Arbeidsnotat nr. 195 / 2012
Working paper no. 195 / 2012

The Pediatric Evaluation of Disability Inventory (PEDI) – Differences between the US and the Norwegian Normative Samples

Jo Kleiven, Marie Berg, Lena Krumlinde-Sundholm
og Anne-Stine Dolva

Høgskolen i Lillehammer
Lillehammer University College • hil.no

ISSN 0806-8348

Kr. 150,-

The Pediatric Evaluation of Disability Inventory (PEDI) – Differences between the US and the Norwegian Normative Samples

Jo Kleiven
Lillehammer University College, Lillehammer, Norway

Marie Berg
Sunnaas Rehabilitation Hospital, University of Oslo, Norway

Lena Krumlinde-Sundholm
Department of Woman's and Children's Health
Karolinska Institutet, Stockholm, Sweden

Anne-Stine Dolva
Lillehammer University College, Lillehammer, Norway

Sammendrag: The Pediatric Evaluation of Disability Inventory (PEDI) er en standardisert prosedyre for vurdering av barns ferdigheter som ble utviklet i USA. Den er mye brukt ved rehabilitering av barn, også i Norge. Et representativt utvalg av norske barn i ti aldersgrupper ble sammenlignet med det originale amerikanske standardiseringsutvalget. På svært mange ulike punkter viste det seg at norske og amerikanske barn er ulike. Ved vurdering av norske barn bør de derfor sammenlignes med det norske utvalget, ikke med de amerikanske normene for PEDI-testen.

Emneord: PEDI, testing, barns ferdighetsnivå

Summary: The Pediatric Evaluation of Disability Inventory (PEDI) is a standardized procedure for assessing childrens' abilities, developed in the USA. It is widely used in child rehabilitation, also in Norway. A representative sample of Norwegian children in ten age groups is compared to the original American standardizing sample. On a large number of points, Norwegian and American children were shown to be different. In the assessment of Norwegian children, therefore, they should be compared to the Norwegian sample, not to the American norms for the PEDI test.

Key words: PEDI, testing, ability levels of children

Contents:

Innhold

Contents: 3
List of figures: 5
List of tables: 7
Introduction 9
1. Self-care function 11
1.1 Self-care domain 11
A. Raw scores 11
B. Scaled scores 12
1.2 Self-care subdomains 15
A. Types of Food Textures (4 items) 15
B. Use of Utensils (5 items) 16
C. Use of Drinking Containers (5 items) 17
D. Toothbrushing (5 items) 18
E. Hairbrushing (4 items) 19
F. Nose Care (5 items) 20
G. Handwashing (5 items) 21
H. Washing Body and Face (5 items) 22
I. Pullover/Front-Opening Garments (5 items) 23
J. Fasteners (5 items) 24
K. Pants (5 items) 25
L. Shoes/socks (5 items) 26
M. Toileting Tasks (5 items) 27
N. Management of Bladder (5 items) 28
O. Management of Bowel (5 items) 29
1.3 Self-care items 30
A. All-over sample differences in Self-care capability 30
B. Age and sample differences on Self-care single items 32
C. Sample differences of item difficulty 35
D. DIF-tests of Self-Care item difficulty differences 36
2. Mobility function 38
2.1 Mobility domain 38
2.2 Mobility subdomains 39
A. Toilet Transfers (5 items) 39
B. Chair/Wheelchair Transfers (5 items) 40
C. Car Transfers (5 items) 41
D. Bed Mobility/Transfers (4 items) 42
E. Tub Transfers (5 items) 43
F. Indoor Locomotion Methods (3 items) 44
G. Indoor Locomotion – Distance/speed (5 items) 45
H. Indoor Locomotion – Pulls /Carries Objects (5 items) 46
I. Outdoor Locomotion – Methods (2 items) 47
J. Outdoor Locomotion – Distance /Speed (5 items) 48
K. Outdoor Surfaces (5 items) 49
L. Up Stairs (5 items) 50
M. Down Stairs (5 items) 51

2.3 Mobility items 52
A. All-over sample differences in Mobility capability 52
B. Age and sample differences on Mobility single items 54
C. Sample differences of item difficulty 56
D. DIF-tests of Mobility item difficulty differences 57
3. Social function 59
3.1 Social function domain 59
3.2 Social subdomains 60
A. Comprehension of Word Meanings (5 items) 60
B. Comprehension of Sentence Complexity (5 items) 61
C. Functional Use of Communication (5 items) 62
D. Complexity of Expressive Communication (5 items) 63
E. Problem Resolution (5 items) 64
F. Social Interactive Play (5 items) 65
G. Peer Interaction (5 items) 66
H. Play with Objects (5 items) 67
I. Self -Information (5 items) 68
J. Time Orientation (5 items) 69
K. Household Chores (5 items) 70
L. Self-Protection (5 items) 71
M. Community Function (5 items) 72
3.3 Social Function items 73
A. All-over sample differences in Social Function capability 73
B. Age and sample differences on single Social Function items 74
C. Sample differences of item difficulty 76
D. DIF-tests of Social Function item difficulty differences 77
4. Concluding comments 79
5. References 81

List of figures:

Figure 1: Age means for Self-care domain raw score in American and Norwegian samples 11
Figure 2: Age means for Self-care domain scaled score in American and Norwegian samples 13
Figure 3: Age means for *Food Textures* subdomain in US and Norwegian samples 15
Figure 4: Age means for *Use of Utensils* subdomain in US and Norwegian samples 16
Figure 5: Age means for *Use of Drinking Containers* subdomain in two samples 17
Figure 6: Age means for *Toothbrushing* subdomain in two samples 18
Figure 7: Age means for *Hairbrushing* subdomain in two samples 19
Figure 8: Age means for *Nose Care* subdomain in two samples 20
Figure 9: Age means for *Handwashing* subdomain in two samples 21
Figure 10: Age means for *Washing Body and Face* subdomain in two samples 22
Figure 11: Age means for *Pullover/Front-Opening* Garments subdomain in two samples 23
Figure 12: Age means for *Fasteners* subdomain in two samples 24
Figure 13: Age means for *Pants* subdomain in two samples 25
Figure 14: Age means for *Shoes/socks* subdomain in two samples 26
Figure 15: Age means for *Toileting Tasks* subdomain in two samples 27
Figure 16: Age means for *Management of Bladder* subdomain in two samples 28
Figure 17: Age means for *Management of Bowel* subdomain in two samples 29
Figure 18: Per cent mastering the five items (SC69-73) of the Management of Bowel subdomain, in two samples with ten age groups 32
Figure 19: Per cent mastering the five items (SC64-68) of the Management of Bladder subdomain, in two samples with ten age groups 34
Figure 20: Self-care item difficulty scores in the two normative samples 35
Figure 21: DIF analysis of Self-care items in the two normative samples 36
Figure 22: Age means for Mobility domain raw score in two samples 38
Figure 23: Age means for Toilet Transfer subdomain in two samples 39
Figure 24: Age means for Chair/Wheelchair Transfers subdomain in two samples 40
Figure 25: Age means for Car Transfers subdomain in two samples 41
Figure 26: Age means for Bed Mobility/Transfers subdomain in two samples 42
Figure 27: Age means for Tub Transfers subdomain in two samples 43
Figure 28: Age means for Indoor Locomotion Methods subdomain in two samples 44
Figure 29: Age means for Indoor Locomotion – Distance/speed subdomain in two samples 45
Figure 30: Age means for Indoor Locomotion – Pulls/Carries Objects subdomain in two samples 46
Figure 31: Age means for Outdoor Locomotion – Methods subdomain in two samples 47
Figure 32: Age means for Outdoor Locomotion – Distance/Speed subdomain in two samples 48
Figure 33: Age means for Outdoor Surfaces subdomain in two samples 49
Figure 34: Age means for Up Stairs subdomain in two samples 50
Figure 35: Age means for Down Stairs subdomain in two samples 51
Figure 36: Per cent mastering the four last items of the Indoor Locomotion – Pulls /Carries Objects subdomain, in two samples with ten age groups 55
Figure 37: Mobility item difficulty scores in the two normative samples 56
Figure 38: DIF analysis of Mobility items in the two normative samples 57
Figure 39: Age means for Social function domain raw score in two samples 59

Figure 40: Age means for Comprehension of Word Meanings subdomain in two samples 60
Figure 41: Age means for Comprehension of Sentence Complexity subdomain in two samples 61
Figure 42: Age means for Functional Use of Communication subdomain in two samples 62
Figure 43: Age means for Complexity of Expressive Communication subdomain in two samples 63
Figure 44: Age means for Problem Resolution subdomain in two samples 64
Figure 45: Age means for Social Interactive Play subdomain in two samples 65
Figure 46: Age means for Peer Interaction subdomain in two samples 66
Figure 47: Age means for Play with Objects subdomain in two samples 67
Figure 48: Age means for Self Information subdomain in two samples 68
Figure 49: Age means for Time Orientation subdomain in two samples 69
Figure 50: Age means for Household Chores subdomain in two samples 70
Figure 51: Age means for Self Protection subdomain in two samples 71
Figure 52: Age means for Community Function subdomain in two samples 72
Figure 53: Per cent mastering the five items of the *Time orientation* subdomain, in two samples with ten age groups 75
Figure 54: Social function item difficulty scores in the two normative samples 76
Figure 55: DIF analysis of Social Function items in the two normative samples 77

List of tables:

Table 1: ANOVA of Self-care raw score in two samples and ten age groups.....12
Table 2: ANOVA of Self-care scaled scores in two samples and ten age groups.....13
Table 3: Mixed-design ANOVA of ten (American) age groups with Self-care scaled scores based on 10 vs. 14 groups.....14
Table 4: ANOVA of *Food Textures* subdomain in two samples and ten age groups.....15
Table 5: ANOVA of *Use of Utensils* subdomain in two samples and ten age groups 16
Table 6: ANOVA of Use of *Drinking Containers* subdomain in two samples and ten age groups.....17
Table 7: ANOVA of *Toothbrushing* subdomain in two samples and ten age groups 18
Table 8: ANOVA of *Hairbrushing* subdomain in two samples and ten age groups 19
Table 9: ANOVA of *Nose Care* subdomain in two samples and ten age groups.....20
Table 10: ANOVA of *Handwashing* subdomain in two samples and ten age groups.....21
Table 11: ANOVA of Washing Body and Face subdomain in two samples and ten age..22
Table 12: ANOVA of *Pullover/Front-Opening* Garments subdomain in two samples and ten age groups 23
Table 13: ANOVA of *Fasteners* subdomain in two samples and ten age groups 24
Table 14: ANOVA of *Pants* subdomain in two samples and ten age groups.....25
Table 15: ANOVA of *Shoes/socks* subdomain in two samples and ten age groups.....26
Table 16: ANOVA of *Toileting Tasks* subdomain in two samples and ten age groups 27
Table 17: ANOVA of *Management of Bladder* subdomain in two samples and ten age groups.....28
Table 18: ANOVA of *Management of Bowel* subdomain in two samples and ten age groups.....29
Table 19: Self-care items; proportion mastering items in two samples (US=313, Norw=224).....30
Table 20: Self-care items with significant sample differences 37
Table 21: ANOVA of Mobility raw score in two samples and ten age groups.....38
Table 22: ANOVA of *Toilet Transfer* subdomain in two samples and ten age groups.....40
Table 23: ANOVA of *Chair/Wheelchair Transfers* Transfer subdomain in two samples and ten age groups 41
Table 24: ANOVA of *Car Transfers* subdomain in two samples and ten age groups 42
Table 25: ANOVA of *Bed Mobility/Transfers* subdomain in two samples and ten age groups.....43
Table 26: ANOVA of *Tub Transfers* subdomain in two samples and ten age groups 44
Table 27: ANOVA of *Indoor Locomotion Methods* subdomain in two samples and ten age groups 45
Table 28: ANOVA of *Indoor Locomotion – Distance/speed* subdomain in two samples and ten age groups 46
Table 29: ANOVA of *Indoor Locomotion – Pulls/Carries* Objects subdomain in two samples and ten age groups.....47
Table 30: ANOVA of *Outdoor Locomotion – Methods* subdomain in two samples and ten age groups 48

Table 31: ANOVA of Outdoor Locomotion – Distance/Speed subdomain in two samples and ten age groups 49
Table 32: ANOVA of *Outdoor Surfaces* subdomain in two samples and ten age groups. 50
Table 33: ANOVA of *Up Stairs* subdomain in two samples and ten age groups..... 51
Table 34: ANOVA of *Down Stairs* subdomain in two samples and ten age groups 52
Table 35: Mobility items; proportion mastering items in two samples (US=313, Norw=224) 52
Table 36: Mobility items with significant sample differences..... 58
Table 37: ANOVA of Social Function raw score in two samples and ten age groups..... 59
Table 38: ANOVA of Comprehension of Word Meanings subdomain in two samples and ten age groups 60
Table 39: ANOVA of *Comprehension of Sentence Complexity* subdomain in two samples and ten age groups 61
Table 40: ANOVA of Functional Use of Communication subdomain in two samples and ten age groups 62
Table 41: ANOVA of *Complexity of Expressive Communication* subdomain in two samples and ten age groups..... 63
Table 42: ANOVA of *Problem Resolution* subdomain in two samples and ten age groups 64
Table 43: ANOVA of *Social Interactive Play* subdomain in two samples and ten age groups..... 65
Table 44: ANOVA of *Peer Interaction* subdomain in two samples and ten age groups... 66
Table 45: ANOVA of *Play with Objects* subdomain in two samples and ten age groups 67
Table 46: ANOVA of *Self Information* subdomain in two samples and ten age groups... 68
Table 47: ANOVA of *Time Orientation* subdomain in two samples and ten age groups . 69
Table 48: ANOVA of *Household Chores* subdomain in two samples and ten age groups 70
Table 49: ANOVA of *Self Protection* subdomain in two samples and ten age groups..... 71
Table 50: ANOVA of *Community Function* subdomain in two samples and ten age groups..... 72
Table 51: Social Function items; proportion mastering items in two samples (US=313, Norw=224) 73
Table 52: Social Function items with significant sample differences 78

Introduction

During the last couple of years, a stimulating cooperation around the well-known PEDI inventory (S. M. Haley, Coster, Ludlow, Haltiwanger, & Andrellos, 1992) has taken much of our time. Utilizing Marie Berg's two samples (Berg, Frey Frøislie, & Hussain, 2003; Berg, Aamodt, Stanghelle, Krumlinde-Sundholm, & Hussain, 2008), we have made an effort to provide Norwegian norms for the PEDI.

A central question in this work has been the differences between American and Norwegian PEDI results. While certain cross-cultural problems have been documented (Berg et al., 2003; Berg et al., 2008), an even more precise knowledge of the differences may be needed to avoid mistakes in the clinical use of the PEDI in Norway. To allow exact and detailed comparisons with our data, professor Wendy Coster at Boston University has graciously provided access to the original American normative material.

A journal article in preparation (Berg, Dolva, Kleiven, & Krumlinde-Sundholm) will report the main results of the Norwegian norming project. All our discussions, arguments and conclusions will be provided there – not in the present informal paper. Also, the complete set of relevant norms will be available to Norwegian PEDI users at this web site of the Sunnaas Hospital (*http://www.sunnaas.no/aktuelt/rapporter/*).

But neither the journal article nor the final set of norms will provide a suitable home for a large number of detailed analyses that have become necessary in the process. To be sufficiently concise, just a few selected examples and simplifications could be used to support the conclusions of our publications. But, admittedly, "killing your darlings" was somewhat painful.

The need for brevity may perhaps also be unfortunate for more advanced PEDI users. Some experts may feel the need for a closer scrutiny of the basis of the norms, to ascertain that their diagnostic conclusions are sufficiently well founded.

We have decided, therefore, that some of our more comprehensive analyses should be made available, including rather detailed information on the differences between the American (US) and the Norwegian samples that form the bases for the two national norms. Although the arguments as well as the conclusions of the Norwegian PEDI project will be provided elsewhere, it is our opinion that the present analyses further support the need for Norwegian norms for the PEDI.

The PEDI research team at Boston University has recently developed a revised version of the PEDI named PEDI-CAT, which was just published (S.M. Haley, Coster, Dumas, Fragala-Pinkham, & Moed, 2012). The new PEDI-CAT is based on previous PEDI applications. This new instrument will, like the original PEDI, need both translation and Norwegian validation before it is applicable to Norwegian clinicians and researchers. In the meantime, there is not only a need for Norwegian norms on the original PEDI. Hopefully, our Norwegian norm analyses will also be useful to those who initiate translation and validation of the PEDI-CAT.

The organization of this paper assumes that the reader is somewhat familiar with the PEDI. Our exposition will be limited to the three functional skills areas: Self-care, Mobility and Social Function.

Analyses will be provided at the ***domain*** level (complete summed scores for 73, 59 or 65 items, respectively), the ***subdomain*** level (13-15 scores summing 2-5 related items), as well as the level of ***single items***. In addition, a few examples of single-item ***sample by age*** analyses will be given.

1. Self-care function

1.1 Self-care domain

A. Raw scores

With this scale (with max. score = 73), the curve of means for the US sample may suggest a ceiling effect. This is less pronounced in the Norwegian sample, however.

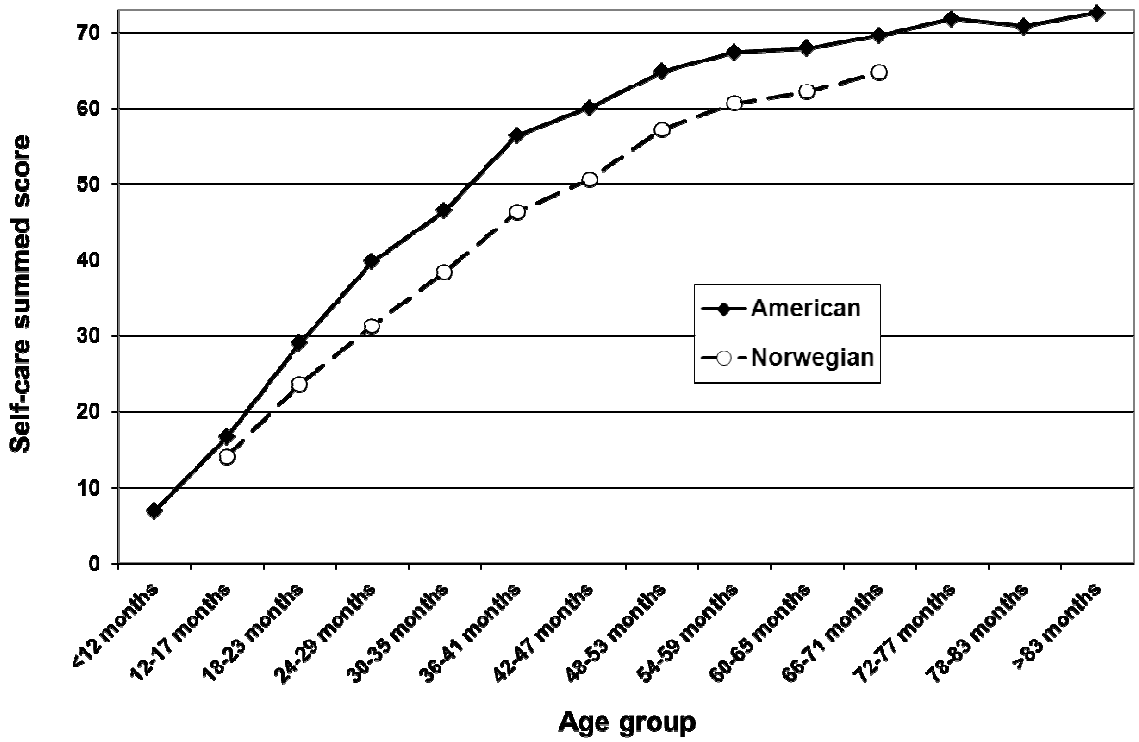

Figure 1: Age means for Self-care domain raw score in American and Norwegian samples

Limiting ourselves to the ten age groups employed in Norway, a two-way ANOVA was used to assess the differences between the two national samples as well as the differences between the age groups. The results (Table 1) indicate significant effects of *age group* as well as significant *national* differences. The non-significant interaction effect (Group by Nation) shows that the Nation and the Age differences are largely independent of each other; i.e., the national differences are rather consistent across all ten age groups.

Table 1: ANOVA of Self-care raw score in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:Self-Care summed raw score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	157944.630[a]	19	8312.875	187.447	.000
Intercept	1182441.278	1	1182441.278	26662.898	.000
group	139359.533	9	15484.393	349.158	.000
Nation	6036.281	1	6036.281	136.112	.000
group * Nation	557.073	9	61.897	1.396	.187
Error	22927.821	517	44.348		
Total	1466586.000	537			
Corrected Total	180872.451	536			

a. R Squared = .873 (Adjusted R Squared = .869)

B. Scaled scores

The raw scores of the PEDI, however, are normally not used in testing. Typically, the raw scores are used as a basis for computing 1-100 scaled "Rasch" scores (Bond & Fox, 2007), using the tables of the original PEDI manual (S. M. Haley et al., 1992) or suitable computer programs like *Winsteps* (Linacre, 2010). In addition, these scaled scores need to be transformed into T-scores (Mean = 50; SD = 10) to enable comparisons to the test norms.

While the *T* transformation is simple, linear and comprehensible, understanding the Rasch transformation is perhaps a less simple matter. It may be worth checking, therefore, if this transformation in any way influences the *nation* or *age* differences observed with the raw data.

A potential complication, however, is that the American and the Norwegian do not cover the same age span. The original American sample includes 14 half-year groups, starting at <12 months and peaking at >83 months. The Norwegian sample, however, is limited to 10 half-year groups, omitting the group <12 months as well as the groups >72 months.

To be on the safe side, therefore, it may be prudent to first compare the 10-group Norwegian Rasch scale scores to American scaled scores derived from the original 14-group sample used to compute the PEDI test norms. Secondly, the Norwegian scores should also be compared to their age-matched American counterparts. All three versions of scaled Rasch scores were obtained with the *Winsteps* program (Linacre, op.cit.), and results are displayed in figure 2.

First of all, figure 2 clearly shows that all three mean scores increase with age. It may also suggest some difference between the two American scores, possibly decreasing with age.

In addition, the Norwegian scores in all age groups are clearly lower than their American counterparts, closely resembling the difference shown in figure 1.

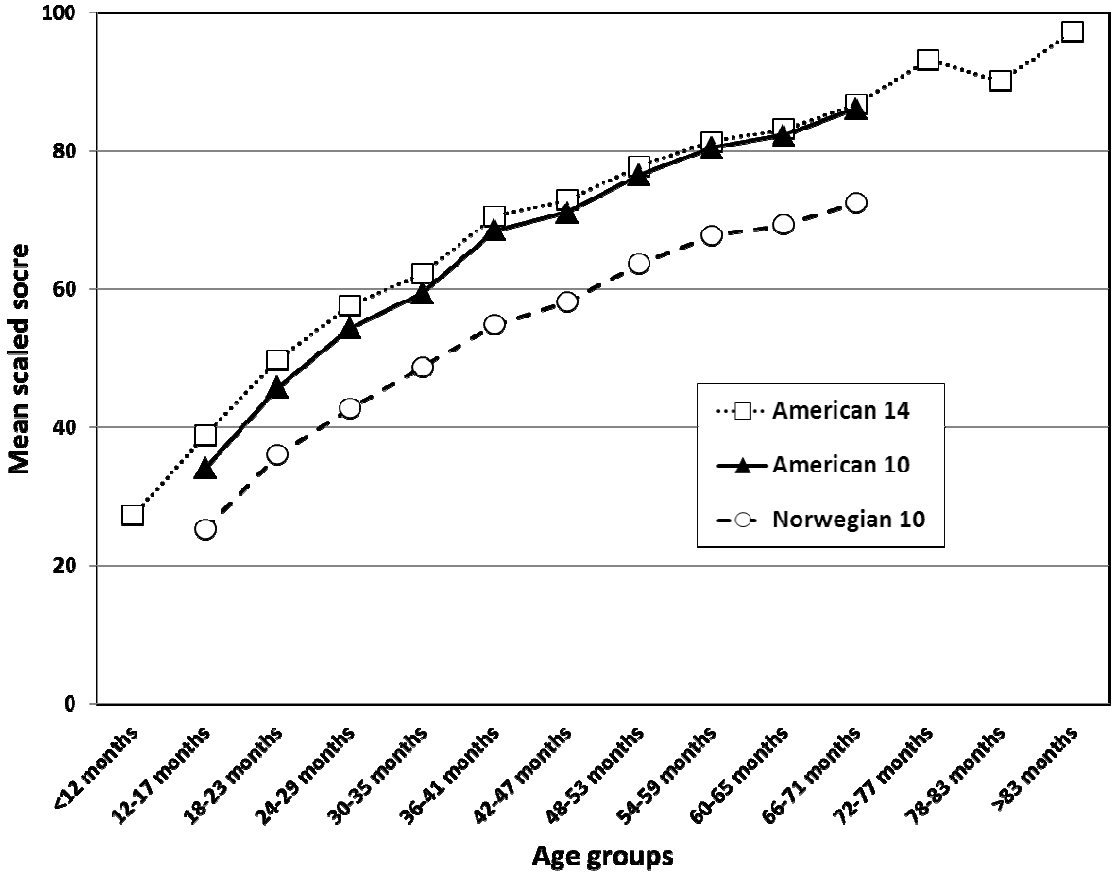

Figure 2: Age means for Self-care domain scaled score in American and Norwegian samples

Table 2: ANOVA of Self-care scaled scores in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:Rasch10SC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	140978.734[a]	19	7419.933	153.726	.000
Intercept	1812885.423	1	1812885.423	37559.306	.000
Nation	18061.324	1	18061.324	374.194	.000
group	116715.572	9	12968.397	268.679	.000
Nation * group	307.553	9	34.173	.708	.702
Error	24954.182	517	48.267		
Total	2154594.292	537			
Corrected Total	165932.916	536			

a. R Squared = ,850 (Adjusted R Squared = ,844)

The ANOVA results in table 2 show that with scaled scores based on 10 groups, there are significant differences between age groups as well as between nations.

Table 3: Mixed-design ANOVA of ten (American) age groups with Self-care scaled scores based on 10 vs. 14 groups

Tests of Within-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
10- vs.14-group scaled scores	764,70	1	764,70	4281,72	,000
10 vs. 14 factor * Age groups	298,84	9	33,20	185,92	,000
Error (10 vs. 14)	54,11	303	0,18		

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	2735803,51	1	2735803,51	25522,41	,000
Age groups	160151,22	9	17794,58	166,01	,000
Error	32479,23	303	107,19		

The powerful repeated-measurement ANOVA used to test the difference between the two scaled scores (based on 10 and 14 age groups, respectively), yield interesting results. First of all, the scores based on the full age range are generally significantly higher than the scores based on 10 groups only. Secondly, this difference significantly decreases with age. Thirdly – and perhaps less surprising – the differences between the age groups are also statistically significant.

All in all, therefore, fairly clear conclusions may be drawn. By using scaled self-care scores based on a more limited age range than what was used in the original PEDI, the difference between the American and the Norwegian samples may be somewhat overestimated. This effect, however, is obviously smaller than the differences in the raw data. It is to be expected, therefore, that Norwegian children should have lower PEDI self-care scores than the original American normative sample. Consequently, there may be cases where using the original PEDI norms with Norwegian children have lead to inaccurate results.

1.2 Self-care subdomains

A. Types of Food Textures (4 items)

Here, the Norwegian sample scores higher, mainly with the younger groups.

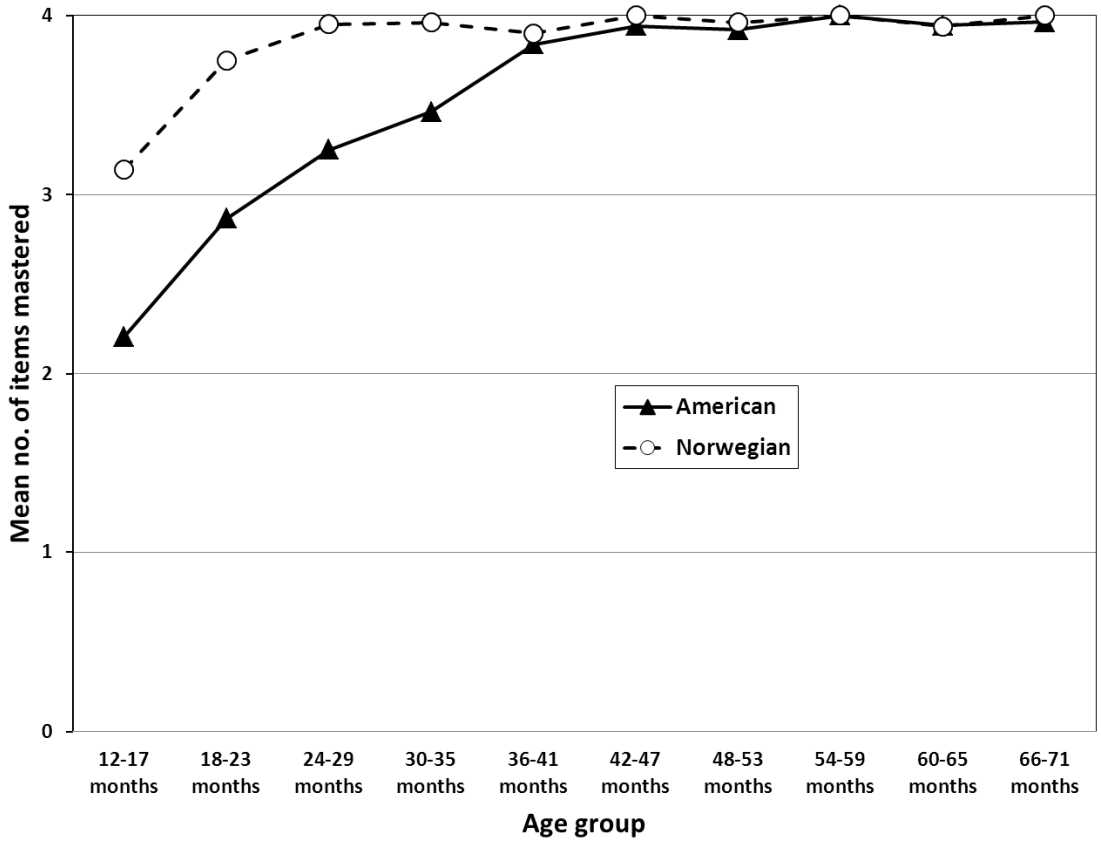

Figure 3: Age means for *Food Textures* subdomain in US and Norwegian samples

Table 4: ANOVA of *Food Textures* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	143.059[a]	19	7.529	22.092	.000
Intercept	6917.394	1	6917.394	20295.940	.000
Nation	12.970	1	12.970	38.054	.000
group	72.048	9	8.005	23.488	.000
Nation * group	16.930	9	1.881	5.519	.000
Error	176.207	517	.341		
Total	7495.000	537			
Corrected Total	319.266	536			

B. Use of Utensils (5 items)

No sample differences were found.

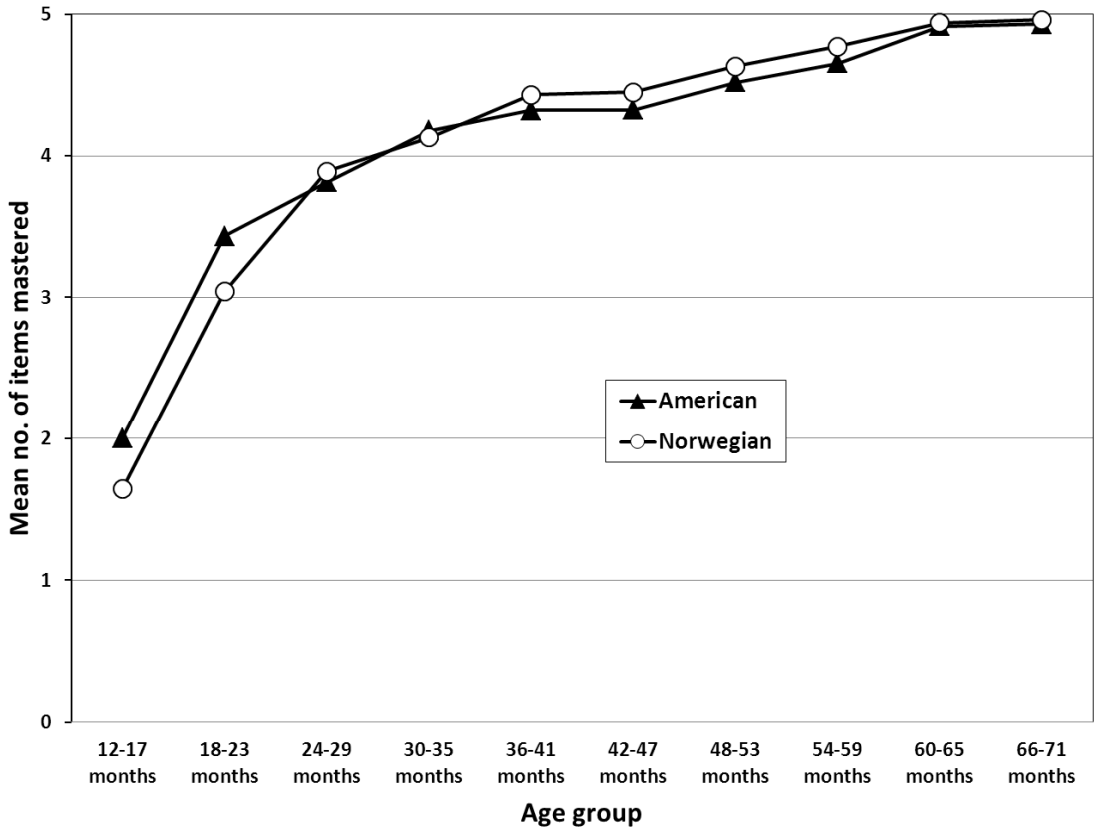

Figure 4: Age means for *Use of Utensils* subdomain in US and Norwegian samples

Table 5: ANOVA of *Use of Utensils* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCB

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	418.994[a]	19	22.052	55.594	.000
Intercept	8490.648	1	8490.648	21404.942	.000
Nation	.051	1	.051	.129	.720
group	373.248	9	41.472	104.551	.000
Nation * group	4.123	9	.458	1.155	.322
Error	205.077	517	.397		
Total	9711.000	537			
Corrected Total	624.071	536			

a. R Squared = .671 (Adjusted R Squared = .659)

C. Use of Drinking Containers (5 items)

No sample differences were found.

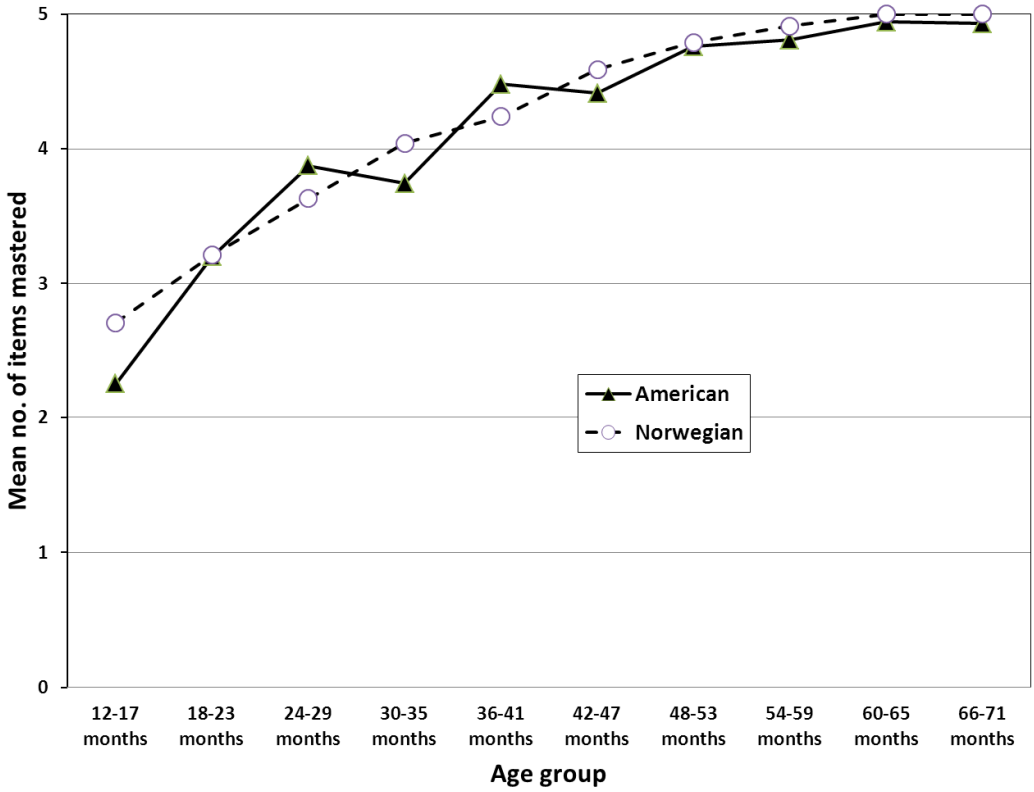

Figure 5: Age means for *Use of Drinking Containers* subdomain in two samples

Table 6: ANOVA of Use of *Drinking Containers* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	362.493[a]	19	19.079	44.615	.000
Intercept	8817.622	1	8817.622	20619.655	.000
Nation	.658	1	.658	1.539	.215
group	301.020	9	33.447	78.214	.000
Nation * group	4.837	9	.537	1.257	.258
Error	221.086	517	.428		
Total	9894.000	537			
Corrected Total	583.579	536			

a. R Squared = .621 (Adjusted R Squared = .607)

D. Toothbrushing (5 items)

The US sample scores
higher.

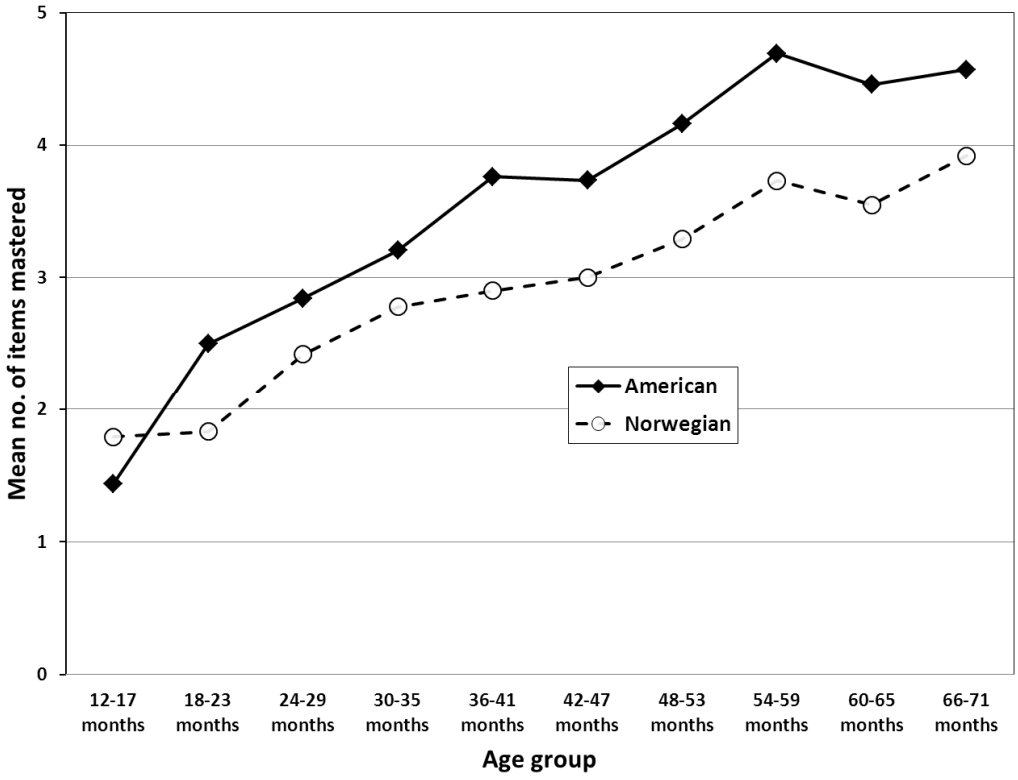

Figure 6: Age means for *Toothbrushing* subdomain in two samples

Table 7: ANOVA of *Toothbrushing* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCD

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	458.373[a]	19	24.125	38.957	.000
Intercept	5268.973	1	5268.973	8508.330	.000
Nation	47.787	1	47.787	77.166	.000
Group	339.481	9	37.720	60.910	.000
Nation * group	15.085	9	1.676	2.707	.004
Error	320.164	517	.619		
Total	6475.000	537			
Corrected Total	778.536	536			

A. R Squared = .589 (Adjusted R Squared = .574)

E. Hairbrushing (4 items)

The US sample scores higher.

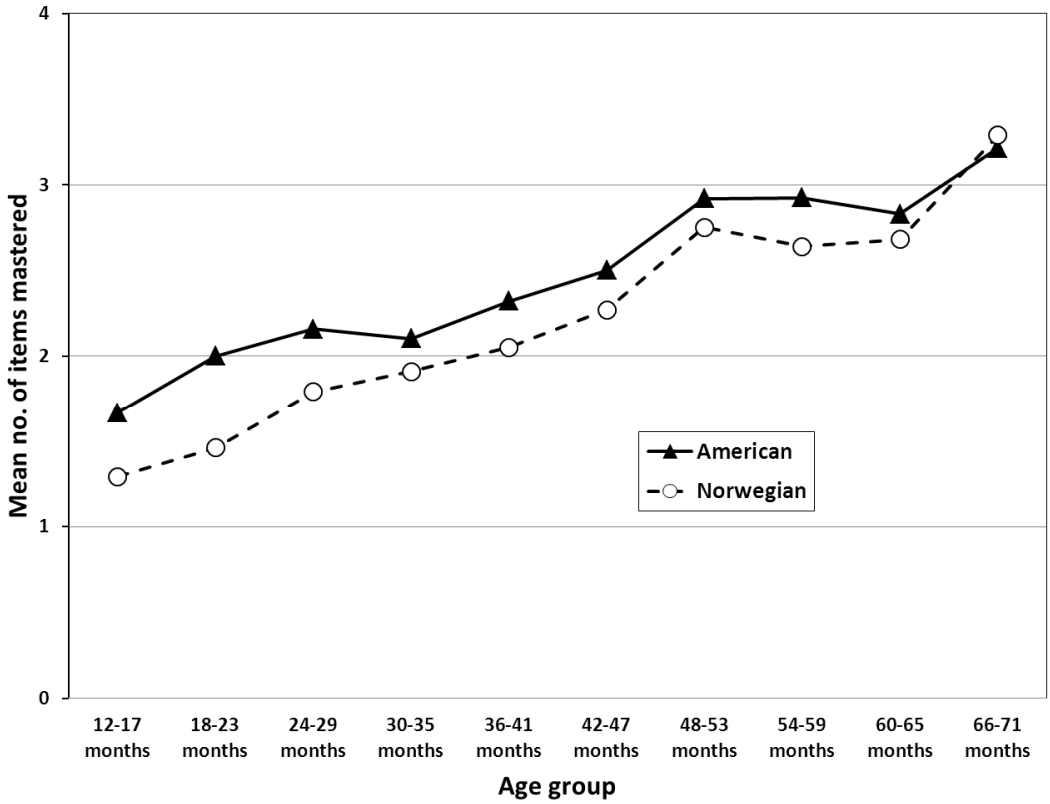

Figure 7: Age means for *Hairbrushing* subdomain in two samples

Table 8: ANOVA of *Hairbrushing* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	149.409[a]	19	7.864	19.714	.000
Intercept	2762.265	1	2762.265	6924.885	.000
Nation	7.955	1	7.955	19.943	.000
group	140.016	9	15.557	39.002	.000
Nation * group	3.175	9	.353	.885	.539
Error	206.226	517	.399		
Total	3345.000	537			
Corrected Total	355.635	536			

a. R Squared = .420 (Adjusted R Squared = .399)

F. Nose Care (5 items)

The US sample scores
higher.

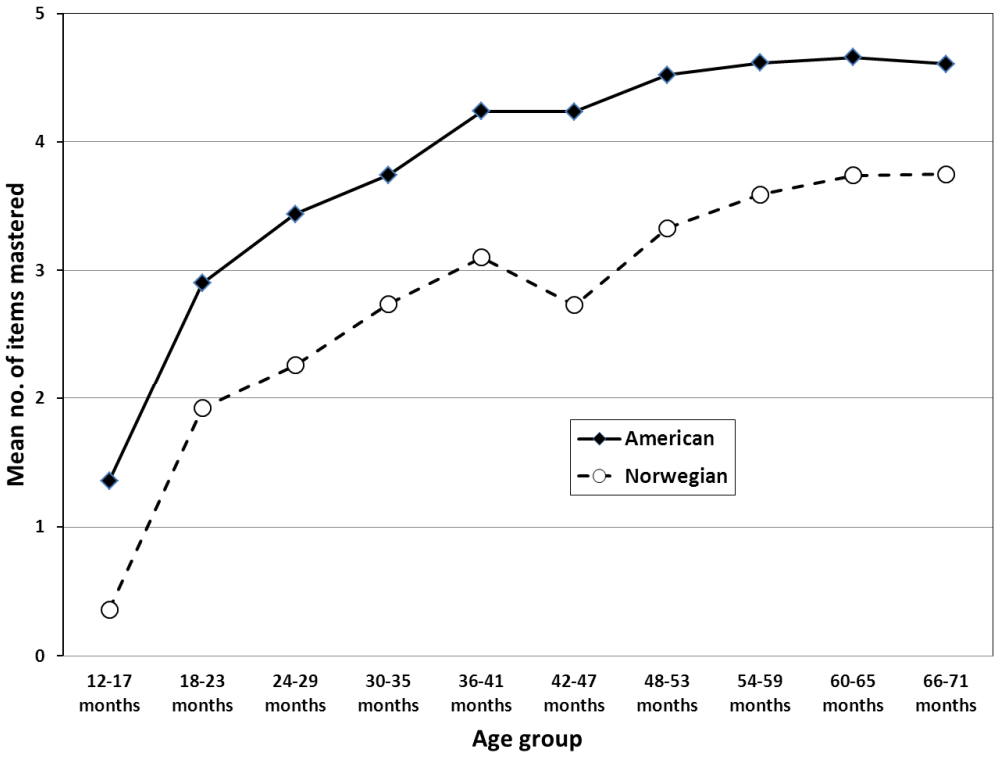

Figure 8: Age means for *Nose Care* subdomain in two samples

Table 9: ANOVA of *Nose Care* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCF

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	618.159[a]	19	32.535	36.014	.000
Intercept	5476.174	1	5476.174	6061.867	.000
Nation	147.400	1	147.400	163.165	.000
group	447.566	9	49.730	55.048	.000
Nation * group	4.133	9	.459	.508	.869
Error	467.048	517	.903		
Total	7240.000	537			
Corrected Total	1085.207	536			

a. R Squared = .570 (Adjusted R Squared = .554)

G. Handwashing (5 items)

The US sample scores higher.

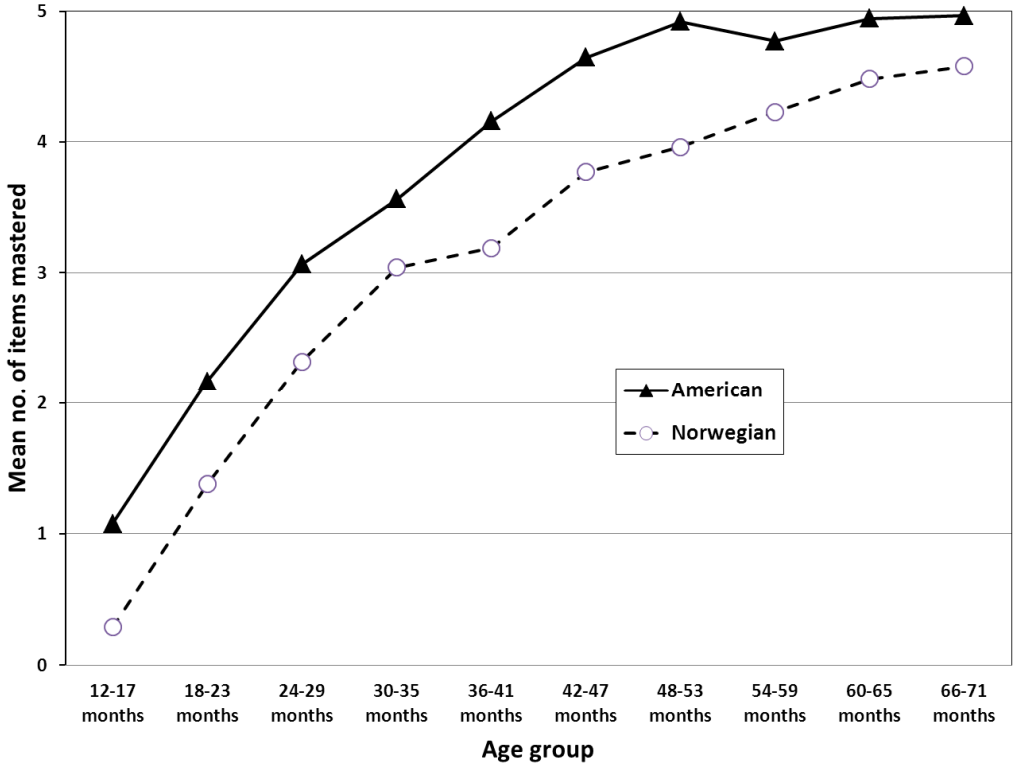

Figure 9: Age means for *Handwashing* subdomain in two samples

Table 10: ANOVA of *Handwashing* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	931.348[a]	19	49.018	69.266	.000
Intercept	6105.508	1	6105.508	8627.425	.000
Nation	62.587	1	62.587	88.439	.000
group	818.763	9	90.974	128.551	.000
Nation * group	5.319	9	.591	.835	.584
Error	365.874	517	.708		
Total	8041.000	537			
Corrected Total	1297.222	536			

a. R Squared = .718 (Adjusted R Squared = .708)

H. Washing Body and Face (5 items)

US sample scores higher, especially in the older age groups.

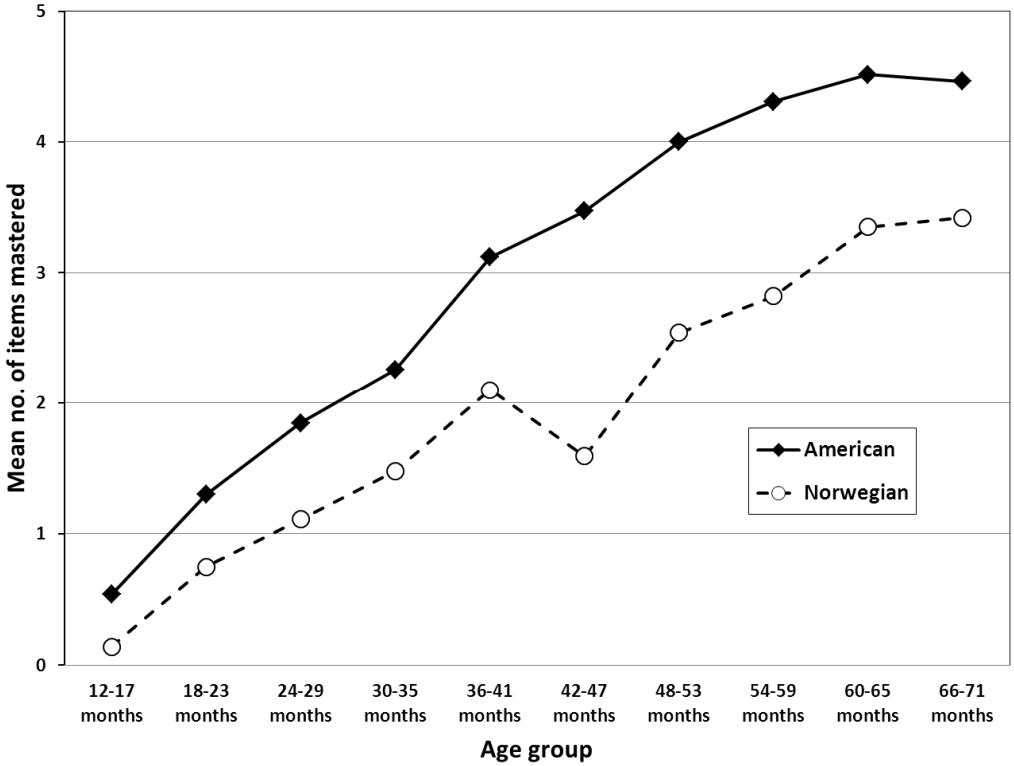

Figure 10: Age means for *Washing Body and Face* subdomain in two samples

Table 11: ANOVA of Washing Body and Face subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCH

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	920.578[a]	19	48.451	39.446	.000
Intercept	3047.617	1	3047.617	2481.152	.000
Nation	139.893	1	139.893	113.891	.000
group	707.392	9	78.599	63.990	.000
Nation * group	23.690	9	2.632	2.143	.025
Error	635.035	517	1.228		
Total	5005.000	537			
Corrected Total	1555.613	536			

a. R Squared = .592 (Adjusted R Squared = .577)

I. Pullover/Front-Opening Garments (5 items)

The US sample scores
higher.

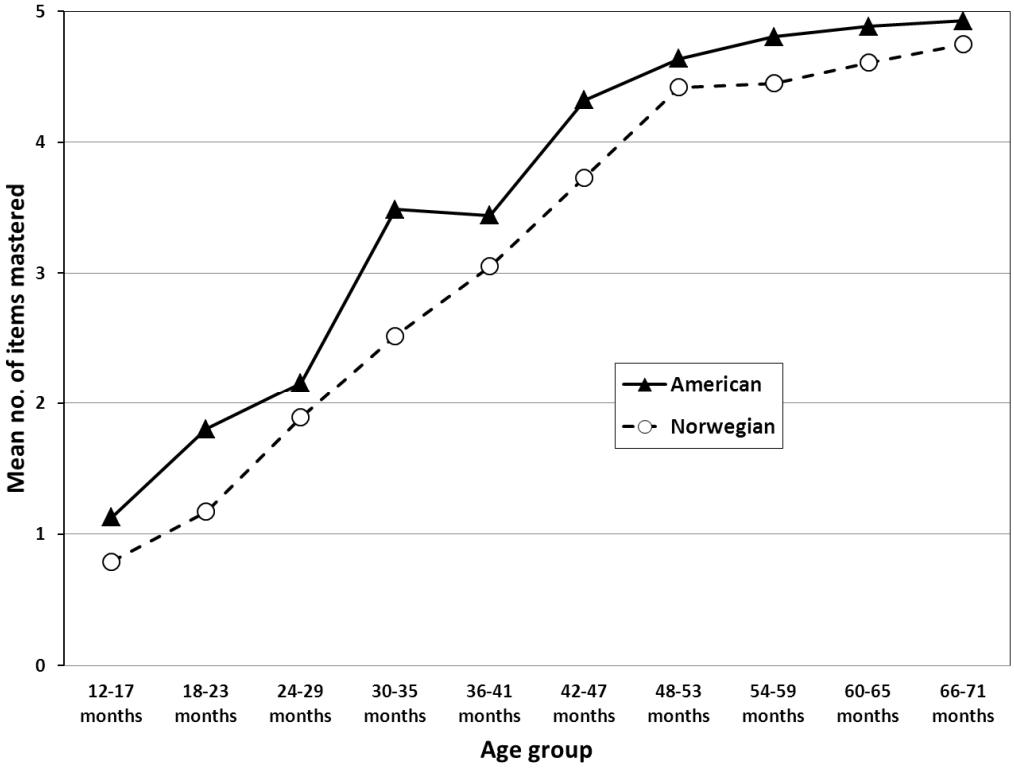

Figure 11: Age means for *Pullover/Front-Opening* Garments subdomain in two samples

Table 12: ANOVA of *Pullover/Front-Opening* Garments subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCI

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1009.321[a]	19	53.122	74.308	.000
Intercept	5668.360	1	5668.360	7928.983	.000
Nation	22.496	1	22.496	31.468	.000
group	924.153	9	102.684	143.635	.000
Nation * group	7.318	9	.813	1.137	.334
Error	369.599	517	.715		
Total	7588.000	537			
Corrected Total	1378.920	536			

a. R Squared = .732 (Adjusted R Squared = .722)

J. Fasteners (5 items)

The US sample scores higher.

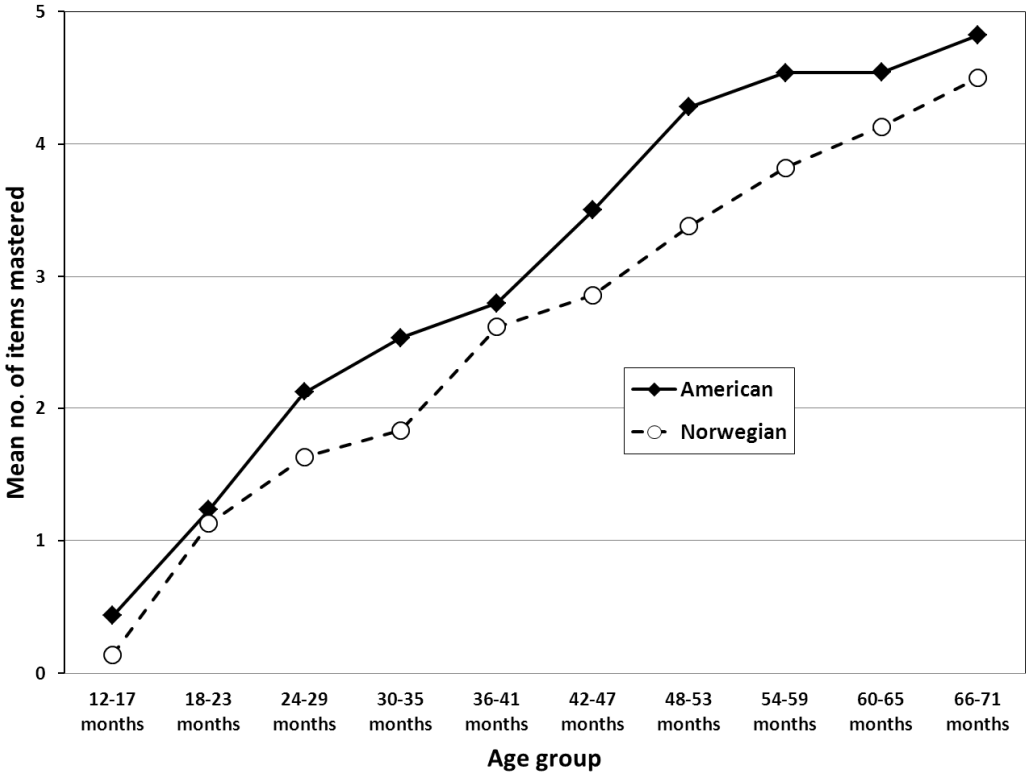

Figure 12: Age means for *Fasteners* subdomain in two samples

Table 13: ANOVA of *Fasteners* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCJ

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1045.138[a]	19	55.007	63.917	.000
Intercept	4083.472	1	4083.472	4744.887	.000
Nation	28.933	1	28.933	33.620	.000
group	939.993	9	104.444	121.361	.000
Nation * group	7.629	9	.848	.985	.451
Error	444.933	517	.861		
Total	5964.000	537			
Corrected Total	1490.071	536			

a. R Squared = .701 (Adjusted R Squared = .690)

K. Pants (5 items)

The US sample scores higher.

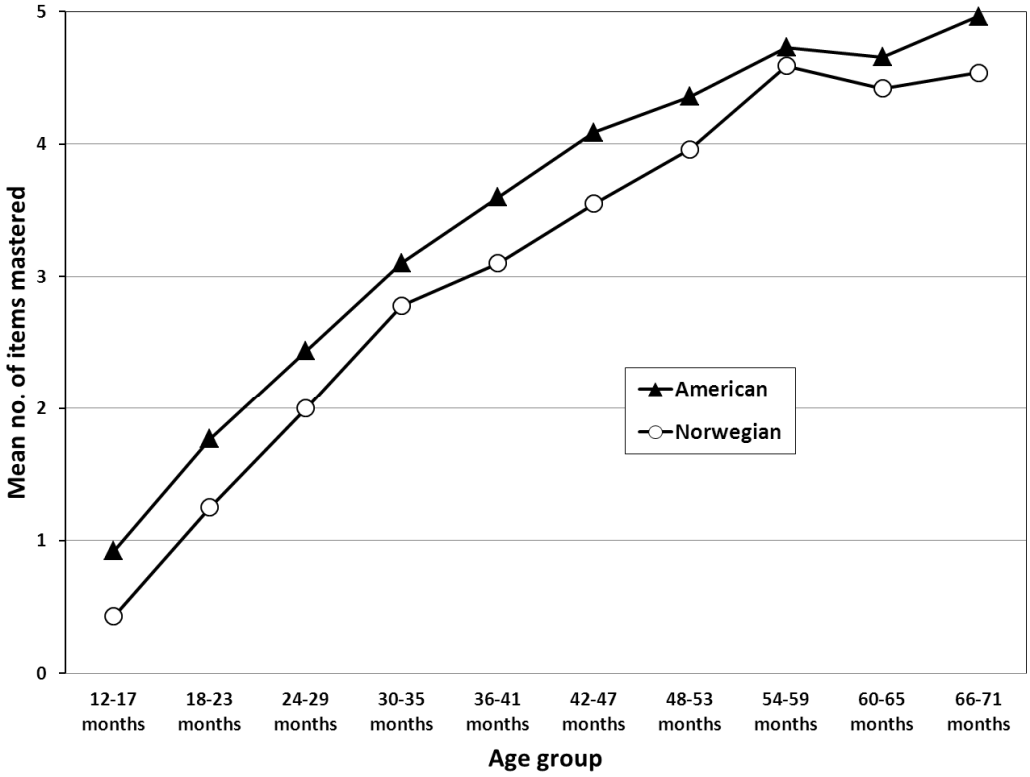

Figure 13: Age means for *Pants* subdomain in two samples

Table 14: ANOVA of *Pants* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCK

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	942.573[a]	19	49.609	82.482	.000
Intercept	5378.875	1	5378.875	8943.163	.000
Nation	20.402	1	20.402	33.921	.000
group	866.286	9	96.254	160.036	.000
Nation * group	2.024	9	.225	.374	.947
Error	310.950	517	.601		
Total	7114.000	537			
Corrected Total	1253.523	536			

a. R Squared = .752 (Adjusted R Squared = .743)

L. Shoes/socks (5 items)

The US sample scores
higher.

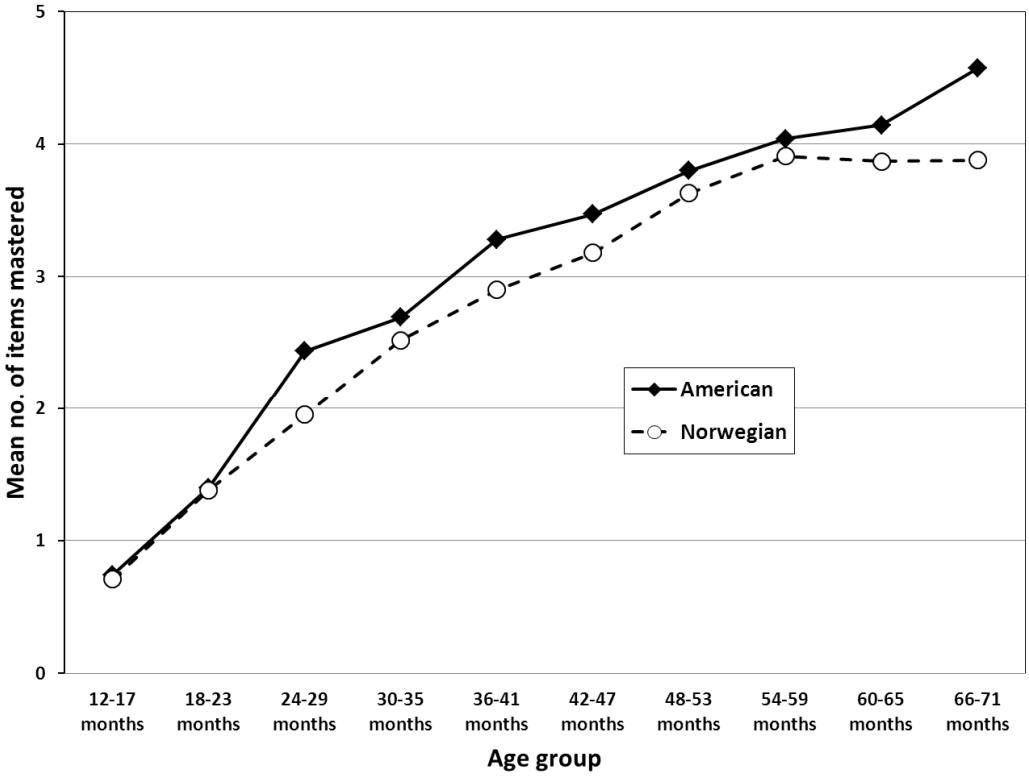

Figure 14: Age means for *Shoes/socks* subdomain in two samples

Table 15: ANOVA of *Shoes/socks* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	689.197[a]	19	36.274	74.943	.000
Intercept	4324.838	1	4324.838	8935.312	.000
Nation	8.885	1	8.885	18.358	.000
group	611.264	9	67.918	140.322	.000
Nation * group	4.944	9	.549	1.135	.336
Error	250.237	517	.484		
Total	5600.000	537			
Corrected Total	939.434	536			

a. R Squared = .734 (Adjusted R Squared = .724)

M. Toileting Tasks (5 items)

The US sample scores higher, especially in the mid-range age groups.

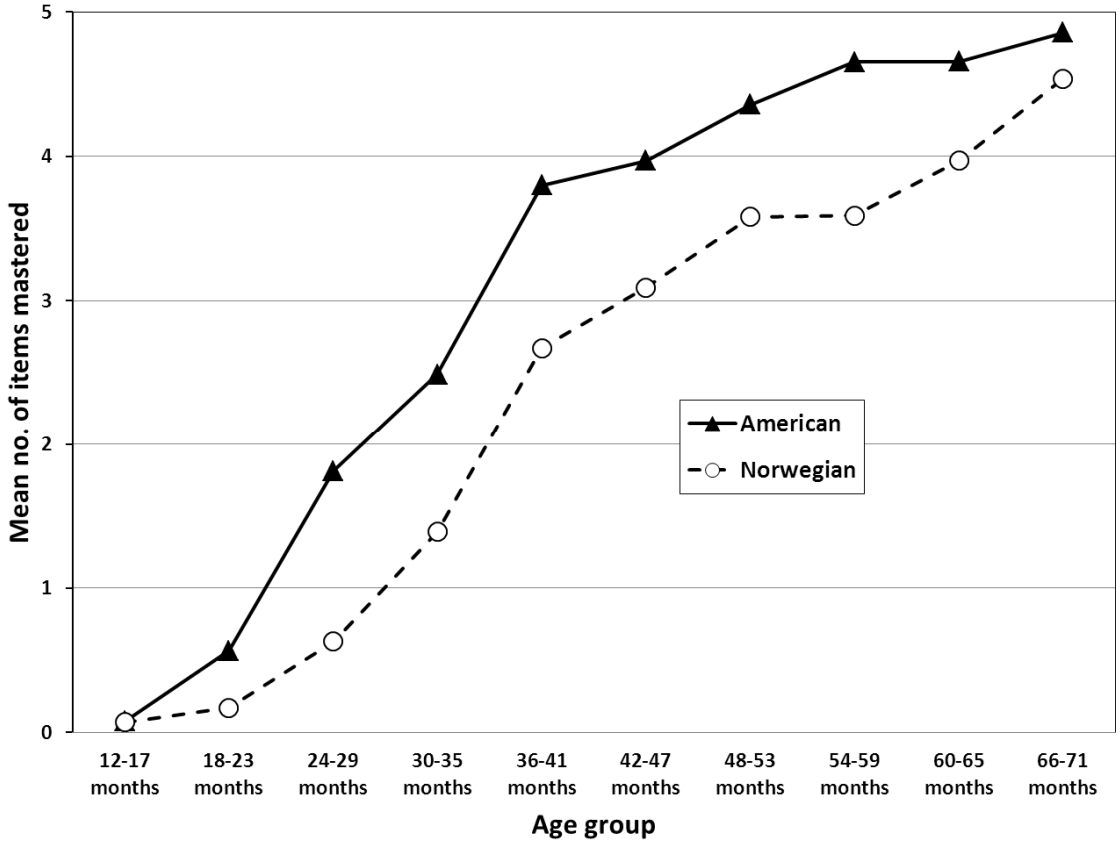

Figure 15: Age means for *Toileting Tasks* subdomain in two samples

Table 16: ANOVA of *Toileting Tasks* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1486.272[a]	19	78.225	81.634	.000
Intercept	3814.833	1	3814.833	3981.103	.000
Nation	71.837	1	71.837	74.968	.000
group	1308.007	9	145.334	151.668	.000
Nation * group	16.852	9	1.872	1.954	.043
Error	495.408	517	.958		
Total	6194.000	537			
Corrected Total	1981.680	536			

a. R Squared = .750 (Adjusted R Squared = .741)

N. Management of Bladder (5 items)

The US sample scores higher, especially in the mid-range age groups.

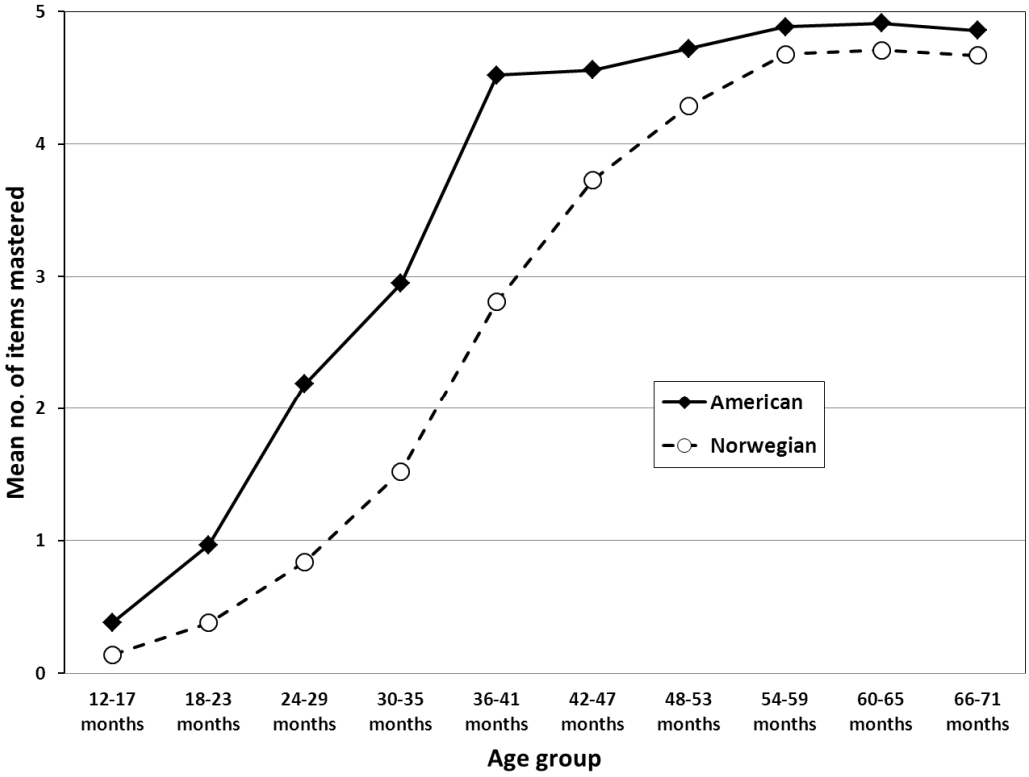

Figure 16: Age means for *Management of Bladder* subdomain in two samples

Table 17: ANOVA of *Management of Bladder* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCN

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1612.961[a]	19	84.893	94.160	.000
Intercept	4969.528	1	4969.528	5512.038	.000
Nation	65.037	1	65.037	72.137	.000
group	1443.013	9	160.335	177.838	.000
Nation * group	38.041	9	4.227	4.688	.000
Error	466.115	517	.902		
Total	7569.000	537			
Corrected Total	2079.076	536			

a. R Squared = .776 (Adjusted R Squared = .768)

O. Management of Bowel (5 items)

The US sample scores higher, especially in the mid-range age groups.

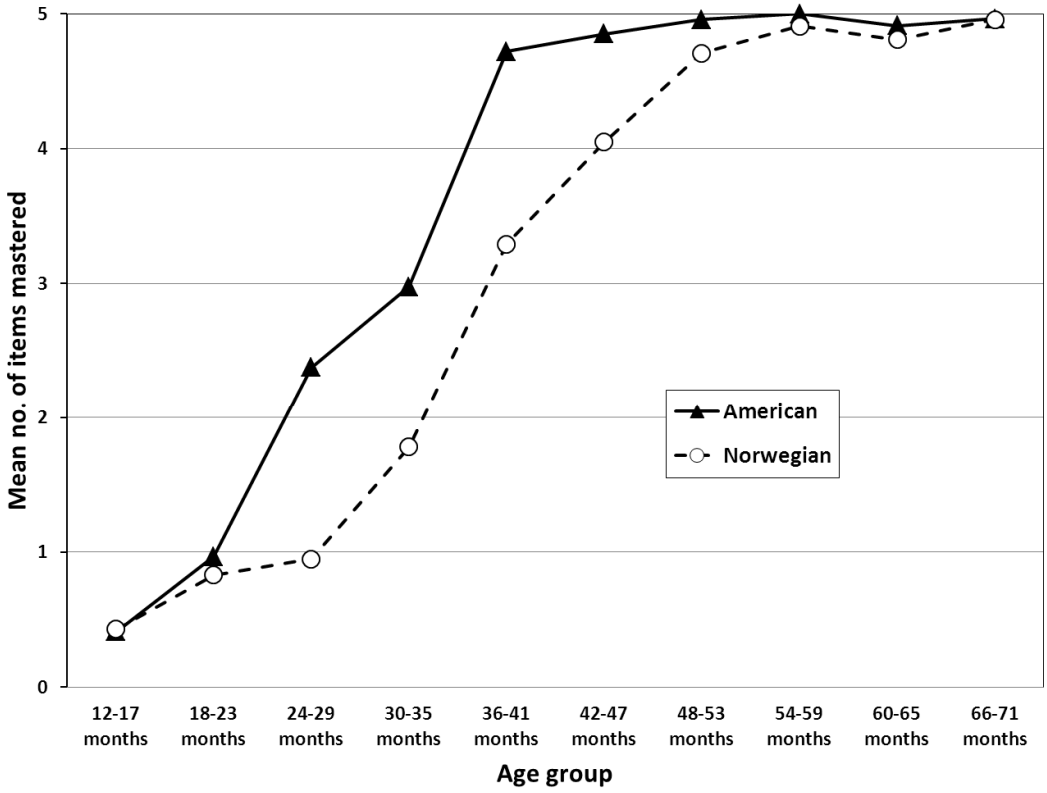

Figure 17: Age means for *Management of Bowel* subdomain in two samples

Table 18: ANOVA of *Management of Bowel* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SCO

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1634.691[a]	19	86.036	96.638	.000
Intercept	5646.047	1	5646.047	6341.738	.000
Nation	37.294	1	37.294	41.889	.000
group	1461.493	9	162.388	182.397	.000
Nation * group	41.215	9	4.579	5.144	.000
Error	460.285	517	.890		
Total	8243.000	537			
Corrected Total	2094.976	536			

a. R Squared = .780 (Adjusted R Squared = .772)

1.3 Self-care items

A. All-over sample differences in Self-care capability

When considering the individual items, it should be kept in mind that "easy" items will be mastered by more respondents than the more difficult items. The proportion of "capable" respondents mastering the behavior in question, therefore, may be read as a proxy for the "difficulty" of the item. When most respondents master an item, it is an easy one.

The Fisher statistic for each item is based on a two-by-two table containing the number of people mastering and not mastering the item in the two national samples. There is no missing data here. Hence, the initial fourfold table may be reconstructed from the percentages given and the number of cases in each sample. The p statistic gives the probability that the observed difference in (raw) numbers is due to random chance variation, and a *p* value < .05 is viewed as statistically significant. Assuming that a two-tailed test is appropriate, 31 out of the 73 items yield significant differences between the two samples.

Please note that the difference may go either way. While most items are more difficult to the Norwegian children, the US children have more trouble with other items. The items "favoring" the Norwegian sample are marked by asterisks in the table.

Table 19: Self-care items; proportion mastering items in two samples (US=313, Norw=224)

Scale	No	Label	% US	% Norw	Fisher p
SC	1	Eats pureed/strained foods	99.7	100.0	1.000
SC	2	* Eats ground/lumpy foods	91.4	100.0	***.000***
SC	3	* Eats cut/chunky foods	88.2	97.8	***.000***
SC	4	* Eats all textures of table food	69.6	91.1	***.000***
SC	5	Finger feeds	100.0	100.0	-
SC	6	Scoops and brings spoon	96.2	96.9	.814
SC	7	Uses spoon well	87.2	87.9	.895
SC	8	Uses fork well	78.0	84.8	.058
SC	9	* Butters and cuts with knife	42.8	51.8	***.044***
SC	10	Holds bottle/spout cup	99.7	100.0	1.000
SC	11	Lifts cup to drink	95.2	98.2	.095
SC	12	* Lifts cup securely w/two hands	89.5	97.3	***.000***
SC	13	Lifts cup securely w/one hand	77.0	77.7	.917
SC	14	* Pours liquid	44.7	57.6	***.004***
SC	15	Opens mouth for toothbrush	96.2	94.6	.406
SC	16	Holds toothbrush	93.9	94.6	.852
SC	17	Brushes some teeth	83.1	73.7	***.010***
SC	18	Brushes teeth thoroughly	38.3	4.5	***.000***
SC	19	Prepares toothbrush/paste	32.9	32.1	.926
SC	20	Holds head for combing	98.1	86.2	***.000***
SC	21	Brings brush/comb to hair	95.5	96.0	.833
SC	22	Brushes/combs hair	40.3	36.2	.369
SC	23	Manages tangled hair	8.0	9.4	.640
SC	24	Allows nose wipe	96.5	92.4	***.048***
SC	25	Blows nose	90.4	66.1	***.000***
SC	26	Wipes nose on request	84.3	75.4	***.011***

SC	27	Wipes nose without request	67.1	37.5	***.000***
SC	28	Blows and wipes without request	35.8	17.4	***.000***
SC	29	Holds hands out for washing	96.5	94.2	.212
SC	30	Rubs hands together in washing	89.1	81.3	***.012***
SC	31	Turns water on, gets soap	73.8	67.0	.101
SC	32	Washes hands thoroughly	58.5	45.5	***.004***
SC	33	Dries hands thoroughly	54.3	41.5	***.004***
SC	34	Tries washing parts of body	91.7	79.5	***.000***
SC	35	Washes body thoroughly, not face	60.7	29.5	***.000***
SC	36	Gets soap and soaps cloth	63.9	50.0	***.001***
SC	37	Dries body thoroughly	36.1	18.3	***.000***
SC	38	Washes/dries face thoroughly	34.5	29.5	.226
SC	39	Assists in dressing	97.4	98.7	.374
SC	40	Removes most pullover garments	81.5	75.0	.087
SC	41	Puts on most pullover garments	66.1	65.6	.927
SC	42	Puts on/removes front-opening garments	61.3	61.2	1.000
SC	43	Puts on/removes fastened garments	40.6	29.9	***.014***
SC	44	Tries assisting with fasteners	84.7	87.5	.381
SC	45	Zips/unzips	75.4	85.7	***.003***
SC	46	Snaps/unsnaps	58.1	45.5	***.005***
SC	47	Buttons/unbuttons	46.6	32.1	***.001***
SC	48	Separates and unhooks zipper	31.9	26.3	.180
SC	49	Assists with pants	94.8	96.0	.679
SC	50	Removes elastic waist pants	83.7	83.5	1.000
SC	51	Puts on elastic waist pants	70.6	72.3	.699
SC	52	Unfastens and removes pants	51.4	43.3	.066
SC	53	Puts pants on and fastens	35.1	27.7	.075
SC	54	* Removes socks and unfastened shoes	93.3	97.3	***.044***
SC	55	Puts on unfastened shoes	77.6	83.5	.101
SC	56	Puts on socks	69.3	65.1	.350
SC	57	Puts shoes on correct foot, manages Velcro	42.5	45.1	.597
SC	58	Ties shoelaces	13.4	1.3	***.000***
SC	59	Assists with clothing	76.4	73.7	.480
SC	60	Tries to wipe self after toilet	68.1	56.3	***.006***
SC	61	Manages toilet seat, paper, flush	64.9	58.0	.125
SC	62	Manages clothes before and after toilet	58.5	55.4	.481
SC	63	Wipes self thoroughly after bowel	30.7	11.2	***.000***
SC	64	Indicates when wet	85.6	79.0	***.049***
SC	65	Occasionally indicates need to urinate	74.8	68.8	.143
SC	66	Consistently indicates need to urinate	64.2	54.9	***.032***
SC	67	Takes self to bathroom to urinate	63.3	55.8	.090
SC	68	Consistently dry day and night	47.6	39.3	.064
SC	69	Indicates need to change	88.8	88.8	1.000
SC	70	Occasionally indicates toilet need	72.2	66.5	.182
SC	71	Consistently indicates toilet need	64.2	56.7	.088
SC	72	Distinguishes urination/bowel	63.6	61.2	.588
SC	73	Takes self into toilet for bowel	57.8	53.6	.334

It may also be worth noting that while there were sample differences on several subdomains, such differences do not necessarily occur with the individual scales within the subdomain.

B. Age and sample differences on Self-care single items

A closer look at the five items of the Management of Bowel subdomain (SC69 through SC 73, cf. Figure 17 above) may provide an illustrative example of the advantage of considering sample and age differences simultaneously.

Blandly overstretching normal measurement assumptions, an ANOVA was performed on the individual item scores (0 or 1), with age groups (10) and samples (2) as factors. Results should of course be interpreted very cautiously, but offer an interesting first look.

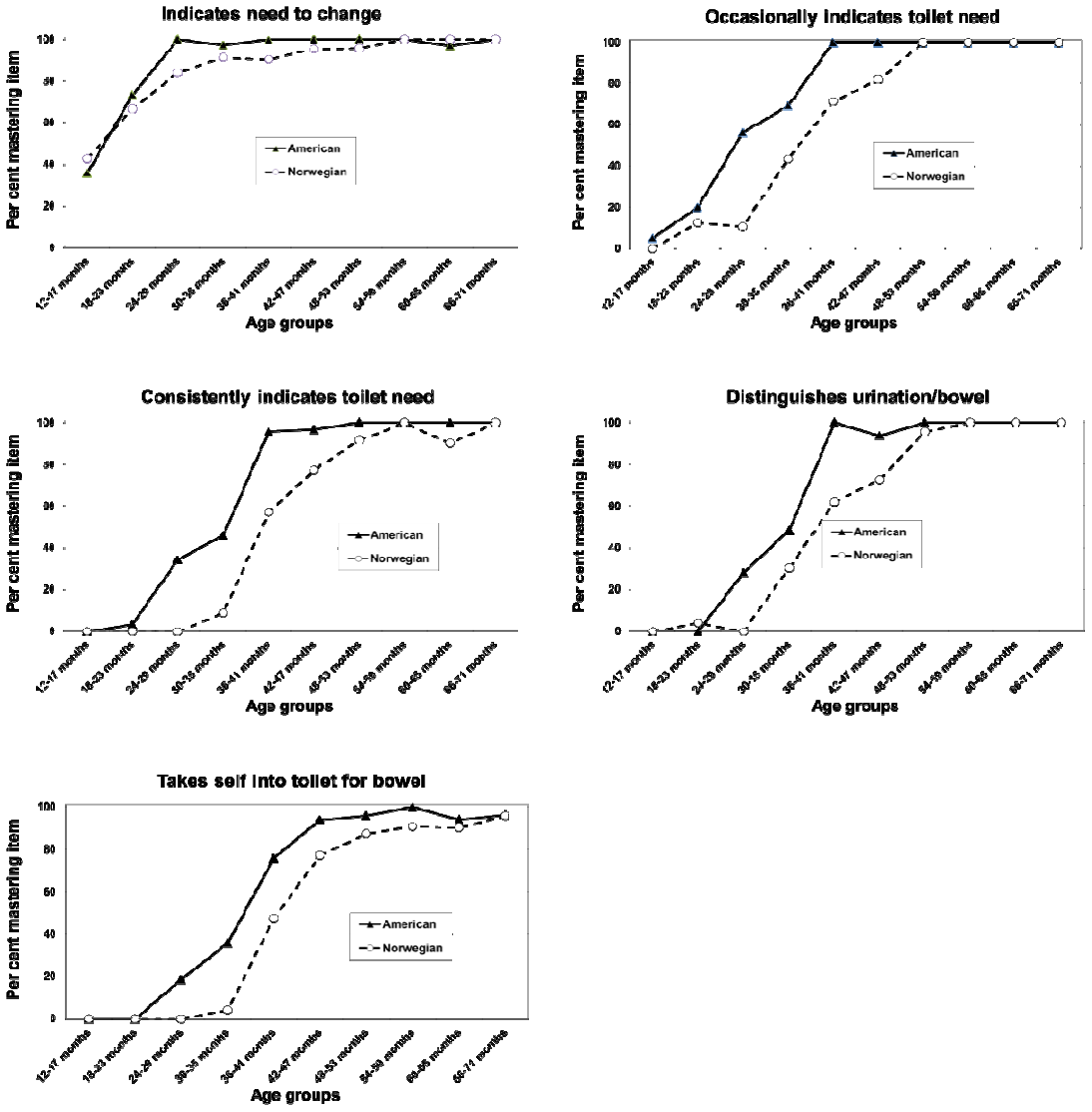

Figure 18: Per cent mastering the five items (SC69-73) of the Management of Bowel subdomain, in two samples with ten age groups.

The first of these items (SC69: *Indicates need to change*), yields an age effect only. This lack of a significant sample effect is consistent with results from the simpler approach used for the Fisher statistics in table 19.

With the remaining four items (SC70 through SC73), however, we obtain three significant effects:

1. An age effect, indicating general age differences across both samples
2. A sample effect, showing higher mean scores with the US sample
3. An interaction effect, indicating that the difference between the samples is not consistent across different ages.

Inspecting all curves in figure 18, we first see that they "peak" at different ages. While most American children indicate a need to be changed at the age of 24-29 months, mastery of the more difficult items comes later. Also the curves for the Norwegian children reflect the increasing difficulty of items SC69 through SC73.

Second, we see that the curve for the US children generally is above that of the Norwegians, showing that their mean score most often is the higher one. This is the *sample* effect.

Third, the percentage of children mastering each item clearly increases with age. While the younger children do not master most items, the older children generally do. This is the *age* effect.

Last, but not least, the sample difference is generally small with the younger groups. It then increases in mid-range, only to decrease again with the older age groups. In the beginning, the items are too difficult for most children. The American children soon face the challenges, however, and rise to master the items within 3-4 half-years. The Norwegian children also do, but a couple of half-years later. After a while, however, the Norwegians catch up, and both samples master the items. This slightly complex pattern is the *interaction* effect.

Since both age, sample and interaction effects are roughly similar with four out of the five items, they combine nicely to yield the very same effects on the subdomain level (cf. figure 17 and table 18, Management of Bowel subdomain).

A rather similar pattern may be found with the *Management of Bladder* subdomain, as shown in figure 19. Across all items, the American sample generally shows a higher percentage mastering the item. This adds up to a significant sample effect for this subdomain. This effect, however, does not extend to all single items. The difference between the proportion mastering the item in the two samples is significant only with two out of the five items (*Indicates when wet* and *Consistently indicates need to urinate*).

Viewed together, the item data from this subdomain suggest a pattern quite similar to that of the *Management of Bowel* subdomain. In the younger and the older age groups, the differences between the national samples are limited. At intermediate ages, however, a larger part of the American sample masters the items. If even younger and older groups had been added to these samples, ceiling as well as floor effects would quite likely become evident, clearly showing a limited age span where the items are relevant and useful.

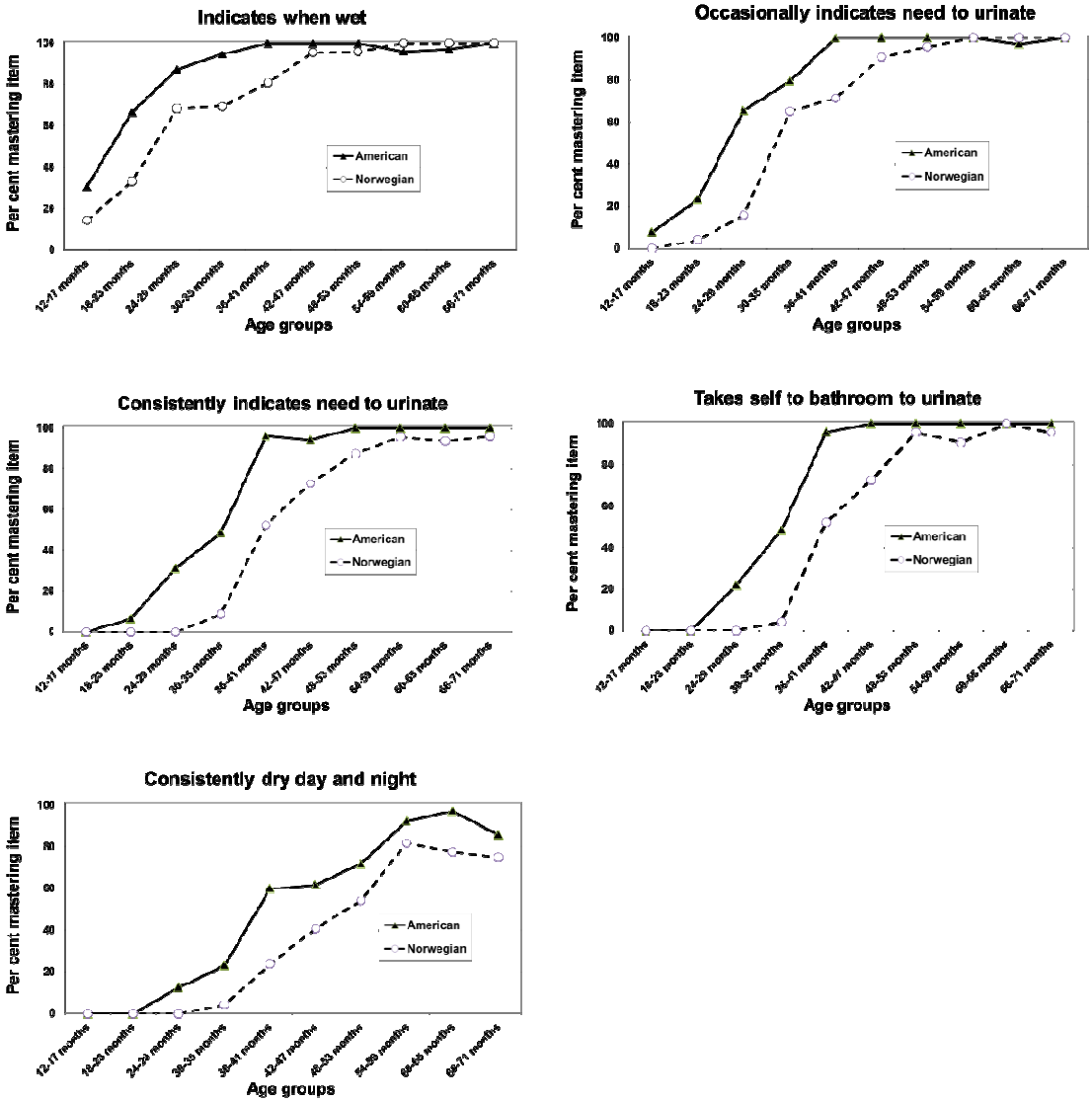

Figure 19: Per cent mastering the five items (SC64-68) of the Management of Bladder subdomain, in two samples with ten age groups

C. Sample differences of item difficulty

For a more sophisticated different way of comparing items across the two samples, consider the *item difficulty calibrations* computed in the Rasch analyses. Here, the "easy" items (mastered by most children) will have low scores, while item mastered by just a few will have high.

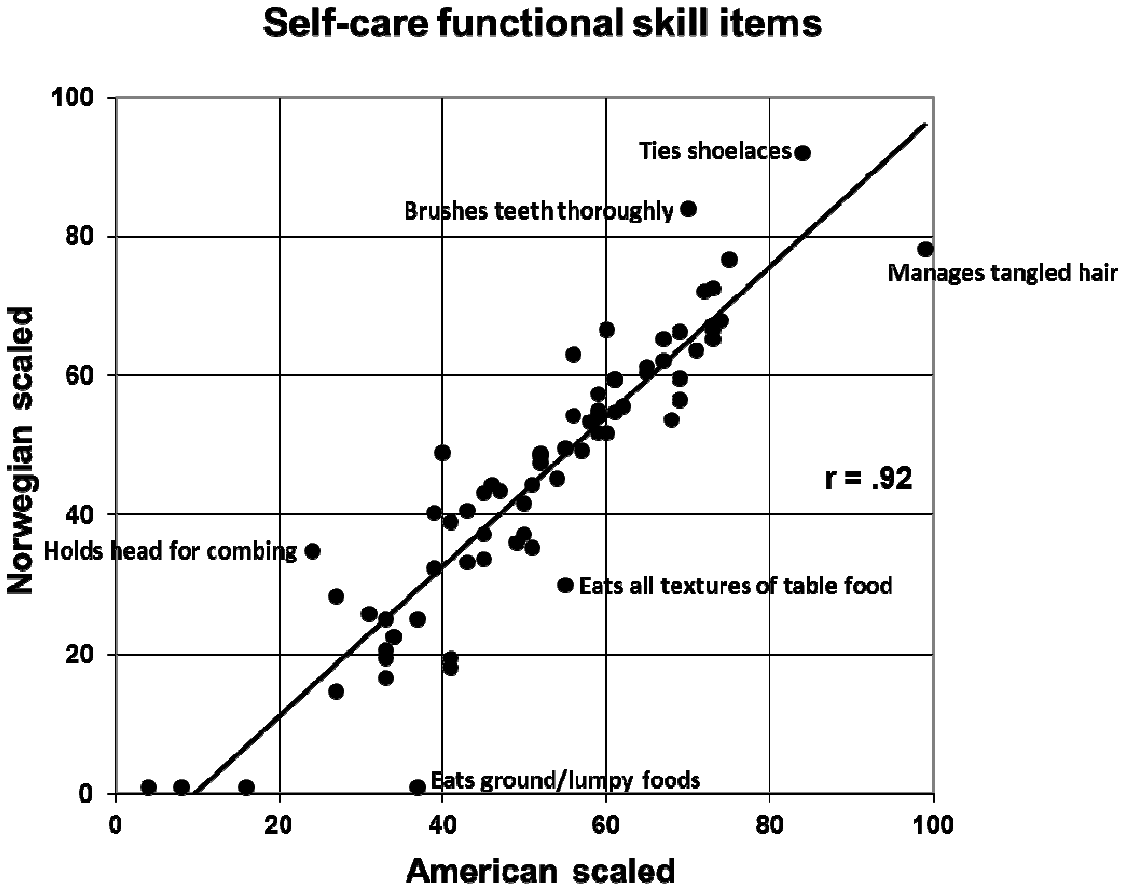

Figure 20: Self-care item difficulty scores in the two normative samples

Roughly speaking, the items are not too far from the regression line. With some exceptions, their relative placement therefore is relatively similar with the two samples.

In the scatterplot, each item is placed according to its difficulty calibration in the American (X-axis) and Norwegian (Y axis) samples. If the relative ranks in the two samples for all items had been equal, all observations would sit close to the regression line. Clearly, this is not the case: a number of items are placed some distance from this ideal line. To indicate the meaning of some of the "deviating" items, texts have been added. It may be worth noting that while some items are more difficult to the American sample (*Eats ground/lumpy foods, Eats all textures of table food, Manages tangled hair*), others are easier (*Holds head for combing, Brushes teeth thoroughly, Ties shoelaces*).

It is also worth noting that the regression line passes through the Y axis well below its zero point. Hence, a regression equation attempting to predict the Norwegian scores from their American counterparts will include some negative constant. In simpler terms, this

implies that item difficulty scores are *generally* higher in the American than in the Norwegian sample.

D. DIF-tests of Self-Care item difficulty differences

Still another way of assessing the difficulty of items in the two samples, is to employ a DIF analysis (Tennant & Pallant, 2007). In figure 21, the item difficulty scores from the two normative samples are plotted. Please cf. table 19 for viewing the content of the different item numbers.

Many items have rather similar scores in the two samples, and fall close to the dotted straight regression line. Quite a few items, however, have appreciably different scores in the two samples. They fall outside the 95% confidence interval shown as a “funnel” formed by heavy black lines. Among these are, e.g., item 25 (Blows nose) and item 4 (Eats all textures of table food).

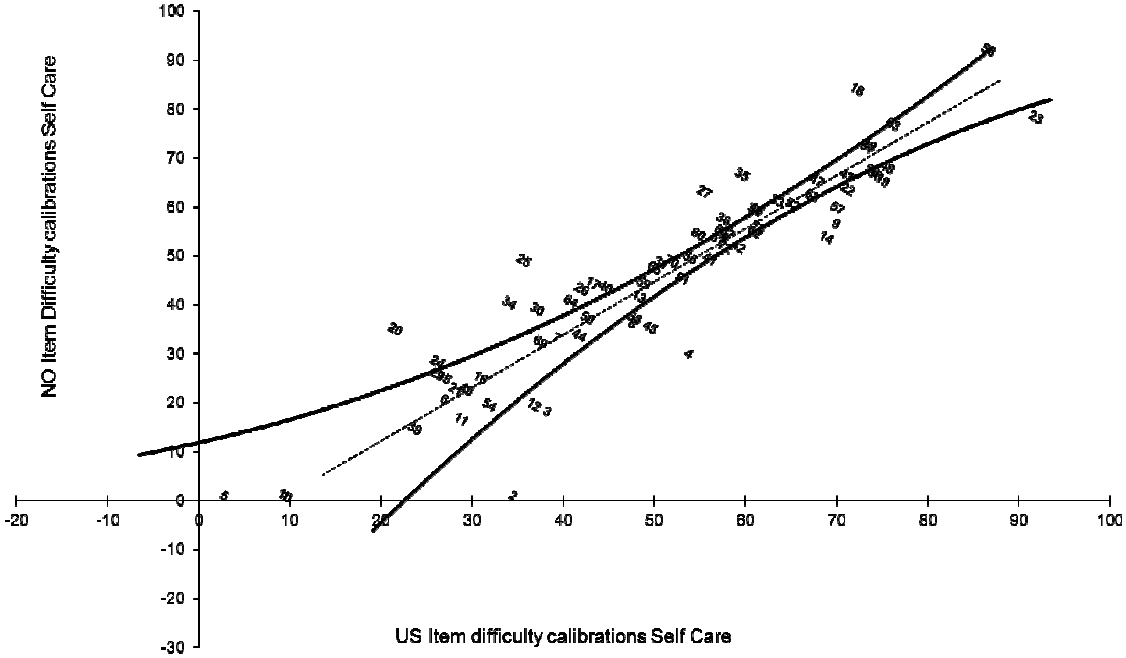

Figure 21: DIF analysis of Self-care items in the two normative samples

The numbers in table 20 confirm that the differences between the samples are substantial, listing the items displaying t-values >2.

It is worth observing that differences are relatively numerous. Also, they go both ways, and generally correspond fairly well to the differences mapped in table 19.

Table 20: Self-care items with significant sample differences

Norwegian			***American***		
Item #	*Item name*	*DIF t-value >2*	*Item #*	*Item name*	*DIF t-value >2*
25	Blows nose	7.19	4	Eats all textures of table food	6.71
27	Wipes nose without request	5.66	14	Pours liquid	5.30
20	Holds head for combing	5.38	9	Butters and cuts with knife	4.55
18	Brushes teeth thoroughly	5.35	45	Zipes/unzips	3.49
35	Washes body thoroughly, not face	5.33	3	Eats cut/chunky foods	3.29
34	Tries washing parts of body	4.19	2	Eats ground /lumpy foods	3.26
30	Rubs hands together in washing	2.62	57	Puts shoes on correct foot, Velcro OK	3.18
17	Brushes some teeth	2.42	23	Manages tangled hair	2.92
26	Wipes nose on request	2.40	19	Prepares toothbrush/paste	2.89
24	Allows nose wipe	2.37	12	Lifts cup securely w/one hand	2.70
			8	Uses fork well	2.44
			55	Puts on unfastened shoes	2.17

2. Mobility function

2.1 Mobility domain

Secondly, consider the Mobility scale (Max. score = 59). Here, the ceiling effect may be observed in both samples.

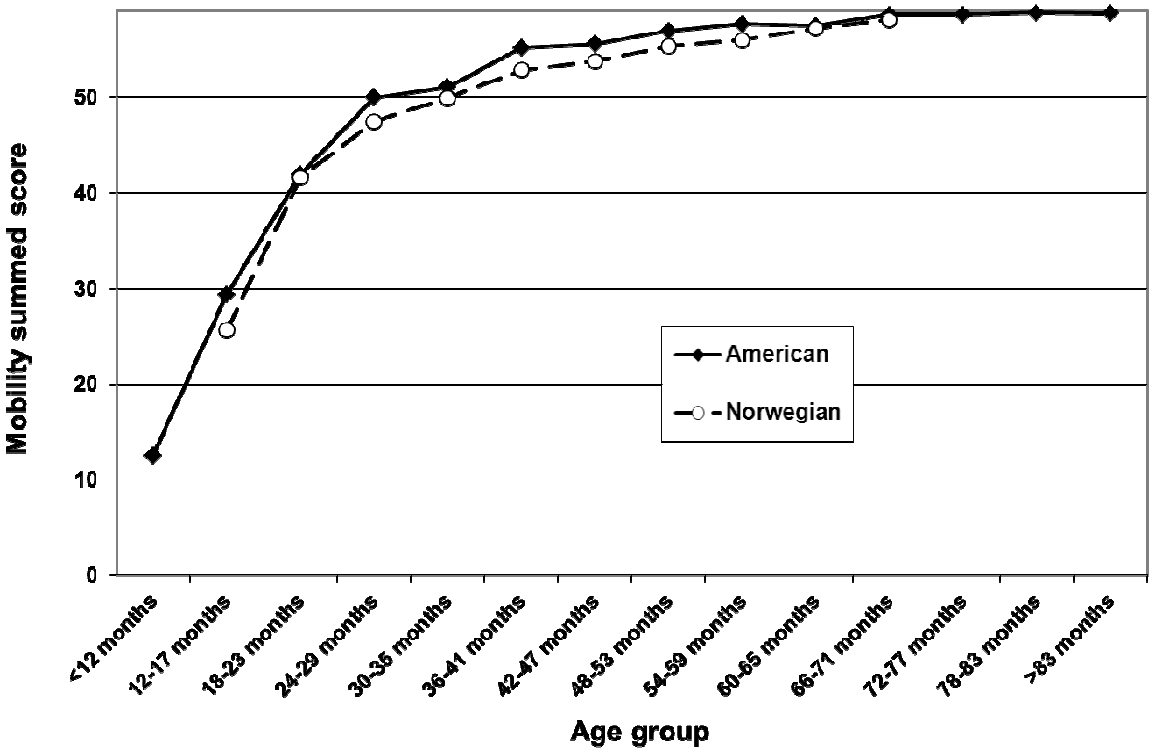

Figure 22: Age means for Mobility domain raw score in two samples

And again, a two-way ANOVA identifies both *age group* and *nation* as significant effects. Consistent with the impression gained from the graphs, however, the national (sample) difference is smaller here than with the Self-care domain data.

Table 21: ANOVA of Mobility raw score in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:Mobility summed raw score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	42148.590[a]	19	2218.347	131.546	.000
Intercept	1294671.825	1	1294671.825	76773.116	.000
group	37114.244	9	4123.805	244.539	.000
Nation	327.018	1	327.018	19.392	.000
group * Nation	127.224	9	14.136	.838	.581
Error	8718.486	517	16.864		
Total	1436813.000	537			
Corrected Total	50867.076	536			

Tests of Between-Subjects Effects

Dependent Variable:Mobility summed raw score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	42148.590[a]	19	2218.347	131.546	.000
Intercept	1294671.825	1	1294671.825	76773.116	.000
group	37114.244	9	4123.805	244.539	.000
Nation	327.018	1	327.018	19.392	.000
group * Nation	127.224	9	14.136	.838	.581
Error	8718.486	517	16.864		
Total	1436813.000	537			
Corrected Total	50867.076	536			

a. R Squared = .829 (Adjusted R Squared = .822)

2.2 Mobility subdomains

A. Toilet Transfers (5 items)

The US sample scores higher.

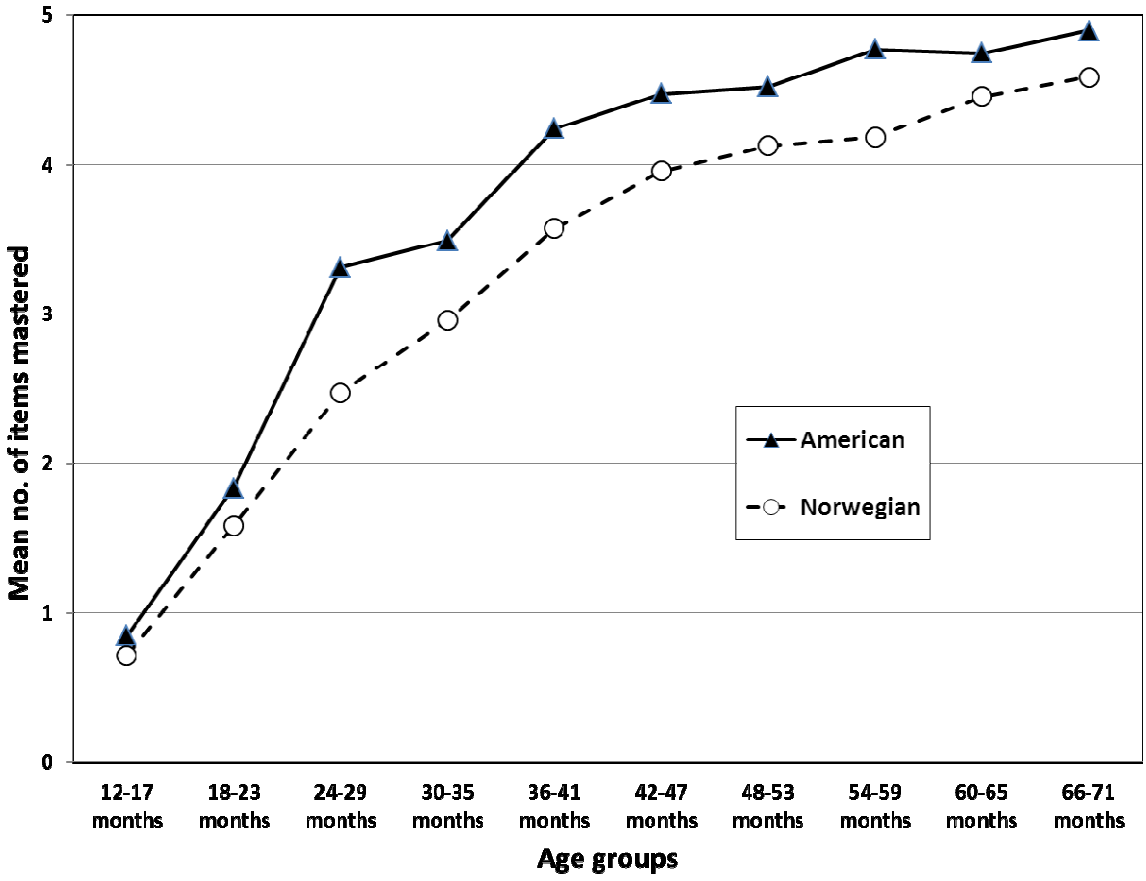

Figure 23: Age means for Toilet Transfer subdomain in two samples

Table 22: ANOVA of *Toilet Transfer* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	884.957[a]	19	46.577	63.202	.000
Intercept	6140.806	1	6140.806	8332.753	.000
Nation	25.807	1	25.807	35.019	.000
group	768.206	9	85.356	115.824	.000
Nation * group	5.007	9	.556	.755	.658
Error	381.002	517	.737		
Total	7932.000	537			
Corrected Total	1265.959	536			

a. R Squared = .699 (Adjusted R Squared = .688)

B. Chair/Wheelchair Transfers (5 items)

The US sample scores higher.

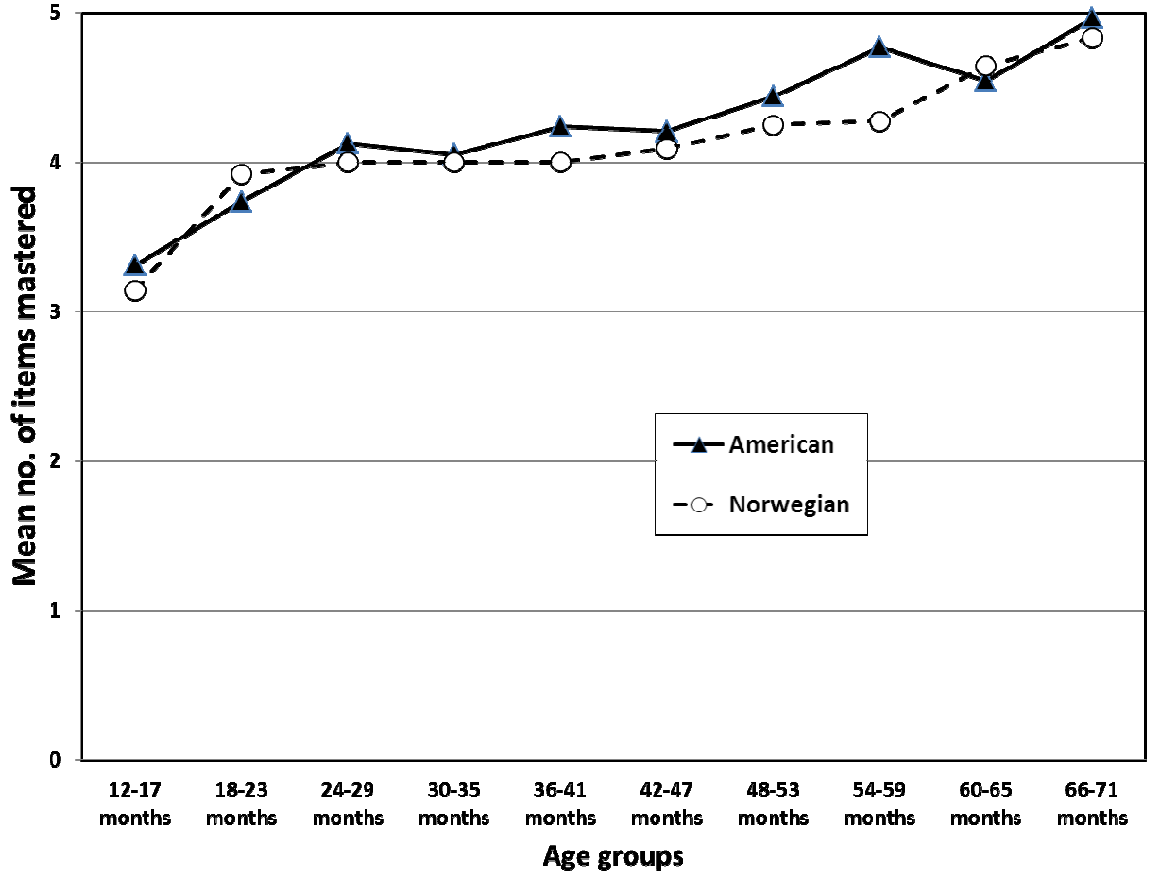

Figure 24: Age means for Chair/Wheelchair Transfers subdomain in two samples

Table 23: ANOVA of *Chair/Wheelchair Transfers* Transfer subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOB

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	105.228[a]	19	5.538	21.895	.000
Intercept	8817.177	1	8817.177	34857.285	.000
Nation	1.905	1	1.905	7.532	.006
group	91.136	9	10.126	40.032	.000
Nation * group	4.029	9	.448	1.770	.071
Error	130.776	517	.253		
Total	9655.000	537			
Corrected Total	236.004	536			

a. R Squared = .446 (Adjusted R Squared = .426)

C. Car Transfers (5 items)

The US sample scores higher, but not in the two oldest groups.

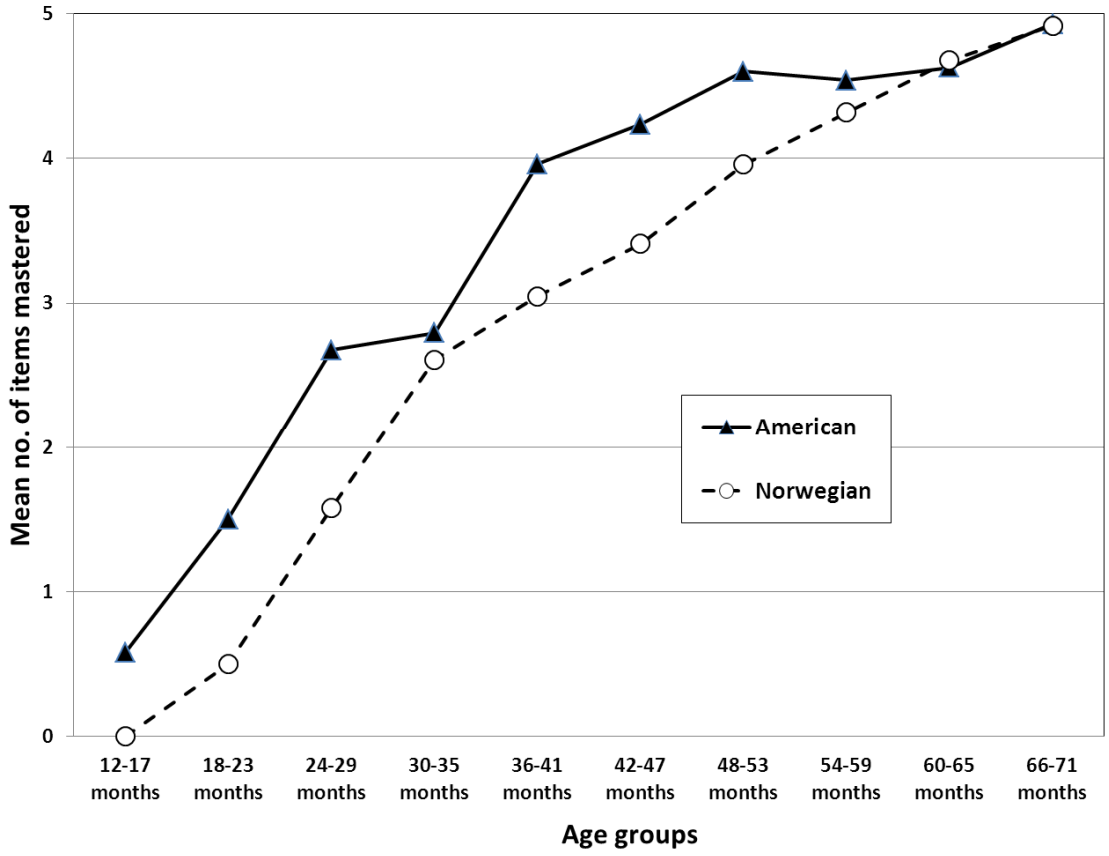

Figure 25: Age means for Car Transfers subdomain in two samples

Table 24: ANOVA of *Car Transfers* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1221.077[a]	19	64.267	87.983	.000
Intercept	5077.731	1	5077.731	6951.486	.000
Nation	37.141	1	37.141	50.847	.000
Group	1117.914	9	124.213	170.049	.000
Nation * group	21.577	9	2.397	3.282	.001
Error	376.183	515	.730		
Total	7185.000	535			
Corrected Total	1597.260	534			

a. R Squared = .764 (Adjusted R Squared = .756)

D. Bed Mobility/Transfers (4 items)

The US sample scores higher, but the sample difference diminishes with age.

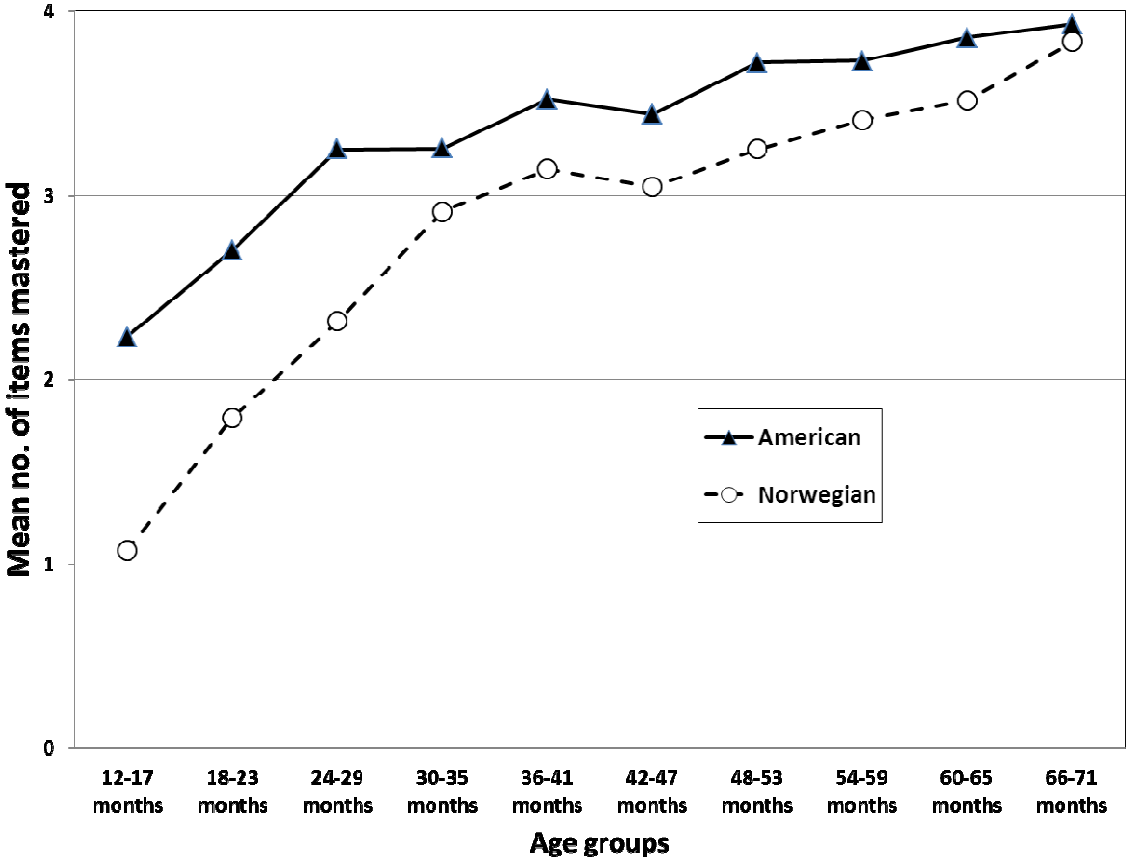

Figure 26: Age means for Bed Mobility/Transfers subdomain in two samples

Table 25: ANOVA of *Bed Mobility/Transfers* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOD

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	234.043[a]	19	12.318	38.176	.000
Intercept	4845.569	1	4845.569	15017.387	.000
Nation	36.116	1	36.116	111.930	.000
Group	205.140	9	22.793	70.641	.000
Nation * group	12.513	9	1.390	4.309	.000
Error	166.817	517	.323		
Total	5751.000	537			
Corrected Total	400.860	536			

a. R Squared = .584 (Adjusted R Squared = .569)

E. Tub Transfers (5 items)

The US sample scores higher.

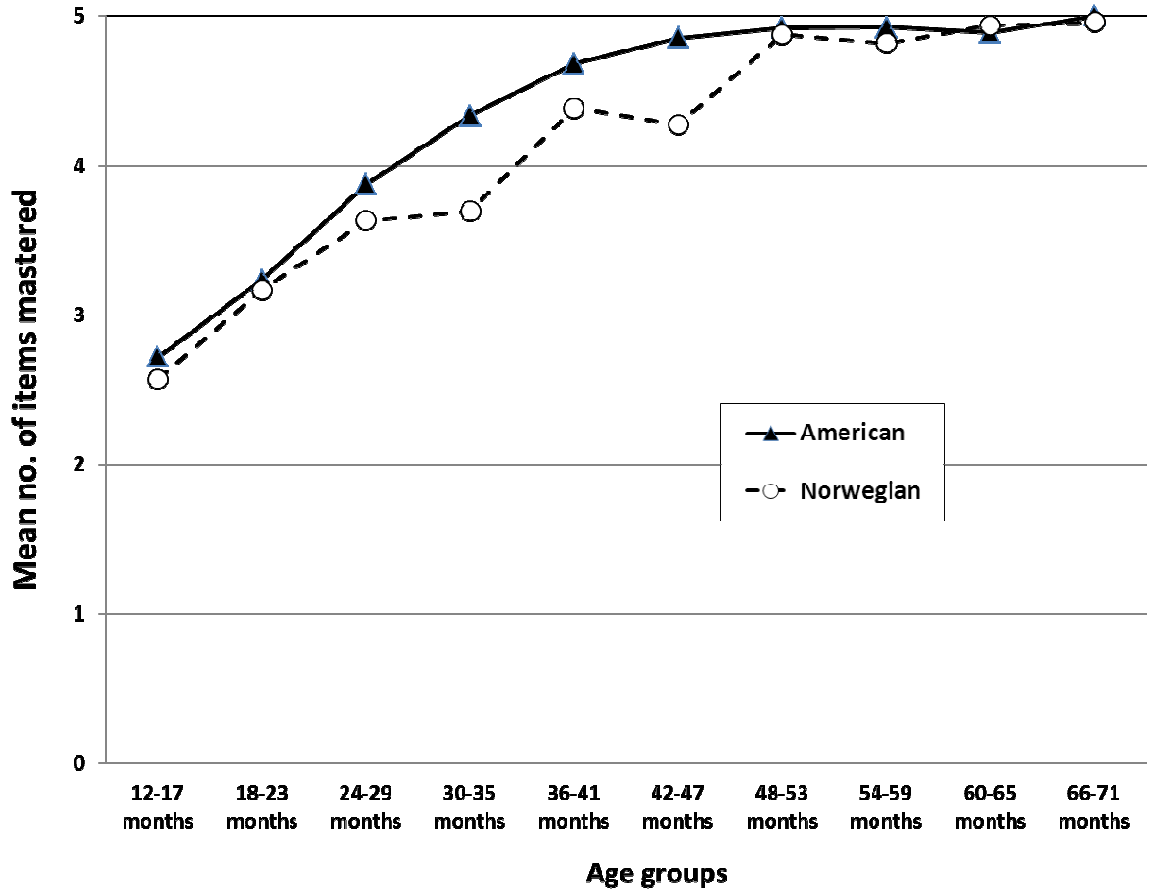

Figure 27: Age means for Tub Transfers subdomain in two samples

Table 26: ANOVA of *Tub Transfers* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	322.560[a]	19	16.977	34.493	.000
Intercept	9071.502	1	9071.502	18431.052	.000
Nation	5.654	1	5.654	11.489	.001
group	283.560	9	31.507	64.014	.000
Nation * group	6.836	9	.760	1.543	.130
Error	254.460	517	.492		
Total	10317.000	537			
Corrected Total	577.020	536			

a. R Squared = .559 (Adjusted R Squared = .543)

F. Indoor Locomotion Methods (3 items)

No significant difference between the two samples.

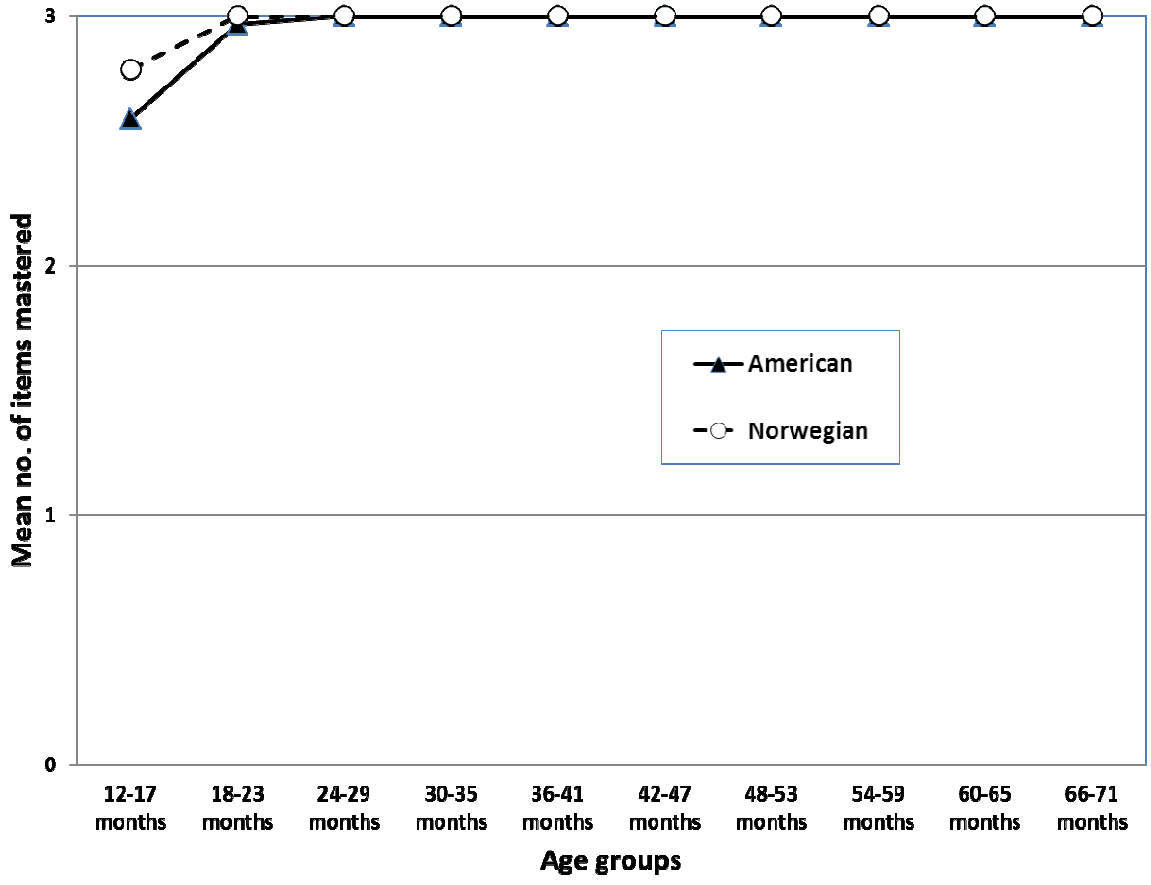

Figure 28: Age means for Indoor Locomotion Methods subdomain in two samples

Table 27: ANOVA of *Indoor Locomotion Methods* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOF

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.495[a]	19	.342	7.766	.000
Intercept	4449.981	1	4449.981	101083.900	.000
Nation	.066	1	.066	1.509	.220
Group	3.664	9	.407	9.249	.000
Nation * group	.363	9	.040	.917	.510
Error	22.760	517	.044		
Total	4743.000	537			
Corrected Total	29.255	536			

a. R Squared = .222 (Adjusted R Squared = .193)

G. Indoor Locomotion – Distance/speed (5 items)

No significant difference between the two samples.

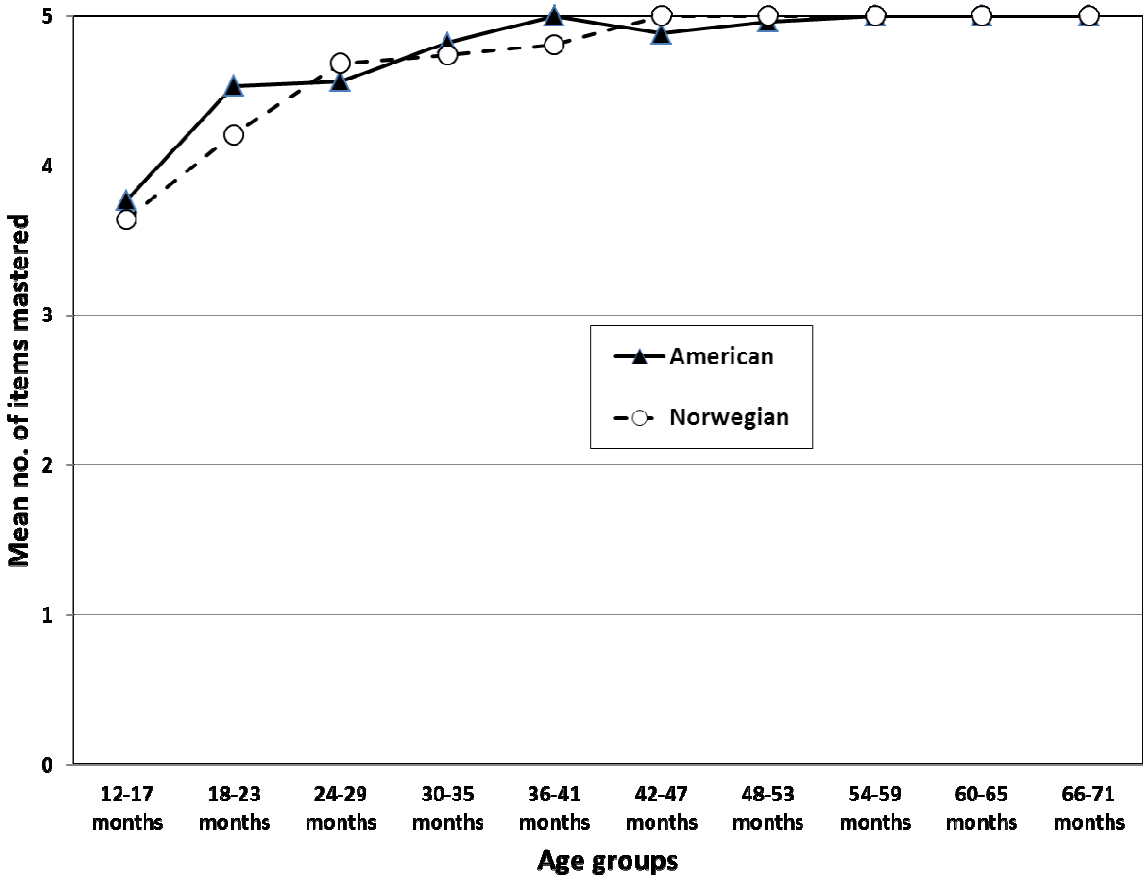

Figure 29: Age means for Indoor Locomotion – Distance/speed subdomain in two samples

Table 28: ANOVA of *Indoor Locomotion – Distance/speed* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	81.002[a]	19	4.263	21.306	.000
Intercept	11311.605	1	11311.605	56531.507	.000
Nation	.249	1	.249	1.244	.265
Group	69.341	9	7.705	38.505	.000
Nation * group	2.230	9	.248	1.238	.269
Error	103.449	517	.200		
Total	12227.000	537			
Corrected Total	184.451	536			

a. R Squared = .439 (Adjusted R Squared = .419)

H. Indoor Locomotion – Pulls /Carries Objects (5 items)

The Norwegian sample scores higher.

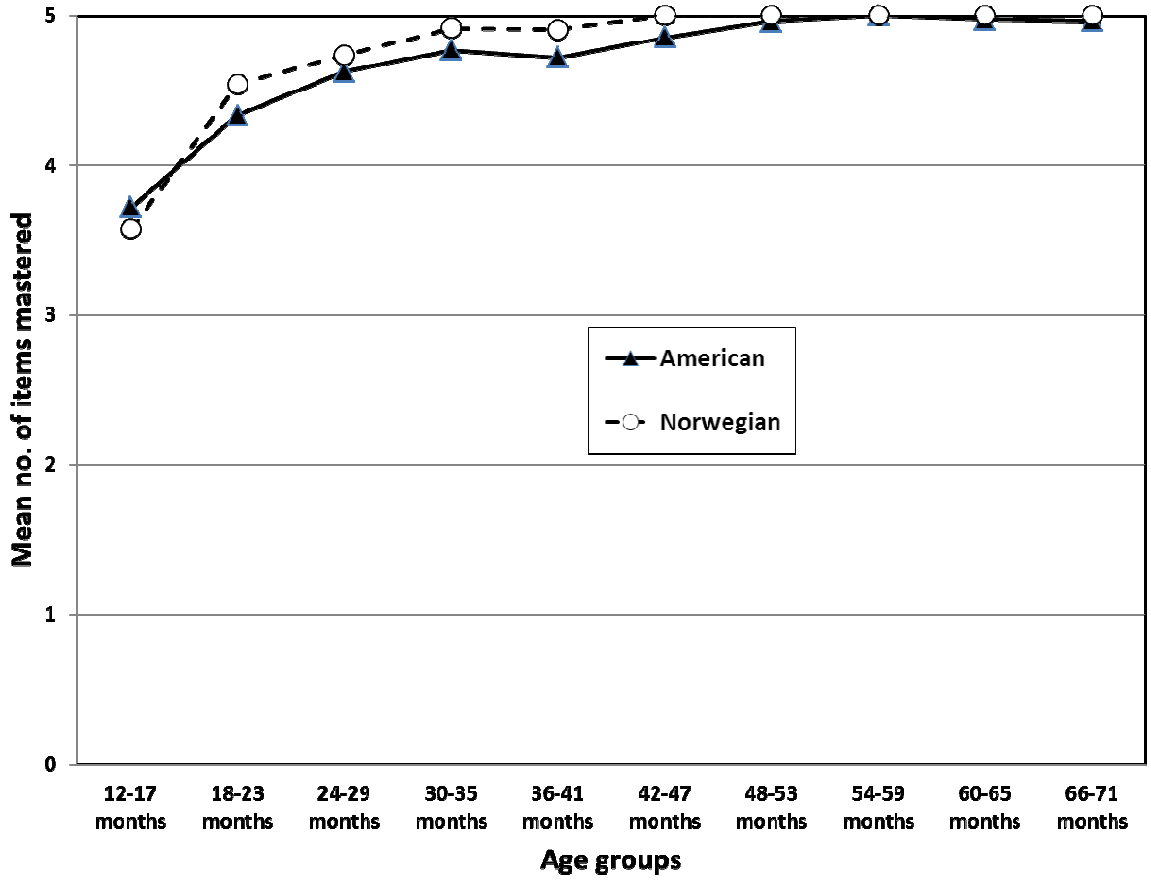

Figure 30: Age means for Indoor Locomotion – Pulls/Carries Objects subdomain in two samples

Table 29: ANOVA of *Indoor Locomotion – Pulls/Carries* Objects subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOH

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	82.411[a]	19	4.337	25.513	.000
Intercept	11304.415	1	11304.415	66493.288	.000
Nation	.718	1	.718	4.221	.040
Group	70.097	9	7.789	45.813	.000
Nation * group	1.169	9	.130	.764	.650
Error	87.894	517	.170		
Total	12175.000	537			
Corrected Total	170.305	536			

a. R Squared = .484 (Adjusted R Squared = .465)

I. Outdoor Locomotion – Methods (2 items)

No significant difference between the two samples.

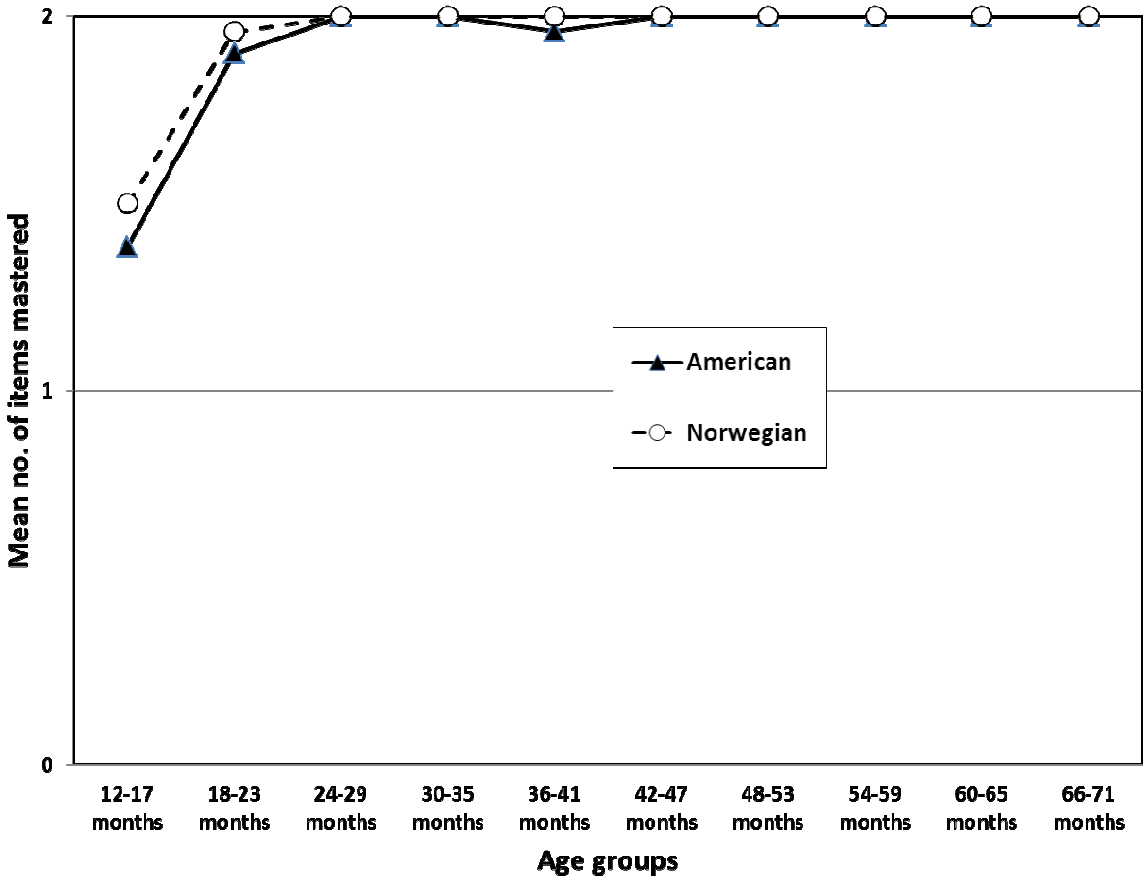

Figure 31: Age means for Outdoor Locomotion – Methods subdomain in two samples

Table 30: ANOVA of *Outdoor Locomotion – Methods* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOI

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	16.237[a]	19	.855	10.685	.000
Intercept	1892.867	1	1892.867	23667.070	.000
Nation	.058	1	.058	.722	.396
group	11.610	9	1.290	16.130	.000
Nation * group	.155	9	.017	.215	.992
Error	41.349	517	.080		
Total	2064.000	537			
Corrected Total	57.587	536			

a. R Squared = .282 (Adjusted R Squared = .256)

J. Outdoor Locomotion – Distance /Speed (5 items)

No significant difference between the two samples.

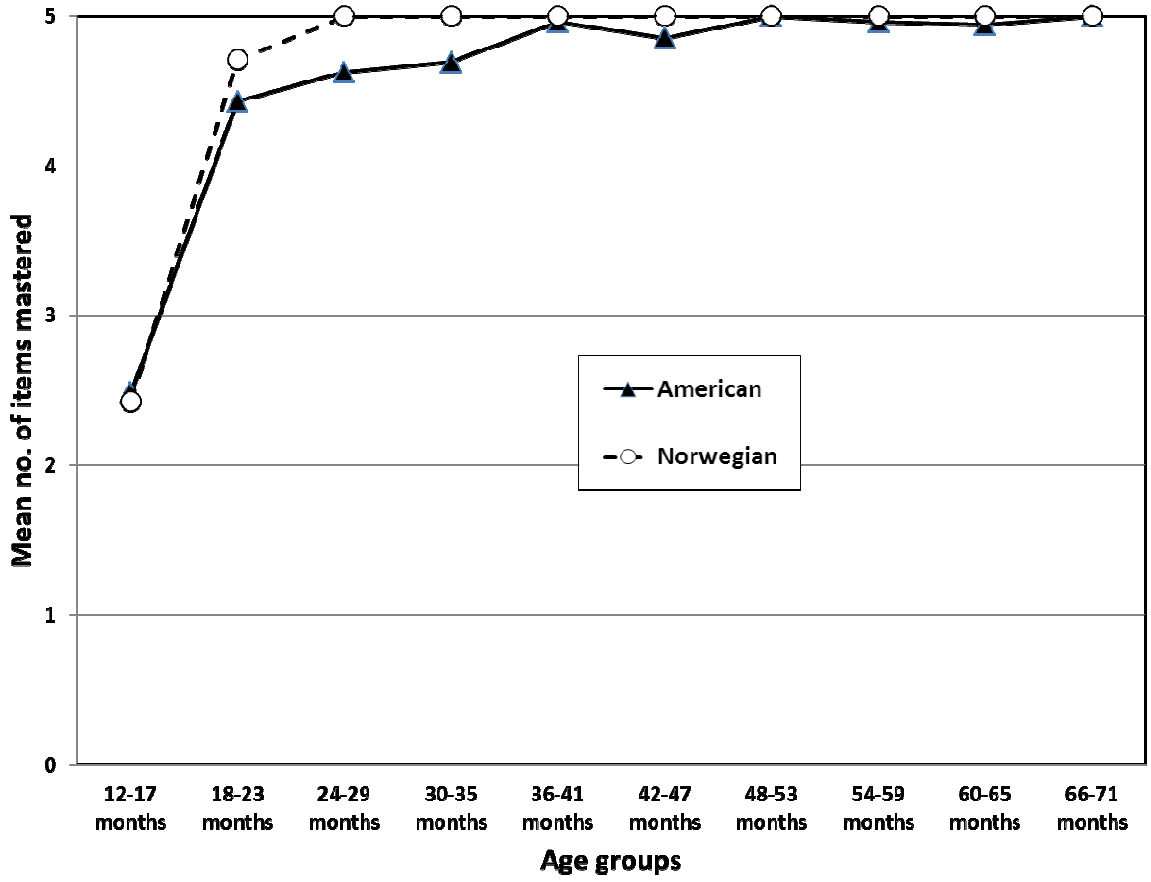

Figure 32: Age means for Outdoor Locomotion – Distance/Speed subdomain in two samples

Table 31: ANOVA of Outdoor Locomotion – Distance/Speed subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOJ

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	291.688[a]	19	15.352	21.723	.000
Intercept	10951.087	1	10951.087	15495.543	.000
Nation	1.765	1	1.765	2.497	.115
Group	233.866	9	25.985	36.768	.000
Nation * group	2.522	9	.280	.397	.937
Error	365.377	517	.707		
Total	12240.000	537			
Corrected Total	657.065	536			

a. R Squared = .444 (Adjusted R Squared = .423)

K. Outdoor Surfaces (5 items)

No significant difference between the two samples.

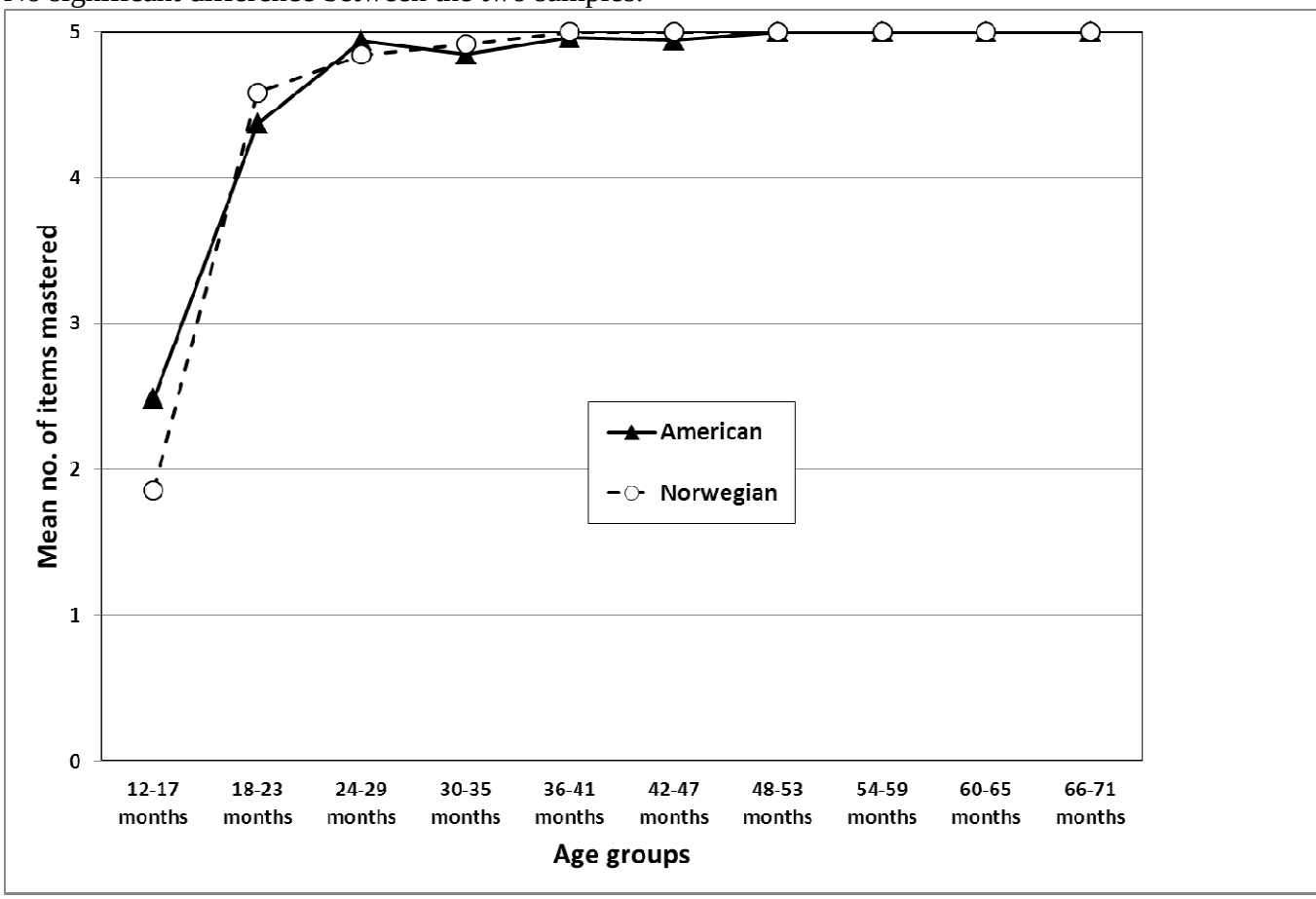

Figure 33: Age means for Outdoor Surfaces subdomain in two samples

Table 32: ANOVA of *Outdoor Surfaces* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOK

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	337.476[a]	19	17.762	36.096	.000
Intercept	10867.077	1	10867.077	22084.034	.000
Nation	.149	1	.149	.302	.583
group	296.305	9	32.923	66.906	.000
Nation * group	4.903	9	.545	1.107	.356
Error	254.405	517	.492		
Total	12212.000	537			
Corrected Total	591.881	536			

a. R Squared = .570 (Adjusted R Squared = .554)

L. Up Stairs (5 items)

No significant difference between the two samples.

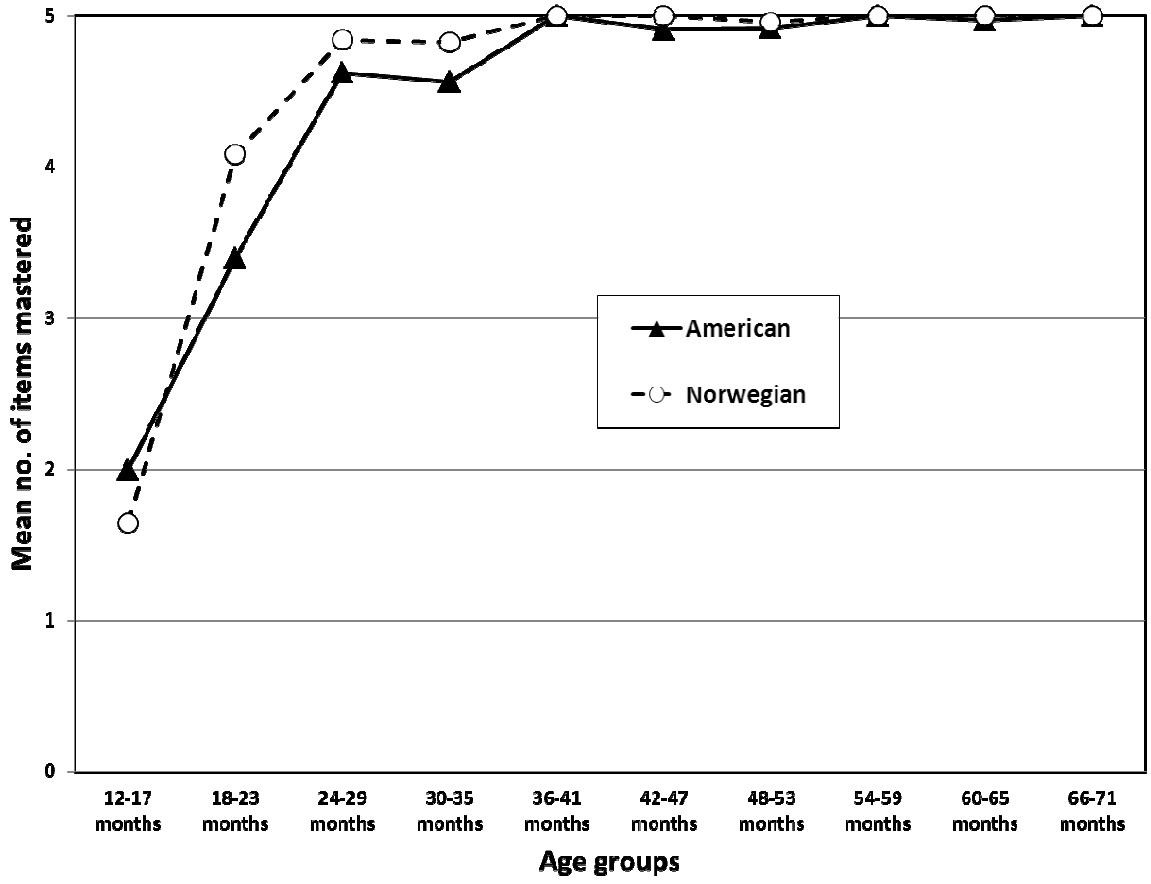

Figure 34: Age means for Up Stairs subdomain in two samples

Table 33: ANOVA of *Up Stairs* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	476.442[a]	19	25.076	61.247	.000
Intercept	10177.769	1	10177.769	24858.647	.000
Nation	1.166	1	1.166	2.847	.092
group	402.509	9	44.723	109.234	.000
Nation * group	7.718	9	.858	2.095	.028
Error	211.673	517	.409		
Total	11486.000	537			
Corrected Total	688.115	536			

a. R Squared = .692 (Adjusted R Squared = .681)

M. Down Stairs (5 items)

No significant difference between the two samples.

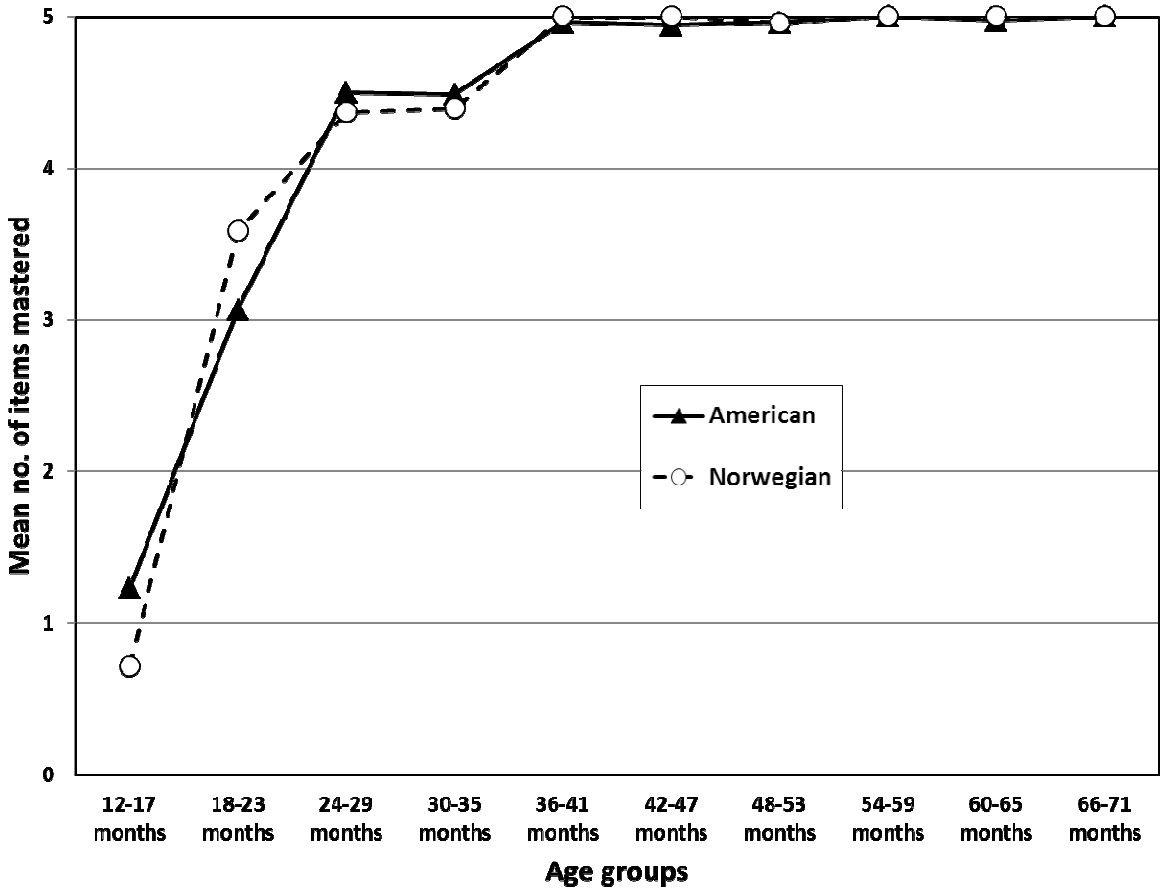

Figure 35: Age means for Down Stairs subdomain in two samples

Table 34: ANOVA of *Down Stairs* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:MOM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	755.327[a]	19	39.754	69.235	.000
Intercept	9374.974	1	9374.974	16327.357	.000
Nation	.013	1	.013	.023	.880
group	651.972	9	72.441	126.163	.000
Nation * group	6.724	9	.747	1.301	.233
Error	296.855	517	.574		
Total	11058.000	537			
Corrected Total	1052.182	536			

a. R Squared = .718 (Adjusted R Squared = .707)

2.3 Mobility items

A. All-over sample differences in Mobility capability

Next, consider the *Mobility* items. Here, 17 out of the 59 items yield significant differences between the "difficulty" scores of the two samples. About half of these differences go in the "unexpected" direction, i.e. the item is most difficult to the US sample – in spite of this sample's higher summed score on the total Mobility scale.

Also here, differences may go either way. While some items are more difficult to the Norwegian children, others are more challenging to the US children. The items that are easiest for the Norwegian sample are marked by asterisks in the table, e.g. items related to walking outdoor or up/down steps.

Table 35: Mobility items; proportion mastering items in two samples (US=313, Norw=224)

Scale	No	Item	% US	% Norw	Fisher *p*
MO	1	Sits supported on toilet	88.5	90.2	0.574
MO	2	Sits unsupported on toilet	85.3	88.4	0.369
MO	3	Climbs/slides low toilet	81.2	81.7	0.911
MO	4	Climbs/slides adult toilet	68.7	64.7	0.353
MO	5	Gets on/off toilet not needing arms	36.7	16.1	**0.000**
MO	6	Sits supported in chair	99.7	100.0	1.000
MO	7	Sits unsupported in chair	99.4	99.1	1.000
MO	8	Climbs on/off low chair	97.4	99.6	0.087
MO	9	* Gets on/off adult chair	88.2	95.1	***0.006***
MO	10	Gets on/off adult chair not needing arms	34.8	24.1	***0.008***
MO	11	Moves in car	92.0	80.4	**0.000**
MO	12	Gets in/out of car	81.4	77.2	0.277

MO	13	Gets in/out of car independently	71.1	71.4	1.000
MO	14	Manages seat belt/restraint	51.1	30.8	***0.000***
MO	15	* Opens/closes car door	37.3	50.0	***0.004***
MO	16	Raises to sitting in bed	100.0	100.0	-
MO	17	Sits and lies down at edge of bed	94.2	87.1	***0.005***
MO	18	Gets in/out of own bed	90.1	81.3	***0.005***
MO	19	Gets in/out of bed not needing arms	47.6	24.6	***0.000***
MO	20	Sits supported in tub	99.7	100.0	1.000
MO	21	Sit unsupported in tub	98.7	99.6	0.407
MO	22	Climbs in/out of tub	82.1	79.0	0.376
MO	23	Sits down/stands up in tub	86.9	90.6	0.218
MO	24	Gets in/out of adult tub	60.7	53.6	0.111
MO	25	Crawls on floor	100.0	100.0	-
MO	26	Walks with support	98.4	100.0	0.079
MO	27	Walks without support	96.2	98.7	0.111
MO	28	Moves in room with difficulty	100.0	100.0	-
MO	29	Moves in room without difficulty	88.0	100.0	0.269
MO	30	Moves between rooms with difficulty	97.8	99.1	0.317
MO	31	Moves between rooms without difficulty	97.1	98.2	0.572
MO	32	Moves and handles doors	78.0	78.6	0.916
MO	33	Changes position on purpose	100.0	100.0	-
MO	34	Moves objects along floor	99.4	100.0	0.513
MO	35	Carries one-hand objects	98.7	99.6	0.407
MO	36	Carries two-hand objects	96.2	97.3	0.628
MO	37	* Carries fragile/spillable	71.9	85.3	***0.000***
MO	38	Walks outdoor with support	96.8	99.1	0.084
MO	39	Walks outdoor without support	94.2	97.3	0.095
MO	40	Moves 10-50 feet outdoor	97.1	97.8	0.786
MO	41	Moves 50-100 feet outdoor	93.6	96.4	0.171
MO	42	* Moves 100-150 feet outdoor	89.8	95.5	***0.015***
MO	43	* Moves 150+ feet with difficulty	87.5	95.5	***0.001***
MO	44	* Moves 150+ feet without difficulty	84.7	95.5	***0.000***
MO	45	Walks level surfaces	97.4	97.8	1.000
MO	46	Walks uneven surfaces	94.6	96.9	0.289
MO	47	Walks rough surfaces	92.0	96.0	0.073
MO	48	Walks up/down inclines	90.1	92.9	0.283
MO	49	Walks up/down curbs	85.0	90.2	0.089
MO	50	* Crawls up 1-11 steps	97.8	100.0	***0.045***
MO	51	Crawls up 12-15 steps	95.8	97.8	0.331
MO	52	Walks up 1-11 steps	87.2	91.5	0.126
MO	53	* Walks up 12-15 steps with difficulty	81.8	88.8	***0.028***
MO	54	* Walks up 12-15 steps without difficulty	73.5	87.5	***0.000***
MO	55	Crawls down 1-11 steps	94.2	96.0	0.427
MO	56	Crawls down 12-15 steps	91.4	94.2	0.246
MO	57	Walks down 1-11 steps	85.0	87.9	0.375
MO	58	Walks down 12-15 steps with difficulty	80.5	85.7	0.133
MO	59	* Walks down 12-15 steps without difficulty	70.3	82.1	***0.002***

B. Age and sample differences on Mobility single items

Also with the mobility items, there may be interesting discrepancies between results at the sub-domain and the single-item level. The five items of the subdomain *Indoor Locomotion – Pulls/Carries Objects* (MO33 through MO37, cf. figure 36 above) give an example of this.

Plots in figure 36 below show the percentage mastering the item within each age group. And again disregarding measurement assumptions, ANOVAs on binary scores were performed to gain an initial impression of the *age* and *sample* effects.

Results are instructive. All subjects in both samples mastered the first item (MO33). Consequently, the figure makes no sense, and is not shown. And, obviously, neither age nor sample differences exist. The following three items (MO34 through MO36) are mastered by all but the youngest group. While this makes understandable figures, it yields no age effects. There also is no sample effect with item MO34, but items MO35 and MO36 do show one. The graphs show, however, that this "general" effect is due to differences within one or two age groups only.

The fifth item in this subdomain (MO37: Carries fragile or spillable objects), displays a more familiar pattern. Here, there is a clear age effect as well as a sample difference.

However, all these five items were used to form one subdomain score (Indoor Locomotion – Pulls /Carries Objects, cf. figure 30 and table 29 above). And, unfortunately, ANOVA of this combined score indicate sample as well as age effects for the subdomain.

This subdomain, therefore, is a case of misleading grouping of different tendencies. Lumping four items without age differences with one case containing such differences, we obtain summed scores suggesting that age differences are important throughout the entire domain. Clearly, results may look different, depending on the level of analysis.

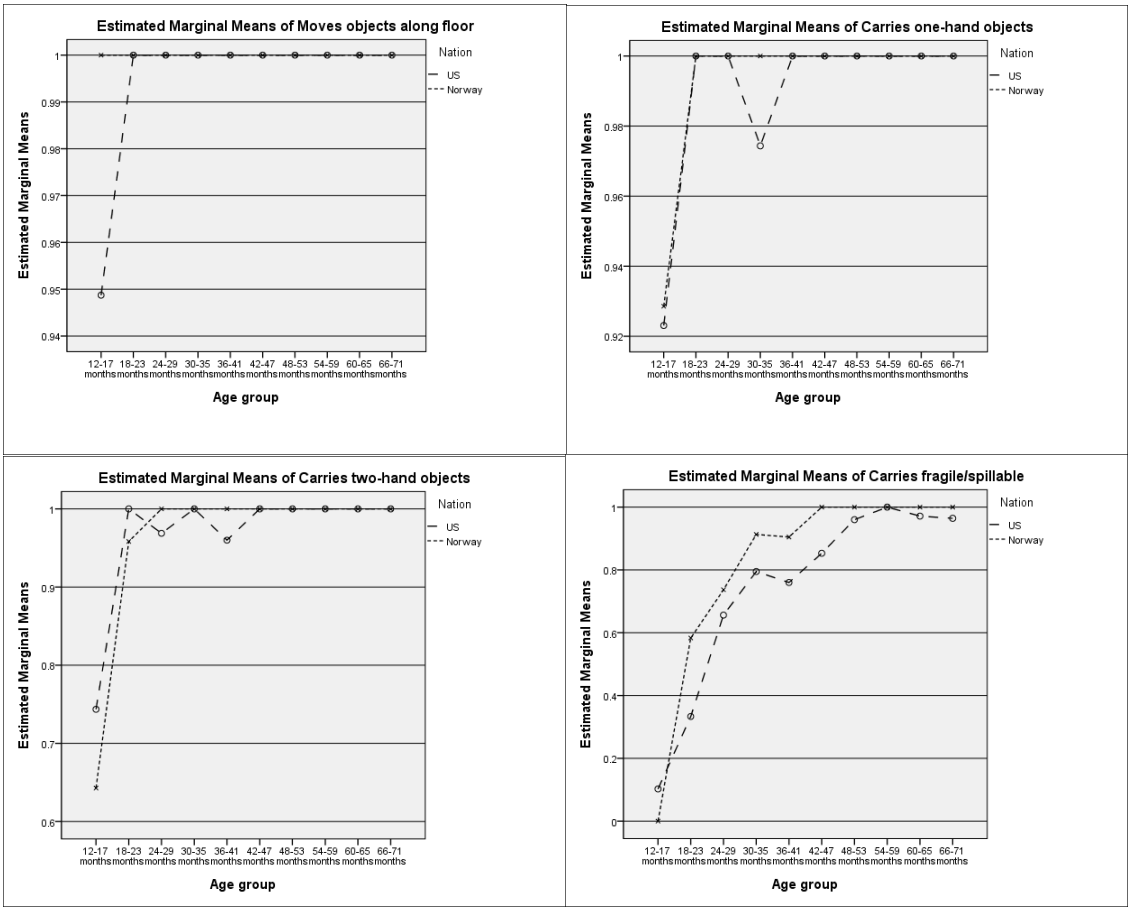

Figure 36: Per cent mastering the four last items of the Indoor Locomotion – Pulls /Carries Objects subdomain, in two samples with ten age groups

C. Sample differences of item difficulty

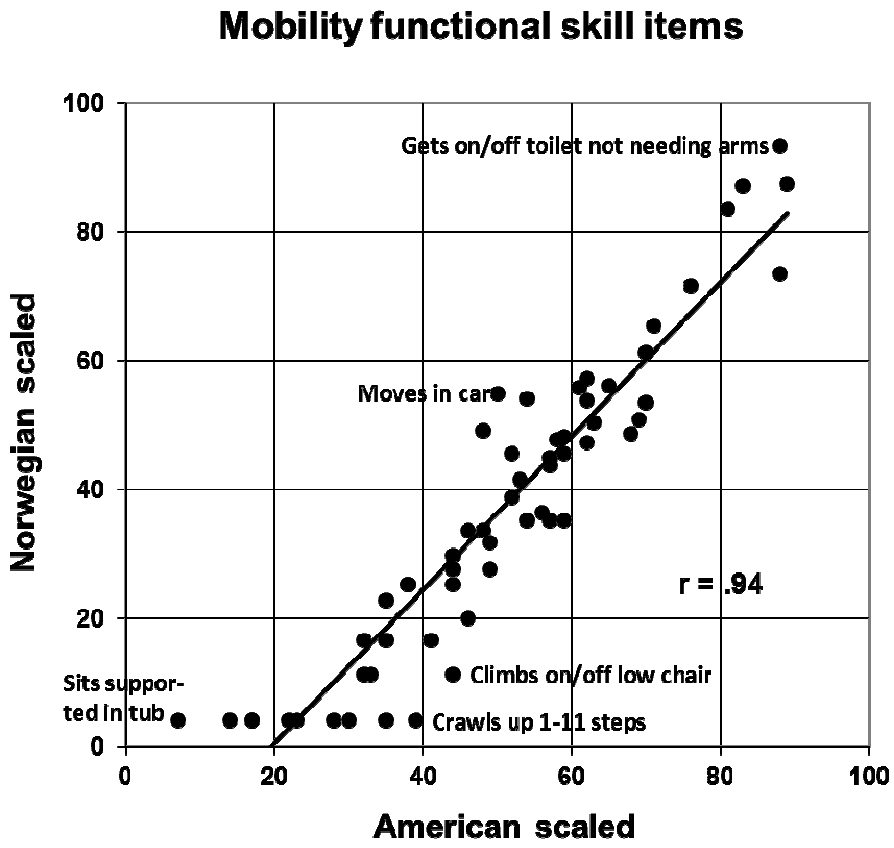

Figure 37: Mobility item difficulty scores in the two normative samples

Items are not very far from the regression line, suggesting that their relative placement is not very different in the two samples.

To indicate the content of interesting items, some texts have been added. Some items are more difficult to the American sample (*Gets on/off toilet not needing arms, Moves in car, Sits supported in tub*), while others are easier (*Crawls up 1-11 steps, Climbs on/off low chair*).

Also here, the regression line passes through the Y axis well below its zero point. Hence, a regression equation attempting to predict the Norwegian scores from their American counterparts will include some negative constant. In simpler terms, this implies that item difficulty scores are *generally* higher in the American than in the Norwegian sample.

D. DIF-tests of Mobility item difficulty differences

In figure 38, the item difficulty scores from the two normative samples are plotted. Please cf. table 35 for viewing the content of the different item numbers.

Also here, many items have rather similar scores in the two samples, falling close to the dotted straight regression line. Quite a few items, however, have appreciably different scores in the two samples. They fall outside the 95% confidence interval. Examples of this are, e.g., item 11 (Moves in car) and item 44 (Moves 150+ feet without difficulty).

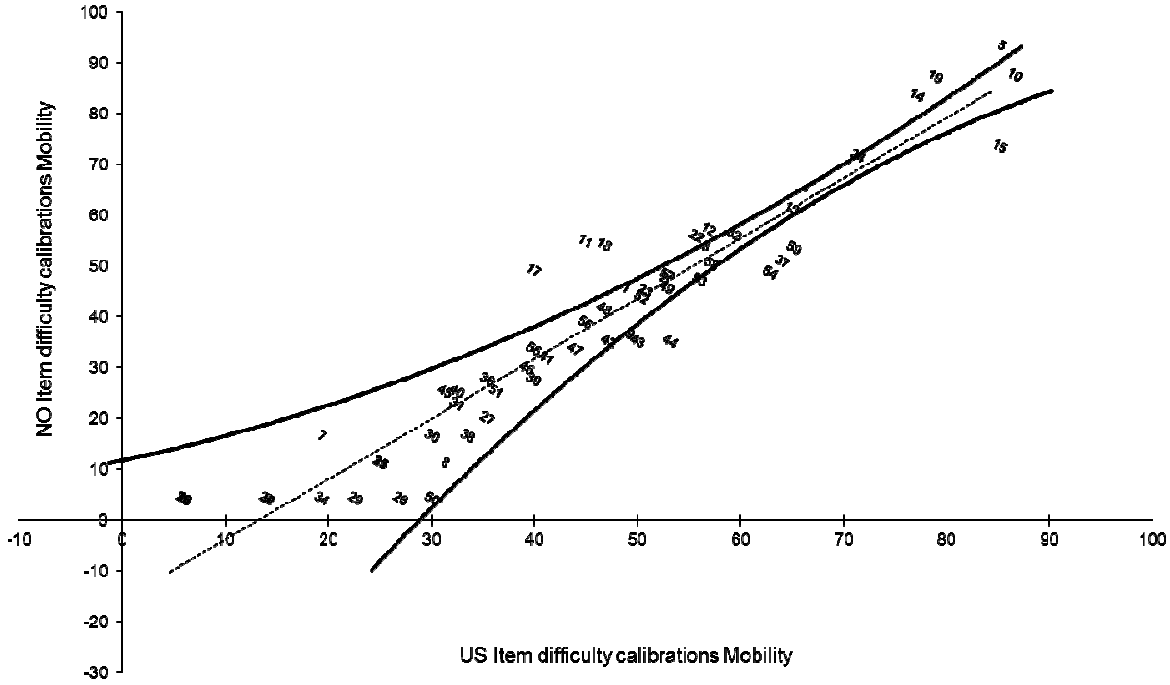

Figure 38: DIF analysis of Mobility items in the two normative samples

The numbers in table 36 confirm that the differences between the samples are substantial, listing the items displaying t-values >2.

It is worth observing that differences are relatively numerous, and go both ways. Also, they generally correspond fairly well to the differences mapped in table 35, including items related to walking outdoor or up/down steps.

Table 36: Mobility items with significant sample differences

Norwegian			***American***		
Item #	*Item name*	*DIF t-value >2*	*Item #*	*Item name*	*DIF t-value >2*
11	Moves in car	6.86	15	Opens/closes car door	6.73
18	Gets in/out of own bed	5.79	54	Walks up 12-15 stepsw.o. diff.	5.07
19	Gets in/out of bed not needing arms	5.70	37	Carries fragile /spillable	4.79
17	Sits and lies down at edge of bed	5.61	44	Moves 150+ feet w.o. diff.	4.44
14	Manages seat belt /restraint	4.89	59	Walks down 12-15 steps w.o. diff.	4.17
22	Climbs in/out of tub	3.54	43	Moves 150+ feet with difficulty	3.28
12	Gets in/out of car	2.65	9	Gets on/off adult chair	2.70
			42	Moves 100-150 feet outdoor	2.27
			53	Walks up 12-15 steps with diff.	2.08

3. Social function

3.1 Social function domain

With the Social functions scale (Max. score = 65), the differences between the US and the Norwegian materials appear even smaller, and clearly less consistent.

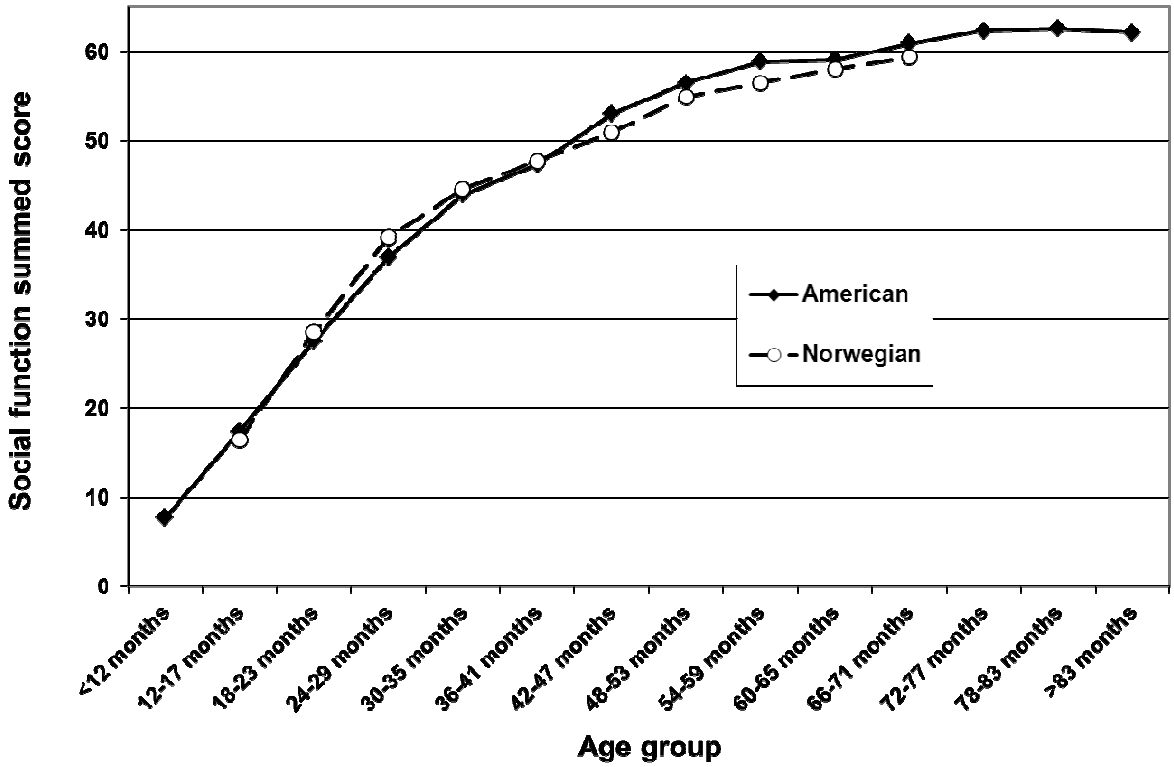

Figure 39: Age means for Social function domain raw score in two samples

This impression is confirmed by the ANOVA, showing no significant difference between the two *national* samples. The magnitude of the significant *age group* effect, however, appears to be comparable to that of the first two scales.

Table 37: ANOVA of Social Function raw score in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:Social functions summed raw score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	100388.565[a]	19	5283.609	184.005	.000
Intercept	1065101.009	1	1065101.009	37092.716	.000
group	89073.990	9	9897.110	344.672	.000
Nation	36.676	1	36.676	1.277	.259
group * Nation	258.892	9	28.766	1.002	.437
Error	14845.427	517	28.715		
Total	1251710.000	537			
Corrected Total	115233.993	536			

a. R Squared = .871 (Adjusted R Squared = .866)

3.2 Social subdomains

A. Comprehension of Word Meanings (5 items)

Norwegian sample scores higher.

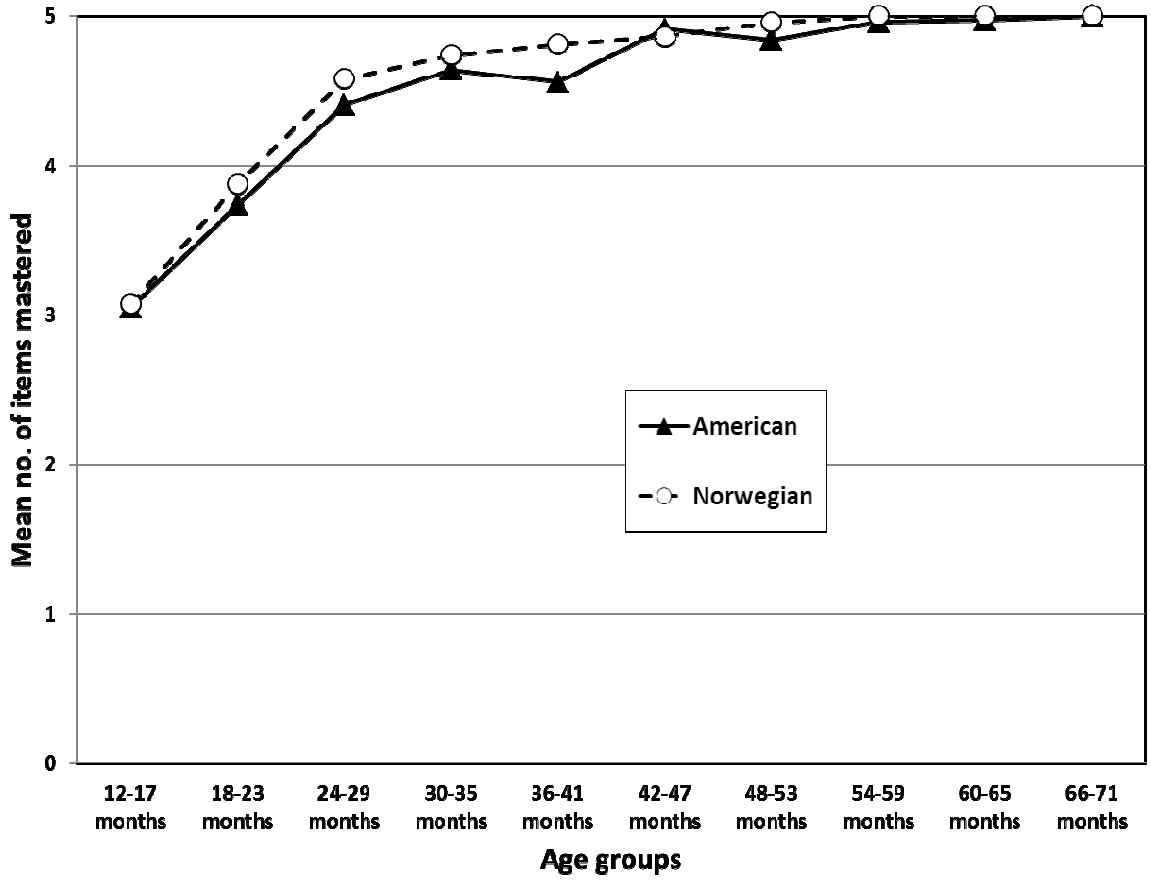

Figure 40: Age means for Comprehension of Word Meanings subdomain in two samples

Table 38: ANOVA of Comprehension of Word Meanings subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOA

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	196.986[a]	19	10.368	47.836	.000
Intercept	10458.114	1	10458.114	48252.677	.000
Nation	.848	1	.848	3.914	.048
group	166.338	9	18.482	85.274	.000
Nation * group	.907	9	.101	.465	.898
Error	112.053	517	.217		
Total	11414.000	537			
Corrected Total	309.039	536			

a. R Squared = .637 (Adjusted R Squared = .624)

B. Comprehension of Sentence Complexity (5 items)

US sample scores higher.

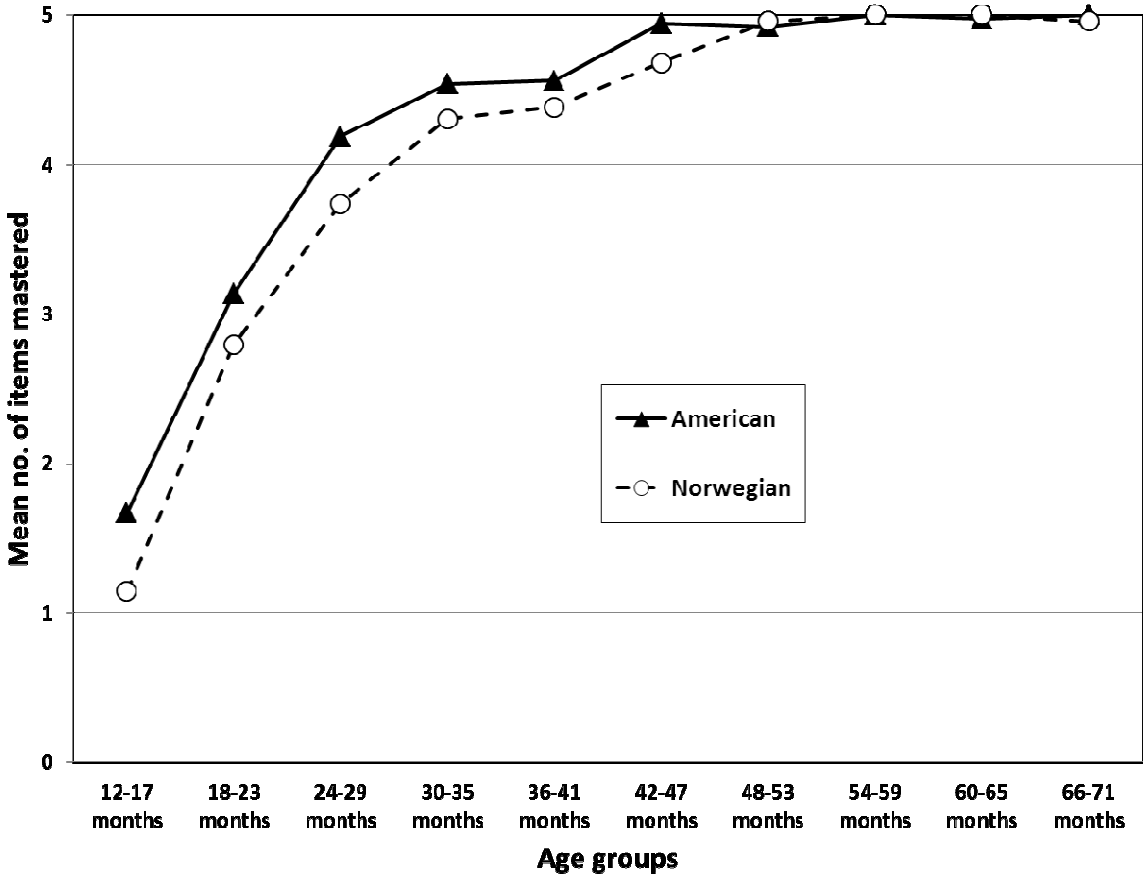

Figure 41: Age means for Comprehension of Sentence Complexity subdomain in two samples

Table 39: ANOVA of *Comprehension of Sentence Complexity* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOB

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	628.860[a]	19	33.098	70.296	.000
Intercept	8889.631	1	8889.631	18880.495	.000
Nation	4.871	1	4.871	10.346	.001
group	559.798	9	62.200	132.105	.000
Nation * group	4.494	9	.499	1.061	.391
Error	243.423	517	.471		
Total	10468.000	537			
Corrected Total	872.283	536			

a. R Squared = .721 (Adjusted R Squared = .711)

C. Functional Use of Communication (5 items)

US sample scores higher.

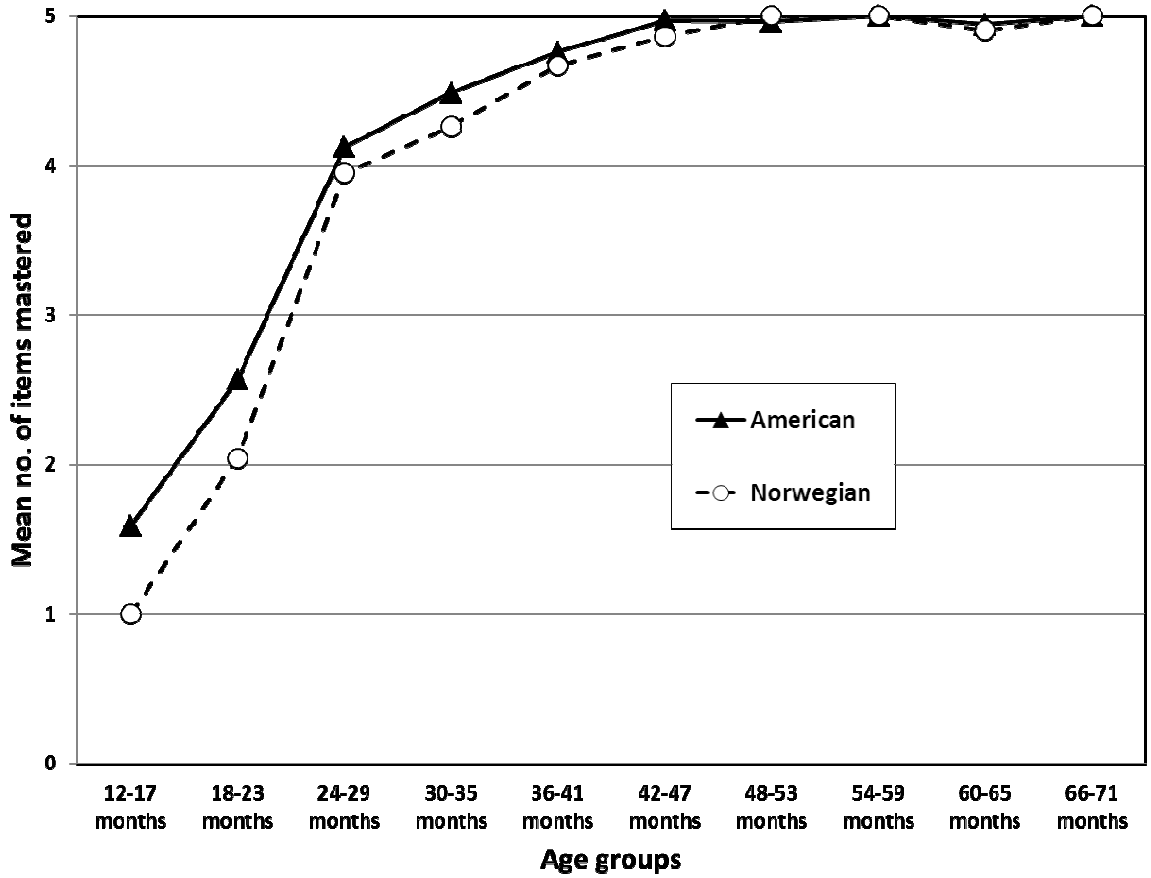

Figure 42: Age means for Functional Use of Communication subdomain in two samples

Table 40: ANOVA of Functional Use of Communication subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	776.108[a]	19	40.848	74.694	.000
Intercept	8723.326	1	8723.326	15951.460	.000
Nation	3.732	1	3.732	6.825	.009
group	709.638	9	78.849	144.183	.000
Nation * group	5.196	9	.577	1.056	.394
Error	282.730	517	.547		
Total	10436.000	537			
Corrected Total	1058.838	536			

a. R Squared = .733 (Adjusted R Squared = .723)

D. Complexity of Expressive Communication (5 items)

No significant difference between the two samples.

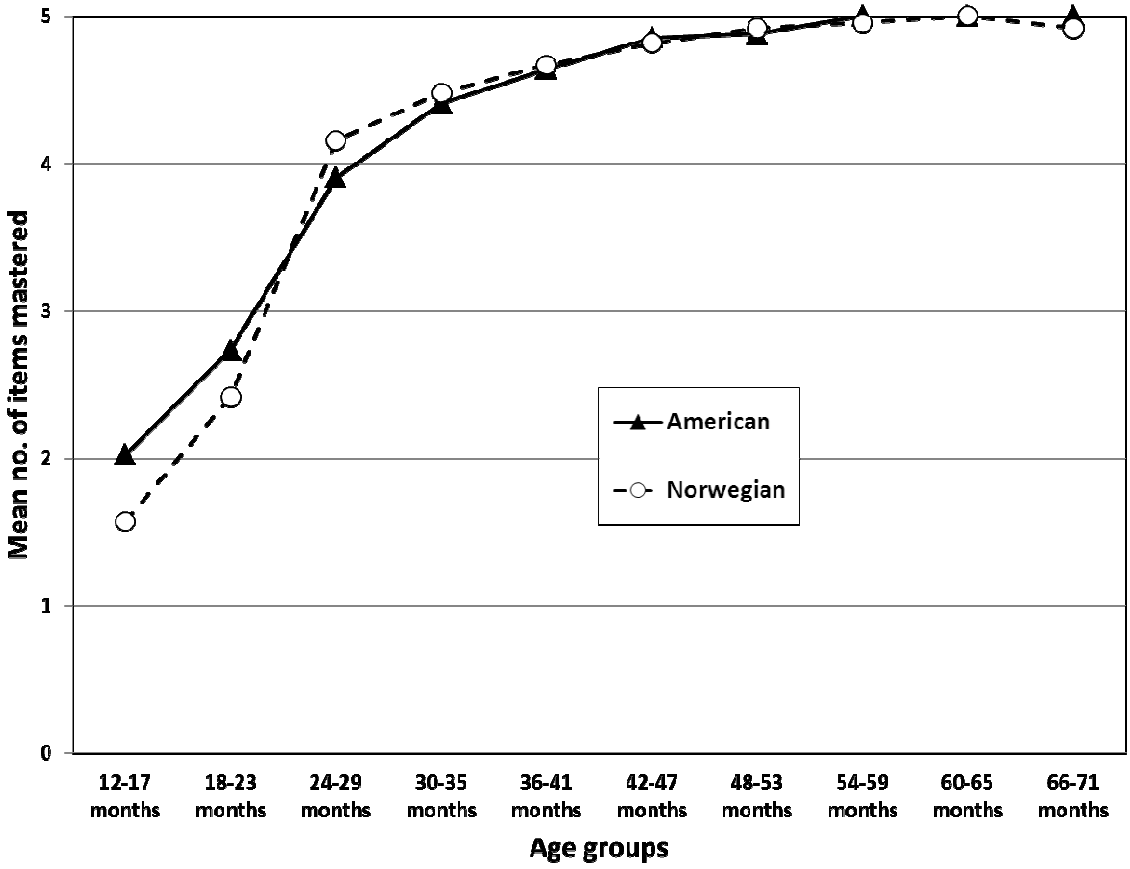

Figure 43: Age means for Complexity of Expressive Communication subdomain in two samples

Table 41: ANOVA of *Complexity of Expressive Communication* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOD

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	585.619[a]	19	30.822	90.669	.000
Intercept	8989.897	1	8989.897	26445.639	.000
Nation	.384	1	.384	1.130	.288
group	535.890	9	59.543	175.159	.000
Nation * group	4.135	9	.459	1.352	.207
Error	175.748	517	.340		
Total	10374.000	537			
Corrected Total	761.367	536			

a. R Squared = .769 (Adjusted R Squared = .761)

E. Problem Resolution (5 items)

No significant difference between the two samples.

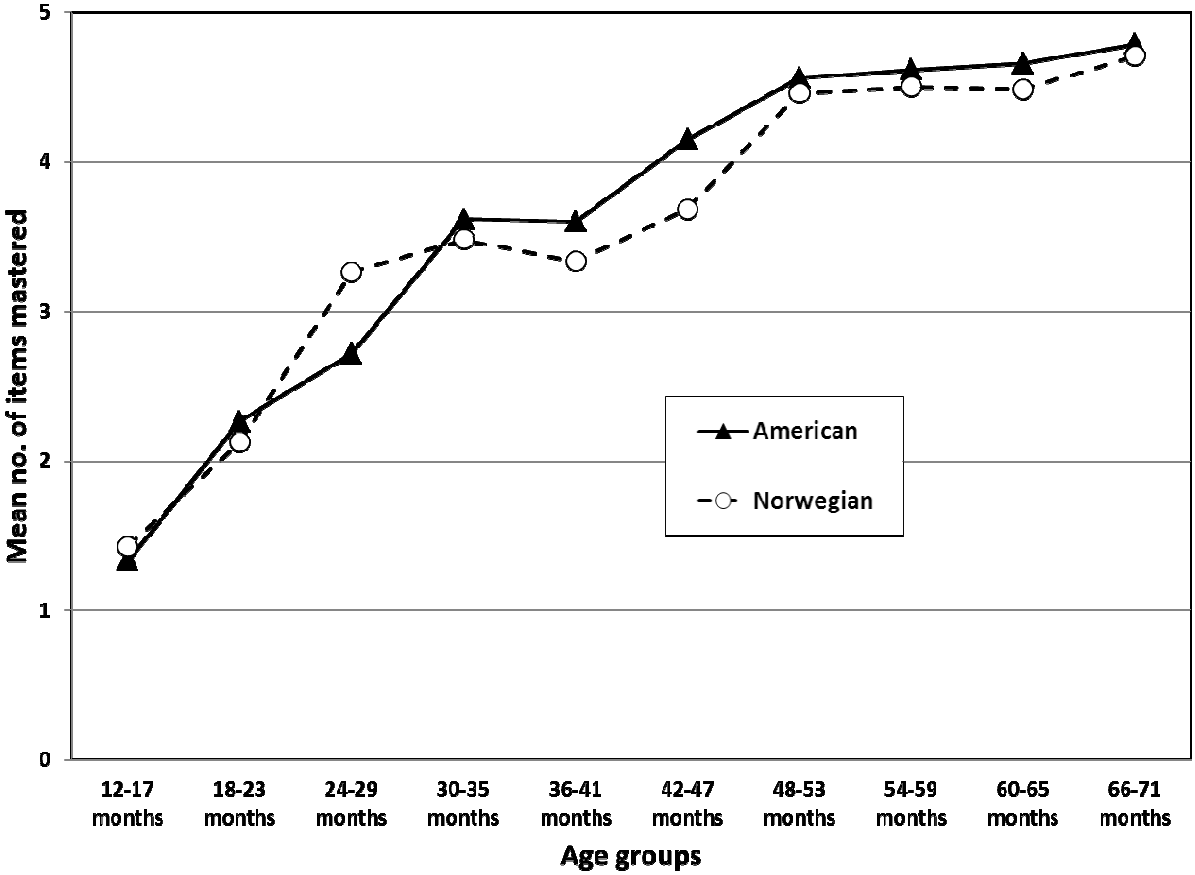

Figure 44: Age means for Problem Resolution subdomain in two samples

Table 42: ANOVA of *Problem Resolution* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	628.631[a]	19	33.086	44.031	.000
Intercept	6507.260	1	6507.260	8659.902	.000
Nation	.889	1	.889	1.183	.277
group	544.894	9	60.544	80.572	.000
Nation * group	7.579	9	.842	1.121	.346
Error	388.486	517	.751		
Total	7968.000	537			
Corrected Total	1017.117	536			

a. R Squared = .618 (Adjusted R Squared = .604)

F. Social Interactive Play (5 items)

The US sample scores higher in some age groups, the Norwegian sample in others.

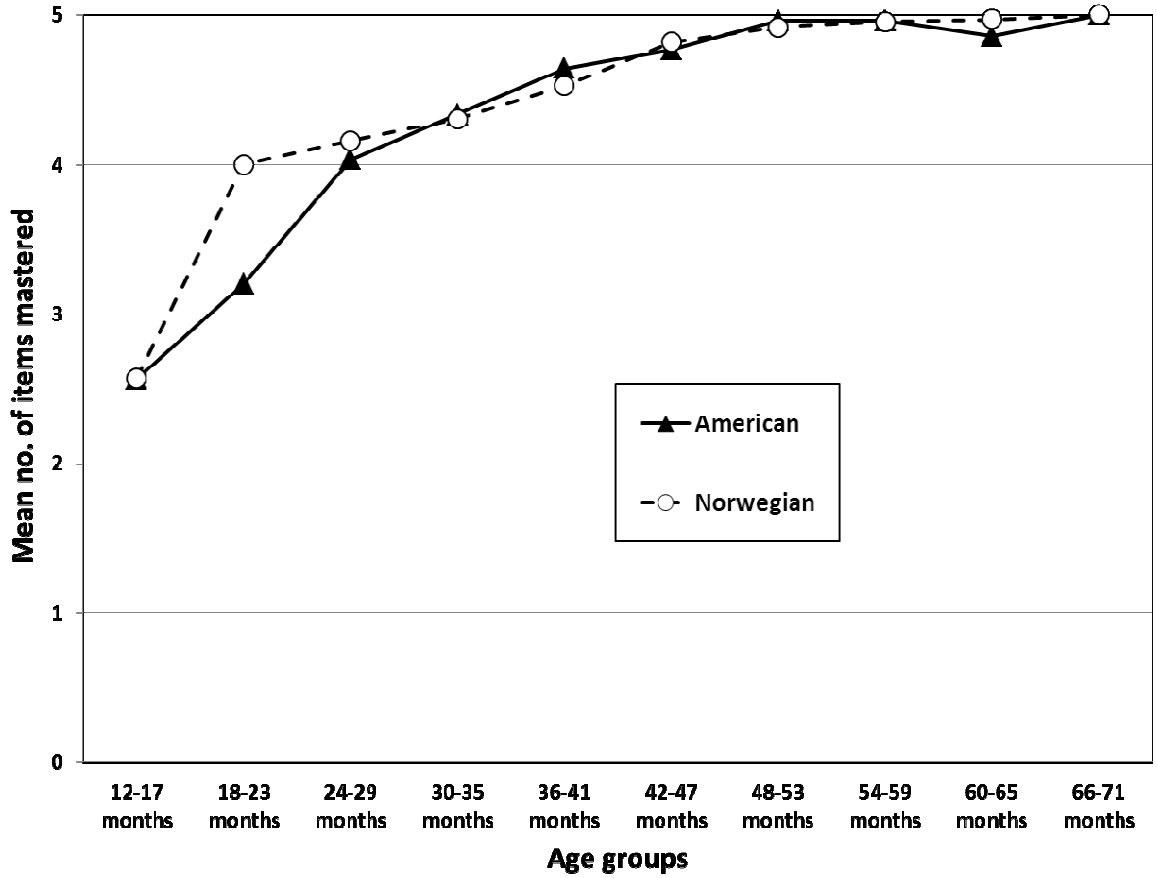

Figure 45: Age means for Social Interactive Play subdomain in two samples

Table 43: ANOVA of *Social Interactive Play* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOF

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	306.054[a]	19	16.108	41.807	.000
Intercept	9680.838	1	9680.838	25125.347	.000
Nation	1.029	1	1.029	2.672	.103
group	251.738	9	27.971	72.595	.000
Nation * group	7.951	9	.883	2.293	.016
Error	199.201	517	.385		
Total	10763.000	537			
Corrected Total	505.255	536			

a. R Squared = .606 (Adjusted R Squared = .591)

G. Peer Interaction (5 items)

No significant difference between the two samples.

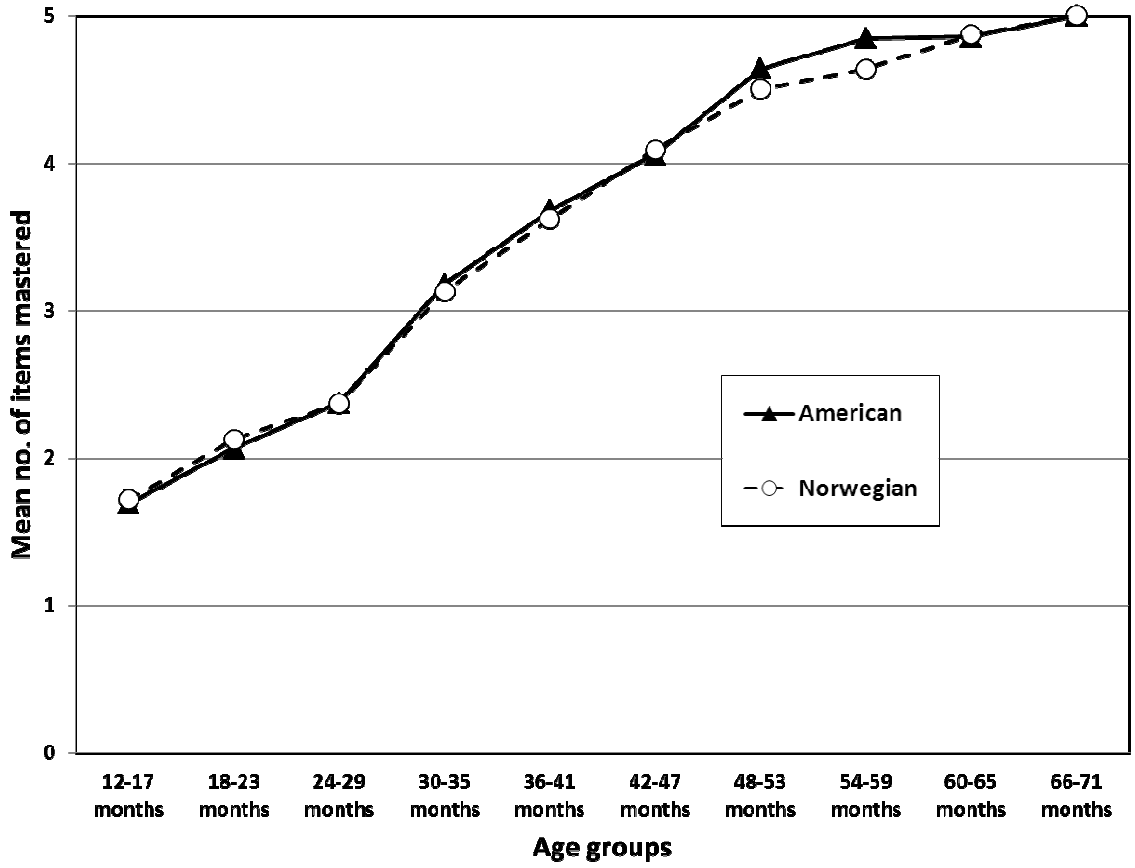

Figure 46: Age means for Peer Interaction subdomain in two samples

Table 44: ANOVA of *Peer Interaction* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	733.227[a]	19	38.591	82.435	.000
Intercept	6633.166	1	6633.166	14169.222	.000
Nation	.146	1	.146	.312	.577
group	675.543	9	75.060	160.338	.000
Nation * group	.782	9	.087	.186	.996
Error	242.028	517	.468		
Total	8049.000	537			
Corrected Total	975.255	536			

a. R Squared = .752 (Adjusted R Squared = .743)

H. Play with Objects (5 items)

No significant difference between the two samples.

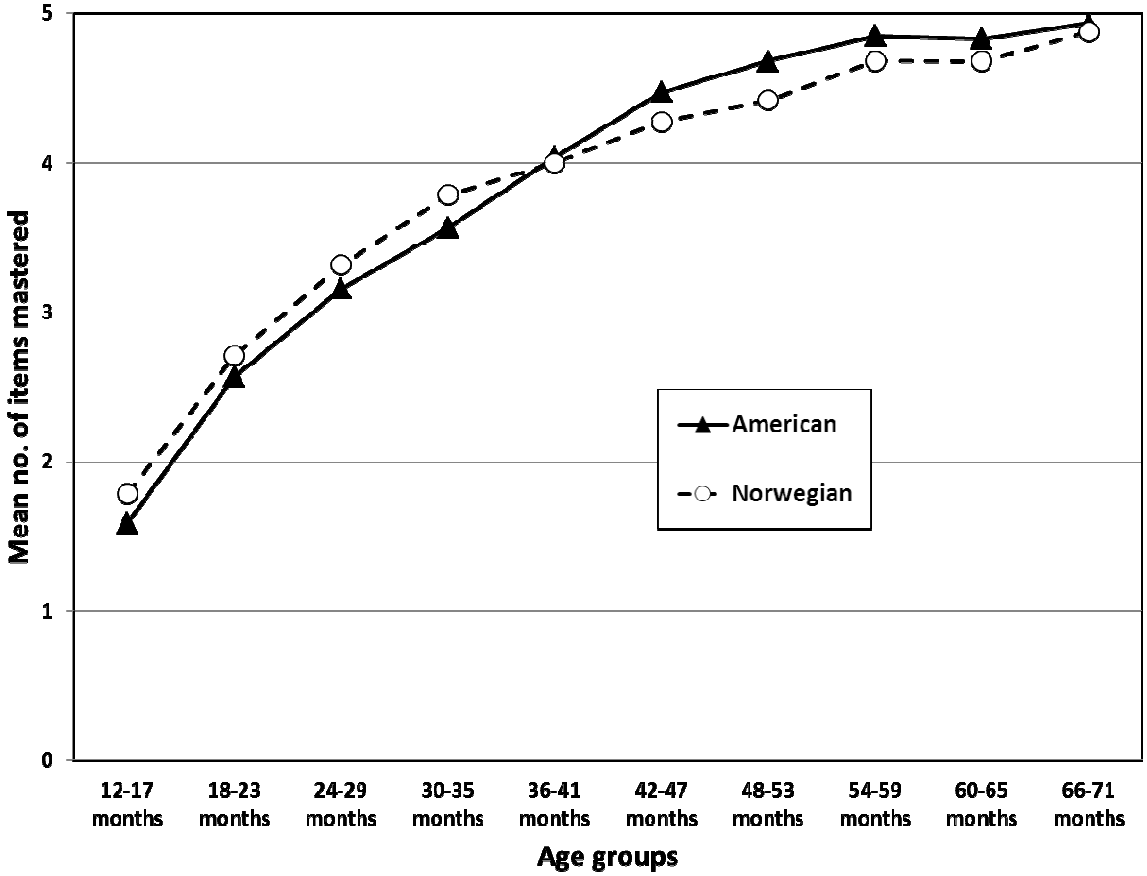

Figure 47: Age means for Play with Objects subdomain in two samples

Table 45: ANOVA of *Play with Objects* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOH

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	560.562[a]	19	29.503	66.492	.000
Intercept	7528.651	1	7528.651	16967.545	.000
Nation	.030	1	.030	.068	.794
group	478.984	9	53.220	119.944	.000
Nation * group	3.734	9	.415	.935	.494
Error	229.398	517	.444		
Total	8777.000	537			
Corrected Total	789.959	536			

a. R Squared = .710 (Adjusted R Squared = .699)

I. Self -Information (5 items)

Norwegian sample scores higher.

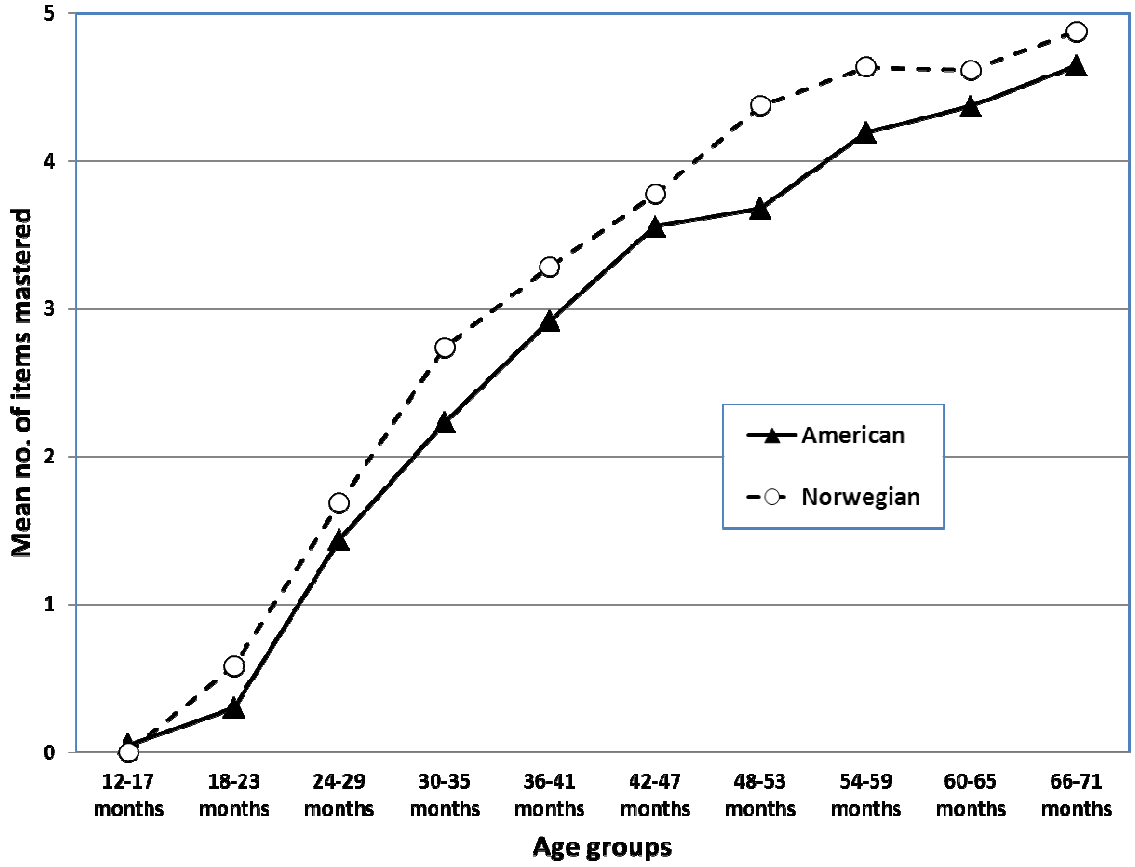

Figure 48: Age means for Self Information subdomain in two samples

Table 46: ANOVA of *Self Information* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOI

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1437.556[a]	19	75.661	128.456	.000
Intercept	4243.548	1	4243.548	7204.622	.000
Nation	12.774	1	12.774	21.687	.000
group	1305.098	9	145.011	246.197	.000
Nation * group	4.295	9	.477	.810	.607
Error	304.515	517	.589		
Total	6216.000	537			
Corrected Total	1742.071	536			

a. R Squared = .825 (Adjusted R Squared = .819)

J. Time Orientation (5 items)

US sample scores higher, mainly with older age groups.

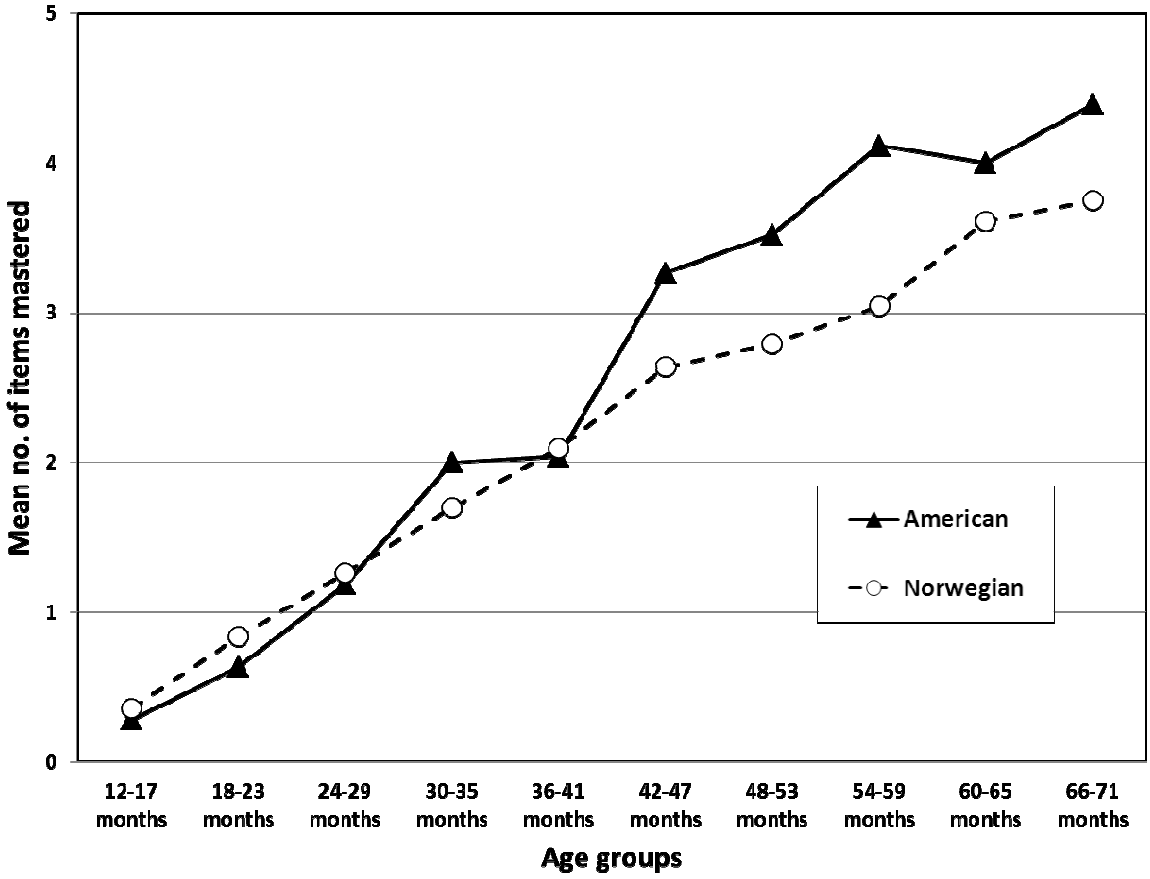

Figure 49: Age means for Time Orientation subdomain in two samples

Table 47: ANOVA of *Time Orientation* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOJ

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	917.842[a]	19	48.307	59.513	.000
Intercept	2853.155	1	2853.155	3514.954	.000
Nation	14.223	1	14.223	17.522	.000
Group	798.960	9	88.773	109.365	.000
Nation * group	20.050	9	2.228	2.744	.004
Error	419.659	517	.812		
Total	4446.000	537			
Corrected Total	1337.501	536			

a. R Squared = .686 (Adjusted R Squared = .675)

K. Household Chores (5 items)

US sample scores higher, mainly with older age groups.

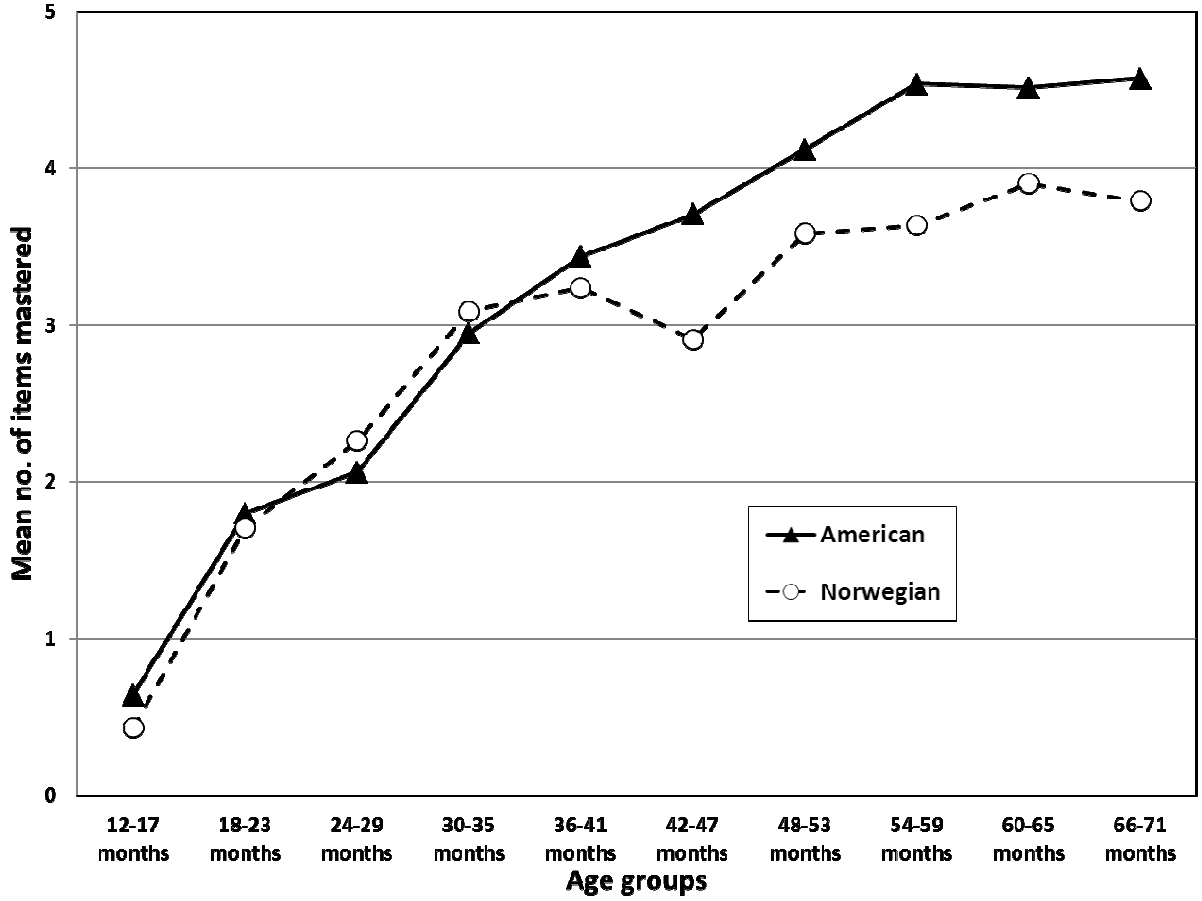

Figure 50: Age means for Household Chores subdomain in two samples

Table 48: ANOVA of *Household Chores* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOK

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	751.583[a]	19	39.557	47.059	.000
Intercept	4685.323	1	4685.323	5573.859	.000
Nation	18.185	1	18.185	21.634	.000
group	630.275	9	70.031	83.311	.000
Nation * group	18.474	9	2.053	2.442	.010
Error	434.584	517	.841		
Total	6256.000	537			
Corrected Total	1186.168	536			

a. R Squared = .634 (Adjusted R Squared = .620)

L. Self-Protection (5 items)

US sample scores higher, mainly with older age groups.

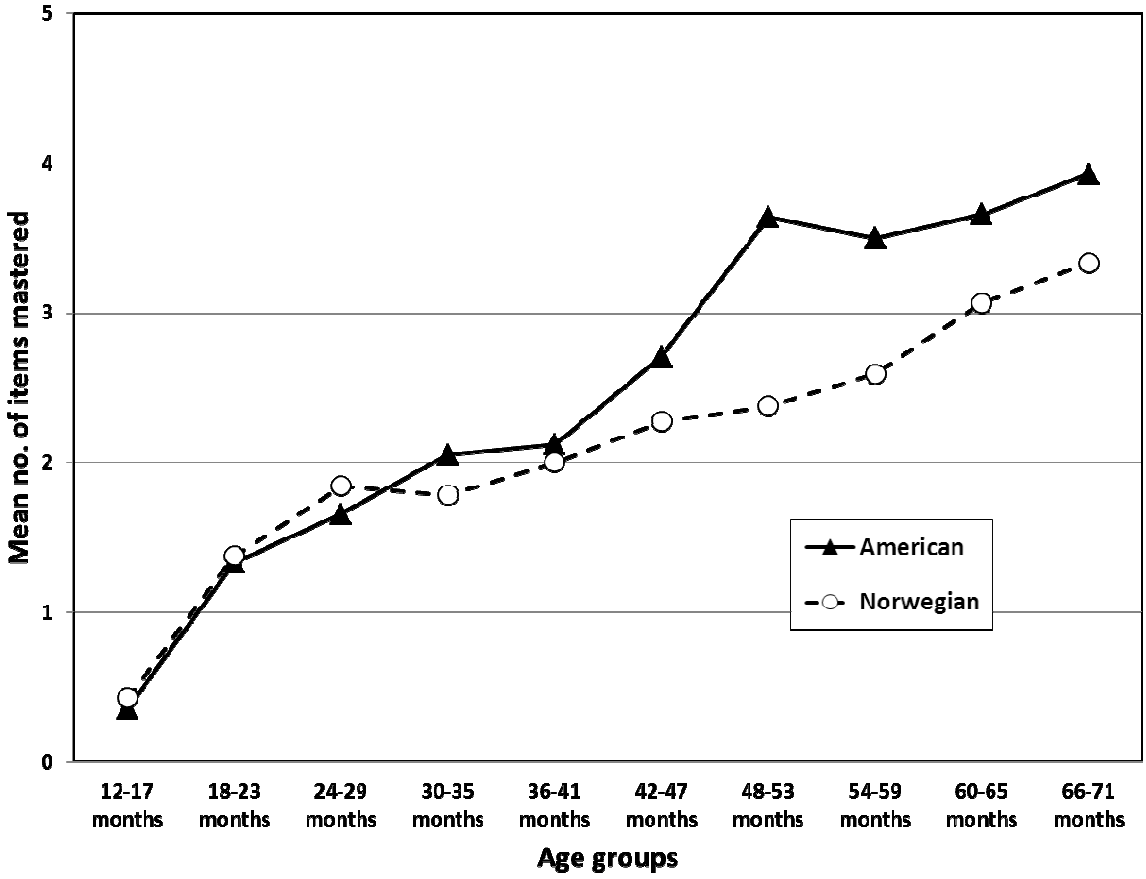

Figure 51: Age means for Self Protection subdomain in two samples

Table 49: ANOVA of *Self Protection* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOL

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	545.157[a]	19	28.692	46.580	.000
Intercept	2675.800	1	2675.800	4343.954	.000
Nation	19.089	1	19.089	30.990	.000
group	439.719	9	48.858	79.317	.000
Nation * group	23.477	9	2.609	4.235	.000
Error	318.463	517	.616		
Total	3764.000	537			
Corrected Total	863.620	536			

a. R Squared = .631 (Adjusted R Squared = .618)

M. Community Function (5 items)

Norwegian sample scores higher.

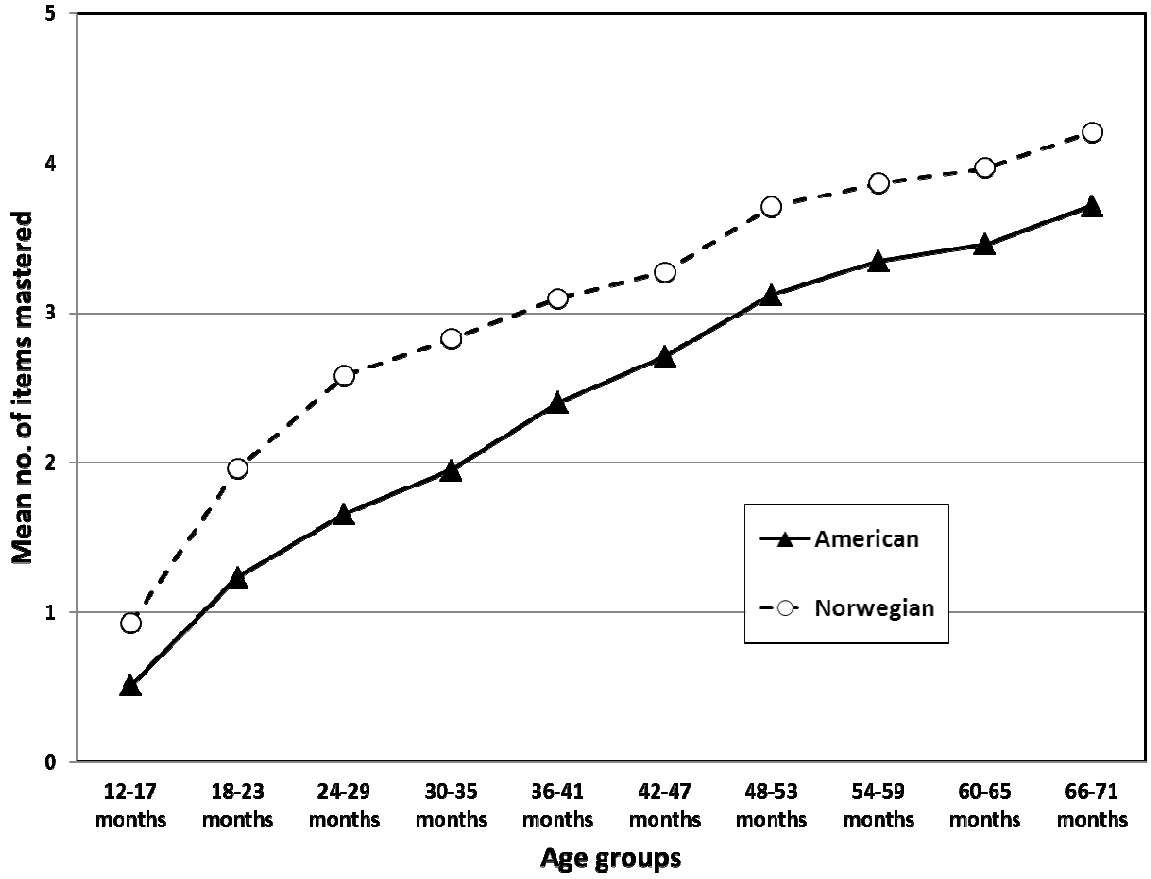

Figure 52: Age means for Community Function subdomain in two samples

Table 50: ANOVA of *Community Function* subdomain in two samples and ten age groups

Tests of Between-Subjects Effects

Dependent Variable:SOM

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	600.939[a]	19	31.628	58.258	.000
Intercept	3753.750	1	3753.750	6914.210	.000
Nation	50.368	1	50.368	92.775	.000
Group	463.211	9	51.468	94.801	.000
Nation * group	3.245	9	.361	.664	.742
Error	280.681	517	.543		
Total	4727.000	537			
Corrected Total	881.620	536			

a. R Squared = .682 (Adjusted R Squared = .670)

3.3 Social Function items

A. All-over sample differences in Social Function capability

The single items of the *Social Function* scale are next. Again, sample differences go both ways. Out of the 65 items, 24 yield significant differences between the two samples. Nineteen of these differences mean that items are easier for the Norwegian sample. These items are marked with an asterisk. As we have seen previously, however, there is no significant sample difference for the total summed Social Function scale.

Table 51: Social Function items; proportion mastering items in two samples (US=313, Norw=224)

Scale	No	Item	US	NO	Fisher's p
SF	1	Orients to sound	99.7	100.0	1.000
SF	2	Responds to "No"	99.7	99.6	1.000
SF	3	Understands 10 words	96.5	98.9	0.170
SF	4	* Understands talk about relationships	87.2	93.8	***0.013***
SF	5	* Understands talk about time and sequence	63.6	74.1	***0.011***
SF	6	Understands short sentences	97.8	98.7	0.533
SF	7	Understands 1-step commands	94.6	93.8	0.711
SF	8	Understands directions with "where"	86.9	87.0	0.793
SF	9	Understands 2-step commands	76.4	72.8	0.365
SF	10	Understands two sentences in different forms	65.8	71.4	0.189
SF	11	Names things	97.8	92.9	***0.008***
SF	12	Direct or requests	92.0	96.0	0.073
SF	13	Seeks information	82.4	82.6	1.000
SF	14	Describes object or action	74.8	79.0	0.258
SF	15	Tells about feelings/thoughts	68.7	70.5	0.704
SF	16	Gestures with meaning	98.7	98.7	1.000
SF	17	Single word with meaning	96.8	96.4	0.813
SF	18	Two words with meaning	87.5	88.8	0.687
SF	19	4-5 word sentences	74.8	79.9	0.178
SF	20	Connects two thoughts in story	59.4	67.4	0.070
SF	21	Tries to show problem	97.1	99.1	0.132
SF	22	* Tackles only immediate help	84.0	95.5	***0.000***
SF	23	Seeks help, tackles short delay	70.3	69.6	0.924
SF	24	Describes problem/feeling	58.5	62.5	0.372
SF	25	Joins adult in solving problem	44.7	40.2	0.331
SF	26	Awareness and interest in others	100.0	100.0	-
SF	27	* Initiates a familiar play routine	94.2	99.1	***0.002***
SF	28	* Takes turn when cued	88.8	95.1	***0.012***
SF	29	* Attempts to imitate adult's action	85.9	92.9	***0.012***
SF	30	Suggest new steps/ideas	57.5	84.7	0.107
SF	31	Notices presence of other children	100.0	100.0	-
SF	32	Interacts with other children	94.6	97.3	0.135
SF	33	* Tries to work out simple plans for play	64.5	75.4	***0.008***
SF	34	Plans and carries out cooperative activity	54.6	60.7	0.185
SF	35	Plays activities or games with rules	40.9	41.1	1.000
SF	36	Intentional manipulation of things	86.7	100.0	0.144

SF	37	Uses objects to pretend	92.7	96.4	0.090
SF	38	* Makes things from materials	82.1	90.2	***0.009***
SF	39	* Extended pretend play	62.9	71.4	***0.042***
SF	40	Elaborate pretend sequences	41.2	38.8	0.593
SF	41	States first name	79.9	85.3	0.111
SF	42	States first and last name	63.3	66.5	0.465
SF	43	* Provides name and family information	63.3	80.8	***0.000***
SF	44	States full home address	32.6	37.5	0.270
SF	45	* Directs an adult to help return home	23.6	54.9	***0.000***
SF	46	* General awareness of daily routines	83.7	92.4	***0.004***
SF	47	Some awareness of weekly events	55.3	58.9	0.427
SF	48	Simple time concept	55.0	48.2	0.137
SF	49	Associates time with actions/events	38.3	33.9	0.318
SF	50	Regularly checks clock/time	12.5	1.3	***0.000***
SF	51	Helping to care for belongings	92.0	92.0	1.000
SF	52	* Helping with simple household chores	79.9	87.1	***0.036***
SF	53	Initiates care for belongings	64.2	57.6	0.127
SF	54	Initiates simple household chores	49.2	57.6	0.066
SF	55	Consistently performs household task	27.5	5.4	***0.000***
SF	56	* Shows caution around stairs	84.0	91.5	***0.013***
SF	57	* Shows caution around hot or sharp objects	76.7	85.3	***0.015***
SF	58	Crossing the street without safety prompting	39.9	21.4	***0.000***
SF	59	Not accepting rides, food or money from strangers	37.1	21.0	***0.000***
SF	60	Crosses busy street safety without an adult	2.6	2.2	1.000
SF	61	Plays safety without const. watch	90.1	94.6	0.075
SF	62	* Plays outside of home, periodic monitoring only	72.5	90.2	***0.000***
SF	63	* Follows school/community guidelines	50.8	82.1	***0.000***
SF	64	* Functions in community without supervision	15.7	39.7	***0.000***
SF	65	* Makes store transaction without assistance	3.5	9.8	***0.003***

B. Age and sample differences on single Social Function items

To illustrate the complex relations between a Social Function subdomain and its single scores, consider the subdomain of *Time orientation* (items SF46 through SF50, cf. figure 49). Here, the summed subdomain score is highest in the US sample, especially in the older age groups. On the level of single items, however, things look partly different (cf. figure 53).

With the first item (SF46: General awareness of mealtimes/routines), Norwegians do slightly better in the younger groups. But then all respondents master the item from age of 3½ years (42 months), suggesting that this item does not distinguish between older children.

The next two items (SF47: Awareness of weekly events; and SF48: Simple time concept) yield patterns similar to that of the summed subdomain score: Americans do better with the older age groups. Item SF49 (Associates time with actions/events) seems not to be relevant to the three youngest groups, but then the Americans do better from the age of 2½ years (30 months) and onwards.

The last item (SF50: Regularly checks clock/time) may not be suitable for the Norwegian sample. The American children begin handling this challenge from the age of 4½ (54 months), while only a small minority among the Norwegians catches on to it.

Although not directly misleading, the summed subdomain score thus hides interesting facts.

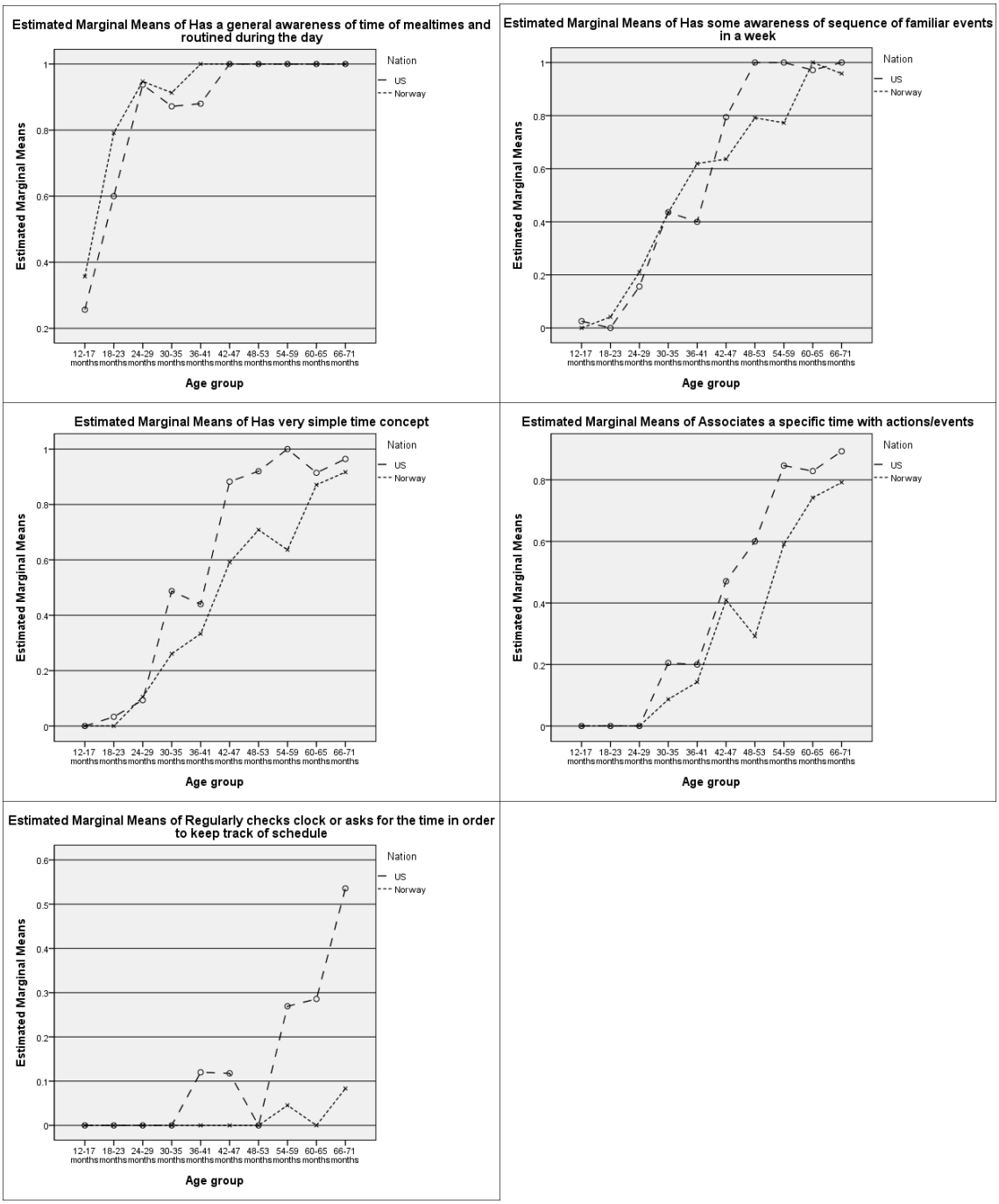

Figure 53: Per cent mastering the five items of the *Time orientation* subdomain, in two samples with ten age groups

C. Sample differences of item difficulty

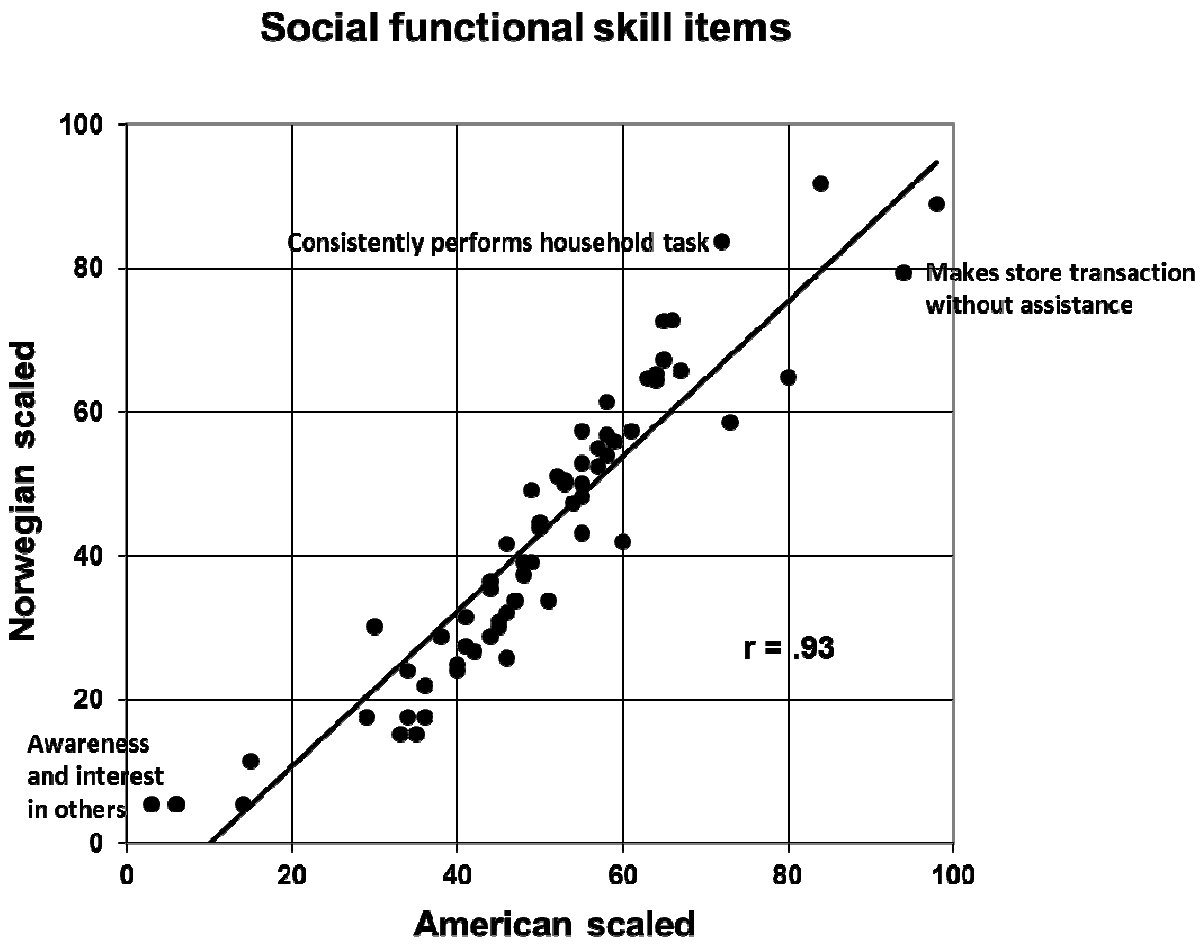

Figure 54: Social function item difficulty scores in the two normative samples

Items are fairly close to the regression line, suggesting that their relative placement is largely similar in the two samples.

And again, some items are not equally difficult in the two samples. More difficult to the American sample is, e.g., *Making store transaction without assistance*, while *Consistently performs household task* yields a higher difficulty score with the Norwegians.

Also here, the regression line suggests that as a whole, the items are more difficult to the Norwegians.

D. DIF-tests of Social Function item difficulty differences

The item difficulty scores from the two normative samples are plotted in figure 55. For the text/content corresponding to the item numbers, please cf. table 51.

Several items have rather similar scores in the two samples, thus falling close to the dotted straight regression line. However, many items clearly yield different scores in the two samples. They fall outside the 95% confidence interval, which is indicated by the "funnel" formed by heavy black lines. Among these are, e.g., item 55 (Consistently performs household task) and item 63 (Follows school/community guidelines).

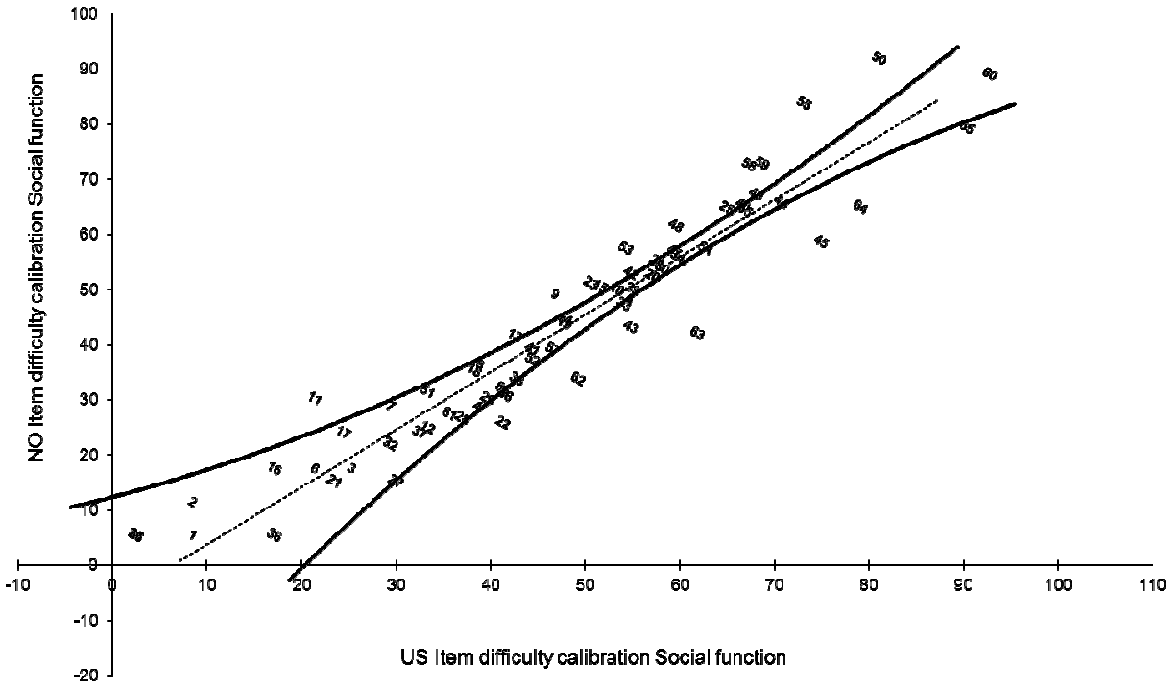

Figure 55: DIF analysis of Social Function items in the two normative samples

Also for this domain, table 52 confirms that the differences between the samples are substantial. Items with significant t-values ($t > 2.0$) are listed.

Again, differences are relatively numerous and go both ways. They also have a great deal in common with the differences mapped in table 51. While the American children perform better on a large number of tasks, Norwegian children apparently score higher on independent behavior outside their home.

Table 52: Social Function items with significant sample differences

Norwegian			*American*		
Item #	*Item name*	*DIF t-value >2*	*Item #*	*Item name*	*DIF t-value >2*
55	Consistently performs househ. task	6.82	45	Directs adult to give help return home	10.49
58	Crossing street w.o. prompting	5.90	63	Follows school/cty uidelines	9.75
11	Names things	5.44	64	Functions in c.ty without supervision	9.22
53	Initiates care for belongings	5.30	62	Plays outside home, per. monitoring	5.21
9	Understands 2-step commands	4.77	65	Makes store transaction w.o. assist.ce	4.46
59	Not accepting (things) from strangers	4.61	22	Tackles only immediate help	3.97
50	Regularly checks clock/time	4.13	46	General awareness of daily routines	2.14
48	Simple time concept	4.01			
23	Seeks help, tackles short delay	3.40			
13	Seeks information	2.84			
7	Understands 1-step commands	2.58			
15	Tells about feelings /thoughts	2.34			
17	Single word with meaning	2.22			
51	Helping to care for belongings	2.20			
25	Joins adult in solving problem	2.09			

4. Concluding comments

The purpose of our present analyses has simply been to show through examples that norms based on the American sample are not likely to be suitable for Norwegian children.

Detailed comparisons show a number of significant differences between the original normative sample of the PEDI and a comparable Norwegian sample. The finding perhaps most important to our concerns is that the two samples have different scores on Self-Care as well as the Mobility domain scores. Consequently, Norwegian standards are clearly needed for the Self-care and Mobility domains.

But all is not well with the Social Functions domain, either, even if the summed domain scores show no difference between the two samples. Within all three domains, *subdomain* scores show sample differences, and give a more detailed picture than the all-over domain scores. Frequently, the subdomain difference between the samples changes across age groups. Sample-by-age differences, therefore, may give interesting suggestions for diagnostic purposes.

In addition, interesting (and complex) sample differences are found with a large number of single items. These are partly independent of the domain and subdomain differences. A few examples are offered, to make it clear that also single-item sample differences should be expected vary across age groups. The selected examples may serve as reminders that sample and age differences with individual items may have implications for the diagnostic use of the PEDI. A full documentation of sample and age differences for all items is beyond the scope of the present paper, however.

The generally high "*R squared*" values obtained in our ANOVAs are also worth noting. They suggest that the multiple regression model implicit in each analysis of variance explains a large proportion of the variance; i.e. *age group* as well as *nation* are powerful predictors of many PEDI scores. And clearly, accounting for the *age* variance is needed to be able to properly assess the *sample* (national) differences.

The effect of the *age* variable is not necessarily consistent over its entire scale, however. A clear example of this is the observed "ceiling effects" (Graziano & Raulin, 1989; Shadish, Cook, & Campbell, 2002). In domains, subdomains as well as single items, differences between the older groups are often limited, indicating that most children master the challenges implied. In practical terms, it means that tests results from older children should be used cautiously. But also cases of "floor effects" are evident, with very few children succeeding. Here, differences between the younger children may go unnoticed.

For the clinical use of the PEDI, then, a fairly complex relationship between age and item difficulty should be noted. Some items – and also certain subdomains – are only applicable within a limited age span. Within a few age classes, the children proceed from mastering nothing to mastering it all. Outside this limited age period, the items are really not applicable. Unfortunately, this *"applicability window"* may appear at different ages in the American and the Norwegian normative samples.

Several analyses including *gender* have also been done, showing that the girls in both samples rather consistently do better than the boys. While this naturally is no surprise, it suggests that the PEDI may need gender-specific norms. The Norwegian sample is too small, however, to support two different sets of norms. The US sample probably also is. To keep things manageable, therefore, this potentially important topic has been left out of our discussion.

The present working paper does not fully describe or explain all differences between American and Norwegian children. Such an endeavor, of course, is far beyond the scope of our present efforts. The examples given should suffice, however, to support our claim that Norwegian norms should be preferred when testing Norwegian children.

5. References

Berg, M., Dolva, A.-S., Kleiven, J., & Krumlinde-Sundholm, L. Are national test norm values needed? Experiences from the development of Norwegian norms of the PEDI (in preparation).

Berg, M., Frey Frøislie, K., & Hussain, A. (2003). Applicability of Pediatric Evaluation of Disability Inventory (PEDI) in Norway. *Scandinavian Journal of Occupational Therapy, 10*, 118-126.

Berg, M., Aamodt, G., Stanghelle, J., Krumlinde-Sundholm, L., & Hussain, A. (2008). Cross-cultural variation of the Pediatric Evaluation of Disability Inventory (PEDI) norms in a randomized Norwegian population. *Scandinavian Journal of Occupational Therapy, 15*(3), 143-152.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. New York: Routledge.

Graziano, A. M., & Raulin, M. L. (1989). *Research Methods: A Process of Inquiry*. New York: Harper & Row.

Haley, S. M., Coster, W. J., Dumas, H. M., Fragala-Pinkham, M. A., & Moed, R. (2012). PEDI-CAT 1.3.6 Development, standardization and administration. Boston, Massachusetts, USA CREcare, LLC.

Haley, S. M., Coster, W. J., Ludlow, L. H., Haltiwanger, J. T., & Andrellos, P. J. (1992). *Pediatric Evaluation of Disability Inventory (PEDI). Development, Standardization and Administration Manual*. Boston: New England Medical Centre Hospitals

Linacre, J. M. (2010). *A User's Guide to Winsteps & Ministep: Rasch-Model Computer Programs*. Chicago: John M. Linacre.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causel Inference*. Boston: Houghton Mifflin Co.

Tennant, A., & Pallant, J. (2007). DIF matters: A practical approach to test if Differential Item Functioning makes a difference. *Rasch Measurement Transactions, 20*(4), 1082-1084.

Høgskolen i Lillehammer / Lillehammer University College
Postboks/P. O. Box 952, 2604 Lillehammer, Norway
Telefon/phone: (+47) 61 28 80 00

www.hil.no