.965-0.994).

Interpretation: Both versions can be used in clinical practice or research. However, the GMFM-66-B&C is recommended as the preferred abbreviated version.

The current "gold standard" measure to obtain an estimate of gross motor functioning for children with cerebral palsy (CP) is the Gross Motor Function Measure (GMFM).[1] The GMFM is an evaluative measure designed to measure change over time or change in response to an intervention. The GMFM initially comprised 85 items; the measure later consisted of 88 items to enable evaluation of some items bilaterally and was subsequently referred to as the GMFM-88.[1] Russell and colleagues[1] demonstrated intra-rater and inter-rater reliability intraclass correlation coefficients (ICC) of 0.99 and an ICC of 0.99 for test-retest reliability of the GMFM-88. Drouin and collegues[2] supported the construct validity of the GMFM by reporting significant linear relationships between gait velocity and dimensions D (standing) (r=0.91) and E (walking, running, jumping) (r=0.93) of the GMFM-88. Furthermore, Damiano and Abel[3] found a strong correlation between computerized gait analysis parameters and GMFM-88 scores. This independently confirmed the construct validity of the GMFM-88.

Following extensive use of the GMFM-88 in both clinical practice and research, efforts were put into improving the scaling of the measure.[4,5] Rasch analysis is a statistical technique that can be employed to allow for interval-level scores to be created and used from ordinal level measures.[4] Rasch analysis is based on a probabilistic model that uses maximum likelihood estimation to order items along a difficulty continuum. Rasch analysis was applied to the GMFM-88 to create the GMFM-66 and allowed for the hierarchical structure of the items to be revealed.[4] A computer program was created (the Gross Motor Ability Estimator (GMAE)) and is necessary for clinicians and researchers to compute GMFM-66 scores.[4] The GMFM-66 was shown to be **highly** reliable (ICCs ranging from 0.97-0.99)[5,6] and sensitive-to-change.[5,6] One research study subsequently

supported the greater sensitivity-to-change of the GMFM-66 compared to the GMFM-88 as determined through receiver operating characteristic curve analysis.[7]

Avery and colleagues[4] demonstrated that as few as 13 items would be needed to provide an accurate estimation of a child's gross motor abilities, and although clinicians and researchers were eager to have a shorter version, no guidelines existed for choosing appropriate subsets of items in the public domain at the onset of this project. As a result, two independent shortened versions of the GMFM-66 were created concurrently. The rationale behind both shortened versions is the same: shorter tests allow for the elimination of items not considered clinically relevant to the individual child, therefore, only items around the child's current ability are tested. The hierarchical ordering of items obtained from the Rasch analysis to create the GMFM-66, informed the creation of both the abbreviated versions.

One of the abbreviated methods consists of item sets (GMFM-66-IS) and was created through the use of an algorithm. This version is administered by assessing the child on three GMFM-66 "decision" items; performance on those items ultimately dictates one of four items sets to be administered and scored.[8] Creation and validation of the item sets and application of the algorithm are described elsewhere.[8] The GMFM-66-IS was applied to an existing data set to estimate the validity of this shortened version. An ICC was used to confirm that the agreement between scores on the GMFM-66-IS and the GMFM-66 was high (ICC=0.994) at a single point in time.[8]

Prior to completion of the algorithm approach, the second author (DJB) required an abbreviated version for a nationally-funded project using methods of comprehensive outcomes research in rehabilitation[9,10] and as a result a second shortened version was

created. This alternate shortened version uses a basal and ceiling approach (GMFM-66-B&C) in which the user tests only items around a given child's ability. Accordingly, a new data collection sheet was developed, placing the items in difficulty order as established by the Rasch analysis[4] and using age and Gross Motor Function Classification System (GMFCS) as guides on where to commence administering items (Appendix). Guidelines for age and Gross Motor Function Classification System (GMFCS) level[11] were extrapolated from the gross motor function curves[12] based on the average GMFM-66 score as related to the item map by difficulty.[1] The basal and ceiling approach is commonly used in developmental testing (e.g. as for the Peabody Developmental Motor Scales[13]). A basal score of 3 consecutive 3s (completes) must be obtained as the start of the test. The ceiling score is reached when the child scores 3 consecutive 0s (does not initiate). For the test to be complete, there must be at least 15 items tested in between (or outside of) the basal and ceiling scores. In this approach, the items are ordered by difficulty which removes them from the dimensional approach previously used in the GMFM-88 and GMFM-66. Pilot testing using existing data (n=50) has also demonstrated strong agreement between scores on the GMFM-66-B&C and the full GMFM-66 with ICC=0.99 (95% CI 0.98-0.99) (Bartlett, unpublished data). A limitation of the previously reported results of both of the abbreviated versions is that all analyses were conducted retrospectively on existing data sets. Whether or not these abbreviated versions perform well in a real practice setting has not yet been determined.

The purposes of this study were to estimate the concurrent validity of the two abbreviated versions with the criterion standard (the GMFM-66) and to estimate both the comparability and the test-retest reliability of the two abbreviated versions.

Methods

This was a measurement study comprising validity, comparability and reliability components of the two shortened versions of the GMFM-66 that were described in the introduction. This study was approved by the Ethics Board at The University of Western Ontario, McMaster University and the Thames Valley Children's Centre prior to data collection.

The children who participated in the study were recruited from 4 sites located in Southwestern Ontario, Canada. Children with a primary diagnosis of CP, between the ages of 2 to 6 years, were included in the study. The lower age range was chosen on the basis of previously established reliability of the GMFCS[11] which was greater after 2 years of age than under this age cut point. The upper age range was chosen based on the understanding that after 7 or 8 years of age, motor development plateaus in children with CP.[11] Recent data also suggested that a decline in motor abilities may be exhibited as early as 6.9 years of age.[14] An attempt was made to have representation across all levels of the GMFCS to be able to generalize findings back to the population of children with CP as a whole. Additionally, other information on the child was gathered including: the child's age and gender as well as limb distribution (note: type of motor disorder was not collected in this study due to issues with respect to lack of reliability[15]). This convenience sample comprised twenty-six participants. An ideal target sample size based on the expected ICC values was 40 participants; [16] however, challenges in recruitment limited the sample size. Table 1 contains participant characteristics including age, gender, distribution of involvement and GMFCS level.

Measures

Three measures were used in this study, the full GMFM-66[1] (score sheet available on the CanChild website www.canchild.ca) the GMFM-66-IS[8] (score sheets available as an online appendix[8]) and the GMFM-66-B&C (score sheet available as an appendix to this paper). All measures were scored by LKB using the GMAE software.

Each child was tested on **two** occasions, the first session contained assessments of both shortened versions (one version was randomly allocated to a therapist assessor and the other to the lead author LKB. LKB always began the assessment, thus the shortened versions were applied in a random order between subjects) followed by administration of the full GMFM-66 (administered by LKB) to establish concurrent validity. The full GMFM-66 was administered last as it was the longest assessment. Prior to its administration the children received a rest break if needed. Additionally, this order required a minimal amount of time from the participating therapist and still ensured random ordering of the two abbreviated versions. The second session was conducted 2 weeks after the first session (a timeframe in which no change was expected) and included assessments of both of the abbreviated versions (administered by the same raters and the same order as in time 1) to establish comparability of the two abbreviated versions and test-retest reliability.

All therapist assessors received a training booklet and participated in a one-hour teleconference in which they were introduced to the study protocol. Before commencing the study, the investigator and each physical therapist (PT) passed a criterion test in scoring selected videotaped items of the GMFM-66. Each rater obtained ≥ 80% item agreement on this test prior to collecting data in the study. The 9 therapists ranged from 2-36 years of experience in pediatric physical therapy practice with a mean score of 15

years and all therapists were familiar with the GMFM-66, using it at least occasionally in their practice prior to participating in this study. None were familiar with either abbreviated version. Each therapist referred clients from their caseload and as a result the number of children that each therapist tested varied; one therapist assessed only one child, and the other therapists assessed between 2 and 5 children each.

Concurrent validity, comparability and test-retest reliability were examined using an intraclass correlation coefficient (ICC 2,1). Time to completion was recorded for both shortened versions. In addition, we had PTs indicate their preference for one or the other abbreviated versions after their final assessment with the study; this was measured on a 5-point Likert scale with the two ends representing strong preference of the GMFM-66-IS (a score of 1) or the GMFM-66-B&C (a score of 5) with the middle being a neutral opinion. Seven out of the 9 therapists had exposure to both versions (one therapist only tested one child and a second therapist who tested two children had the same version randomly allocated both times), and only their preferences were recorded.

Results

Score means and standard deviations plus the number of items administered and time to completion at times 1 and 2 can be found in Table 2. On average the GMFM-66-B&C involved testing 15 and 16 fewer items compared to the GMFM-66-IS at times 1 and 2, respectively. This was to be expected as the item sets have a predetermined number of items ranging from 15 to 39 items. Time to completion was not significantly different between versions at either of the time points. A two-way ANOVA revealed that there were no significant effects of time ($F=1.00$, $df=1$, $p=0.32$), version ($F=2.99$, $df=1$, $p=0.09$), or time by version interaction ($F=0.26$, $df=1$, $p=0.61$). Table 3 contains the

concurrent validity and test-retest coefficients for both shortened versions, as well as comparability between the two abbreviated versions for the two time points. In addition, the standard error of measurement and the minimal detectable change (at the 95% level) are provided. Figures 1 and 2 visually demonstrate the concurrent validity between the GMFM-66 and both of the abbreviated versions, including the confidence and prediction bands. Six out of the 7 therapists who had experience with both measures preferred the GMFM-66-B&C version **to** the GMFM-66-IS with a median value on the Likert scale of 5.

Discussion

The results of this study illustrate that both abbreviated versions of the GMFM-66 demonstrated high levels of validity and reliability and can be used in clinical practice or research endeavors. According to Portney and Watkins,[17] ICC coefficients above 0.75 indicate good reliability. This study demonstrated excellent validity and reliability with ICC coefficients above 0.98 for concurrent validity and all indices of reliability, on both shortened versions.

All but one of the therapists in the study strongly preferred the GMFM-66-B&C, with most anecdotally citing the second decision item of the GMFM-66-IS as a problem. It appeared that Item 67 (stand 2 arms held: walks forward 10 steps) was problematic as a decision item because most children in GMFCS levels I to IV, regardless of age, completed this item. This finding is partially due to the manner in which PTs facilitated walking in these children in the context of this study. They were observed to facilitate either from in front of or behind the child, despite the guidelines for administration which clearly state that the assessor should be in front of the child “to reduce the inclination to

facilitate walking"[1] (p. 108). From a clinical perspective, even though training emphasizes adherence to the administration and scoring guidelines, therapists may continue to facilitate the children in this manner and as a result some children are being evaluated with item set 3 although they probably should be assessed with item set 2. The choice of another item around the same difficulty level may prevent this misclassification.

Additionally, the misclassification of children in the study could explain why no significant differences were seen between time to completion of the different versions. Many therapists indicated frustration with using item set 3 as there were few items that the children could complete successfully (score of 3), and few items that the children were able to partially complete (score of 1 or 2) and many items the child could not perform or even attain the starting position of (score of 0). This could have influenced the time to completion results as even though 39 items were "tested" many of these items were not attempted as a result of the child being unable to attain the starting position (for example: a child classified as level IV could not maintain their body weight in standing and therefore items 53-58 inclusive were not attempted and still scored as a 0).

Therapists who preferred the GMFM-66-B&C also revealed their thoughts that this version usually contained items that were more clinically relevant to the individual child they were assessing when compared to the GMFM-66-IS as a result of the misclassification. Related to this, many other developmental scales employ the basal and ceiling approach and clinicians are familiar and skilled in using this method. The one therapist who preferred the GMFM-66-IS cited its ease of administration as the items are

still ordered by number as in the full GMFM-66 and items in similar starting positions remain grouped together.

The lead author of this paper also preferred to use the GMFM-66-B&C. The different item sets often contained items representing milestones for which a child had already attained and subsequently obtained a more advanced but related milestone and it became difficult to score the easier item due to lack of cooperation or interest from the child. For example, item set 3 contained items 65 and 66 related to cruising. When a child has the ability to walk independently these items are very hard to elicit due to lack of interest). The results of this study show that the GMFM-66-B&C is the preferred shortened version of the GMFM-66. This version requires on average 20-25 minutes and represents a large reduction from the estimate of 45-60 minutes given for the GMFM-66[1]. The use of the GMFM-66-B&C can therefore reduce the burden of assessment time for both children and therapists while still providing a valid and reliable estimate of gross motor function in both clinical and research settings.

The smaller than expected sample size is a limitation of this study, however, the high estimates of concurrent validity and reliability, along with tight 95% confidence intervals obtained, suggest that the sample size was adequate. Additionally, there was an uneven gender distribution in this study; however, there is no evidence that gender influences GMFM-66 scores, therefore this is not a significant limitation to this study. A relative disadvantage of the GMFM-66-B&C is that the process of entering scores into the GMAE software (conducted by LKB in this study) was more cumbersome due to the nature of the score sheet, in that it is ordered by difficulty and not in the order that the GMAE software presents for scores to be entered. Although there are columns indicating

which dimension the items are from (potentially assisting with the process of entering scores if one is using the GMAE) the process is slower than that of entering the original GMFM-66 scores. There is syntax developed by the authors available for use by researchers to simplify the scoring. This syntax converts the items so that they can be scored using the GMAE and converted for subsequent use in SPSS (this syntax can be found in a supplemental file to this manuscript and is also posted on the *CanChild* website at www.canchild.ca). Some attention needs to be paid to entering scores for the GMFM-66-IS as well because the item sets are comprised of items from different dimensions and, although they are arranged in numerical order, not all items are present in the item sets.

At this time, both versions were shown to have high levels of validity and reliability for use in clinical or research settings, however, the therapists in this study preferred the GMFM-66-B&C **perceiving it as more clinically relevant to their clients**.

Acknowledgements:

This research was supported through the Canadian Institutes of Health Research (MOP-81107). We also thank Barbara Stoskopf for all her efforts related to this project and Dianne Russell for her input at the design stage of this project, and for a review of an earlier manuscript.

References

1. Russell DJ, Rosenbaum PL, Avery LM, Lane M. Gross Motor Function Measure (GMFM-66 & GMFM-88) User's Manual. London: Mac Keith Press; 2002.

2. Drouin LM, Malouin F, Richards CL, Marcoux S. Correlation between the gross motor function measure scores and gait spatiotemporal measures in children with neurological impairments. *Dev Med Child Neurol.* 1996;38:1007-1019.

3. Damiano DL, Abel MF. Relation of gait analysis to gross motor function in cerebral palsy. *Dev Med Child Neurol*. 1996;38:389-396.

4. Avery LM, Russell DJ, Raina PS, Walter SD, Rosenbaum PL. Rasch analysis of the gross motor function measure: Validating the assumptions of the rasch model to create an interval-level measure. *Arch Phys Med Rehabil.* 2003;84:697-705.

5. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. Improved scaling of the gross motor function measure for children with cerebral palsy: Evidence of reliability and validity. *Phys Ther*. 2000;80:873-885.

6. Shi W, Wang SJ, Liao YG, Yang H, Xu XJ, Shao XM. Reliability and validity of the GMFM-66 in 0- to 3-year old children with cerebral palsy. *Am J Phys Med Rehabil*. 2006;85:141-147.

7. Wang HY, Yang YH. Evaluating the responsiveness of 2 versions of the gross motor function measure for children with cerebral palsy. *Arch Phys Med Rehabil*. 2006;87:51-56.

8. Russell D, Avery L, Walter S, Hanna S, Bartlett D, Rosenbaum P, Palisano R. Development and validation of item sets for the GMFM-66 to improve efficiency of administration in children with cerebral palsy. *Dev Med Child Neurol.* 2009;51:89.

9. Bartlett DJ, Lucy SD. A comprehensive approach to outcomes research in rehabilitation. *Physiother Can*. 2004;56:237-247.

10. Bartlett DJ, Chiarello LA, McCoy SW, Palisano RJ, Rosenbaum PL, Jeffries L, LaForme Fiss A, Tieman B, Stoskopf B. The measurement model of the Move & PLAY study: The protocol of a study using comprehensive outcomes research in rehabilitation. Accepted by *Physical Therapy* May 31, 2010.

11. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol*. 1997;39:214-223.

12. Rosenbaum PL, Walter SD, Hanna SE, Palisano RJ, Russell DJ, Raina P, et al. Prognosis for gross motor function in cerebral palsy: Creation of motor development curves. *JAMA.* 2002;288:1357-1363.

13. Folio MR, Fewell RR. Peabody Developmental Motor Scales (2nd Ed). Austin, Texas: Pro-Ed; 2000.

14. Hanna SE, Rosenbaum PL, Bartlett DJ, Palisano RJ, Walter SD, Avery L, et al. Stability and decline in gross motor function among children and youth with cerebral palsy aged 2 to 21 years. *Dev Med Child Neurol.* 2009; 51:295-302.

15. Gorter JW, Rosenbaum PL, Hanna SE, Palisano RJ, Bartlett DJ, Russell DJ, et al. Limb distribution, motor impairment, and functional classification of cerebral palsy. *Dev Med Child Neurol*. 2004;46:461-467.

16. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101-110.

17. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Upper Saddle River, New Jersey: Prentice Hall Health; 2000.

Table 1 - Participant Characteristics

Characteristic		**Total (n=26)**	
		n	**%**
Gender	Male	18	69.2
	Female	8	30.8
Age	Mean years (SD)	4.1 (1.2)	
	Median	3.9	
	Minimum-Maximum	(1.9-6.4)	
GMFCS Level	I	5	19.2
	II	4	15.4
	III	6	23.1
	IV	7	26.9
	V	4	15.4
Distribution of Involvement	Hemiplegia	5	19.2
	Diplegia	8	30.8
	Triplegia	2	7.7
	Quadriplegia	11	42.3

SD = standard deviation; GMFCS = Gross Motor Function Classification System

Table 2 - GMFM-66 Scores, Number of Items Tested and Time to Completion

Version	**Time 1**	**Time 2**
GMFM-66 (Calculated by GMAE)		
Mean Score (SD)	48.4 (16.7)	
Number of Items Mean (SD)	65.7 (1.1)	
GMFM-66-IS		
Mean Score (SD)	48.1 (16.2)	48.1 (16.0)
Number of Items Mean (SD)	32.9 (8.6)	33.1 (7.8)
Time to Completion in Minutes		
Mean (SD)	29.2 (13.9)	27.8 (15.5)
GMFM-66-B&C		
Mean Score (SD)	48.7 (17.1)	49.2 (17.2)
Number of Items Mean (SD)	17.2 (3.7)	16.6 (2.0)
Time to Completion in Minutes		
Mean (SD)	26.0 (9.3)	21.1 (7.8)

GMFM = Gross Motor Function Measure; GMAE = Gross Motor Ability Estimator, SD= Standard Deviation, IS=Item Set Approach, B&C= Basal and Ceiling Approach

Table 3 – Psychometric Properties of the GMFM-66-IS and the GMFM-66-B&C

Statistic	Estimate	95% CI
Concurrent Validity		
GMFM-66-IS	0.994	0.987-0.997
GMFM-66-B&C	0.987	0.972-0.994
Test-retest Reliability		
GMFM-66-IS	0.986	0.969-0.994
GMFM-66-B&C	0.994	0.987-0.997
Comparability Between Abbreviated Versions		
Time 1	0.984	0.965-0.993
Time 2	0.970	0.932-0.986
Standard Error of Measurement (SEM)		
GMFM-66-IS	1.91	N/A
GMFM-66-B&C	1.31	N/A
Minimal Detectable Change (95%)		
GMFM-66-IS	5.29	N/A
GMFM-66-B&C	3.63	N/A

GMFM = Gross Motor Function Measure, IS = Item Set Approach, B&C = Basal and Ceiling Approach

Figure 1

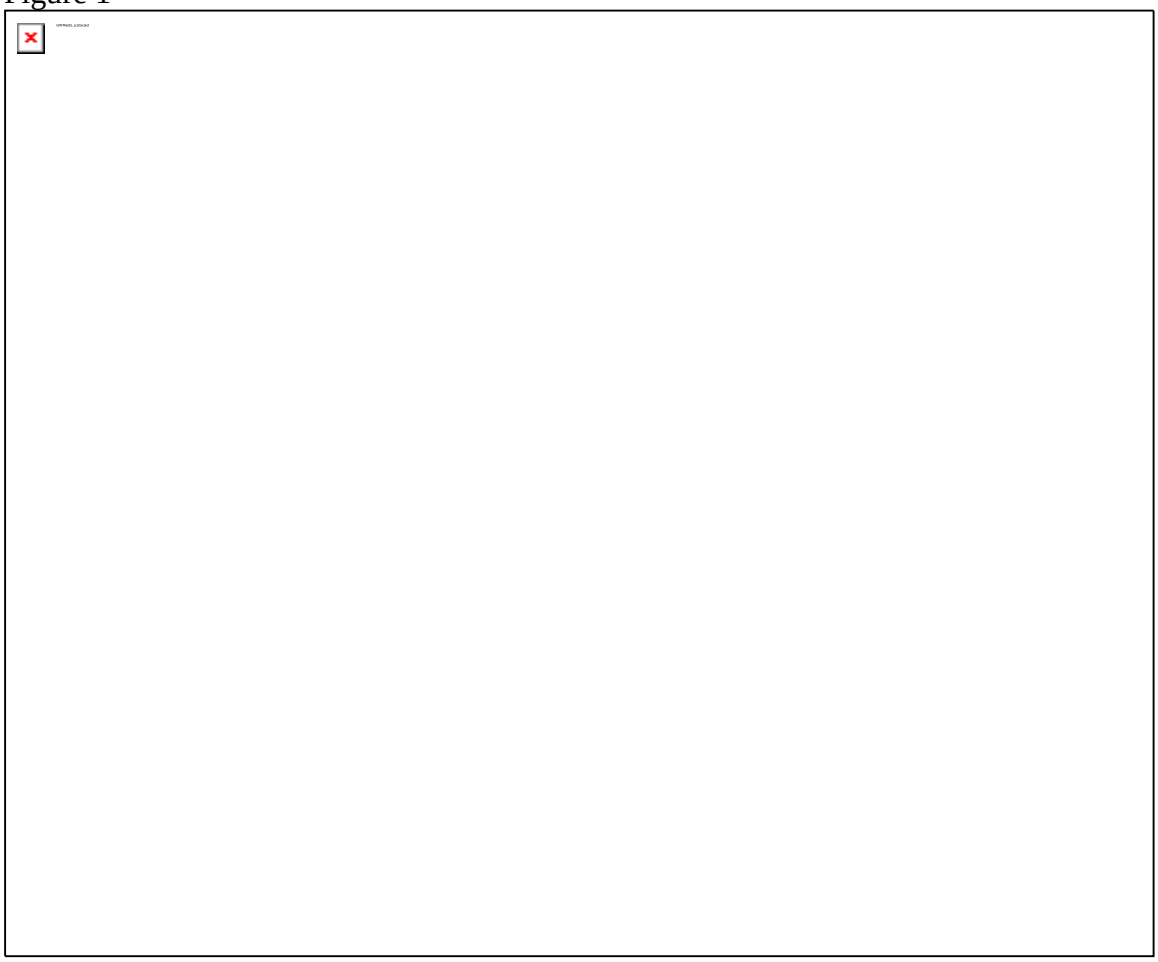

Figure 2