

APPENDICES

Evaluating the impact of a sport-for-development intervention on the physical and mental health of young adolescents in Gulu, Uganda - a post-conflict setting within a low-income country

Justin Richards

University College, University of Oxford

APPENDIX A1: Methods for Systematic Literature Review

A1.1) Criteria for considering studies for this review

Articles considered for this review were screened according to study type, population, intervention and outcome measure.

A1.1a) Types of studies

INCLUDED:

All types of articles found in academic journals, grey literature and international organisation archives. Government documents as well as monitoring and evaluation (M&E) reports from non-governmental organisations (NGO) relevant to the participants, interventions and outcomes of interest were also included.

EXCLUDED:

The anticipated paucity of evidence relevant to the participants, interventions and outcomes of interest meant that articles were not excluded based on study type.

A1.1b) Types of populations

INCLUDED:

Adolescents aged 11-14 years living in a post-conflict region within a developing country.

EXCLUDED:

Adolescents who have relocated from conflict regions to developed countries as refugees or asylum seekers. These were excluded due to the established influence of environmental and social factors on the interventions and outcomes of interest.

Adolescents living in complex humanitarian emergency settings that are not post-conflict (e.g. natural disasters). These were excluded because post-conflict settings often have unique psychosocial and epidemiological characteristics associated with rapid change after a prolonged period of man-made relative impoverishment ^[1-4].

A1.1c) Types of interventions

INCLUDED:

Interventions that refer to sport, PA, PES or play. These fall under the broad description of PA and may be school-based or community-based interventions.

EXCLUDED:

Play that does not involve PA as part of the intervention.

A1.1d) Types of outcome measures

INCLUDED:

PF, MH and PA measured and reported in any form.

EXCLUDED:

The anticipated paucity of evidence relevant to the participants, interventions and outcomes of interest meant that outcome measures were not excluded based on how they were measured or reported.

A1.2) Search methods for identification of studies

All searches were completed from the beginning of existing records for the relevant databases and journals. The literature review was initially completed up until 01 March 2010 and was subsequently updated to 01 October 2011.

A1.2a) Search terms

A broad selection of search terms describing the population and intervention of interest were used (TABLE A1.1).

TABLE A1.1: Search terms used in the systematic review

Population		Intervention
Young adolescent	Post-conflict	Sport, PA, PES, play
- adolescen*	- armed*	- athlet*
- boy*	- army	- basketball*
- child*	- asylum	- bicycl*
- college*	- combat*	- bik*
- girl*	- conflict*	- cricket*
- juvenile*	- disaster*	- danc*
- kid	- displaced people	- exercise
- kids	- displaced person	- fitness
- pube*	- displaced population	- football*
- school*	- emergenc*	- jog*
- student*	- immigrant*	- netball*
- teen*	- immigrat*	- physical activity
- young female*	- migrant*	- physical ed*
- young male*	- militar*	- physical exer*
- young men	- peace	- play*
- young people	- post-conflict	- rugby
- young person	- refugee*	- run*
- youngster*	- soldier*	- soccer
- young women	- veteran*	- sport*
- youth*	- war	- sport program*
	- warfare	- sports program*
	- wars	- swim*
		- track and field
		- volleyball*

NOTE: * denotes the truncation of a search term

A1.2b) Electronic searches

Electronic searches of the following electronic databases were completed:

- Applied Social Sciences Index and Abstracts - ASSIA (CSA)
- CAB Abstracts (OVID)
- Cumulative Index to Nursing and Allied Health Literature – CINAHL (EBSCO)
- Cochrane Controlled Trials Register (Cochrane)
- Cochrane Database of Systematic Reviews (Cochrane)
- Current Controlled Trials (independent search engine)
- Dissertation Abstracts - UK and Ireland (independent search engine)
- Dissertation Abstracts – USA (independent search engine)
- Effective Public Health Practice Project – EPPHPP (independent search engine)
- Excerpta Medica Database – EMBASE (OVID)
- Evidence for Policy and Practice Information and Coordinating Centre - EPPI (independent search engine)
- Education Resources Information Centre – ERIC (CSA)
- Global Health (OVID)
- Google Scholar – first 300 x hits (Google)
- International Bibliography of the Social Sciences – IBSS (EBSCO)
- International Union for Health Promotion and Education – IUHPE (independent search engine)
- MEDLINE (OVID)
- Public Affairs Information Services International – PAIS (CSA)
- Physical Education Index (CSA)
- PsychINFO (OVID)
- System for Information on Grey Literature in Europe – SIGLE (independent search engine)

- Social Services Abstracts (CSA)
- Sociological Abstracts (CSA)
- SportDiscuss (EBSCO)
- Sport for Development Platform (independent search engine)
- Web of Knowledge (independent search engine)
- Worldwide Political Science Abstracts (CSA)

A1.2c) Manual searches

The following journals were individually searched for relevant articles:

- African Journal for Physical, Health Education, Recreation and Dance
- European Journal for Sport and Society
- International Journal of Sport and Health Science
- International Review for the Sociology of Sport
- Journal of Sport and Health Research
- Journal of Sport and Social Issues
- South African Journal for Research in Sport, Physical Education and Recreation
- Sociology of Sport Journal
- Sport, Education and Society
- Sport in Society

The reference lists of all included articles were screened to find additional relevant studies and papers.

A1.2d) Searching other resources

Prominent authors and experts in the field were contacted to identify any additional published or unpublished data and/or articles.

Several NGOs, international bodies and government agencies contributing to activities in this area were also contacted to identify relevant documents. All types of articles were collected from these sources including M&E data and reports that were often the final outputs from "action research".

A1.3) Data collection and analysis

EndNote X1 software was used to compile a central database of articles identified during the literature search. Duplicates were excluded.

A1.3a) Selection of studies

Articles were reviewed and included / excluded in three stages:

- 1) List of titles
- 2) List of titles and abstracts
- 3) Full text of articles

Relevant articles that satisfied the selection criteria were included for data extraction.

A1.3b) Data extraction

The origin, study type, population, intervention and outcomes of each article were collected using a dedicated data extraction form (FIGURE A1.1). The author utilised this information and other relevant details to formulate a brief summary of each article, which included a quality assessment of its contribution to the evidence base.

A1.3c) Data analysis and presentation

The results for each of the PH and MH outcomes were assimilated for analysis. The results were presented in narrative form.

FIGURE A1.1: Data extraction form for systematic review

SYSTEMATIC REVIEW DATA EXTRACTION FORM	
What type of literature?	
<input type="checkbox"/> Academic literature	<input type="checkbox"/> Grey literature

What type of study?	
<input type="checkbox"/> Intervention study	<input type="checkbox"/> Literature review
<input type="checkbox"/> Descriptive article	<input type="checkbox"/> Action research

Where was the article found?	
<input type="checkbox"/> Published in academic literature	<input type="checkbox"/> Sport organisation document
<input type="checkbox"/> Published in non-academic literature	<input type="checkbox"/> NGO document
<input type="checkbox"/> Book	<input type="checkbox"/> Public sector document
<input type="checkbox"/> Unpublished article/data from expert	<input type="checkbox"/> Private sector document
<input type="checkbox"/> International organisation document	<input type="checkbox"/> Other: _____

What population?	
Ages: _____	Location: _____

What intervention?	
<input type="checkbox"/> School-based	<input type="checkbox"/> Community-based
PA description: _____	
Dosage: _____	

What outcome measures?	
<input type="checkbox"/> Physical	<input type="checkbox"/> Mental Health
<input type="checkbox"/> Physical fitness	<input type="checkbox"/> Depression
<input type="checkbox"/> Physical activity level	<input type="checkbox"/> Anxiety
<input type="checkbox"/> Other: _____	<input type="checkbox"/> Self-concept
	<input type="checkbox"/> Other: _____
Measurement tools: _____	

Summary statement:	

APPENDIX A2: Location of Study

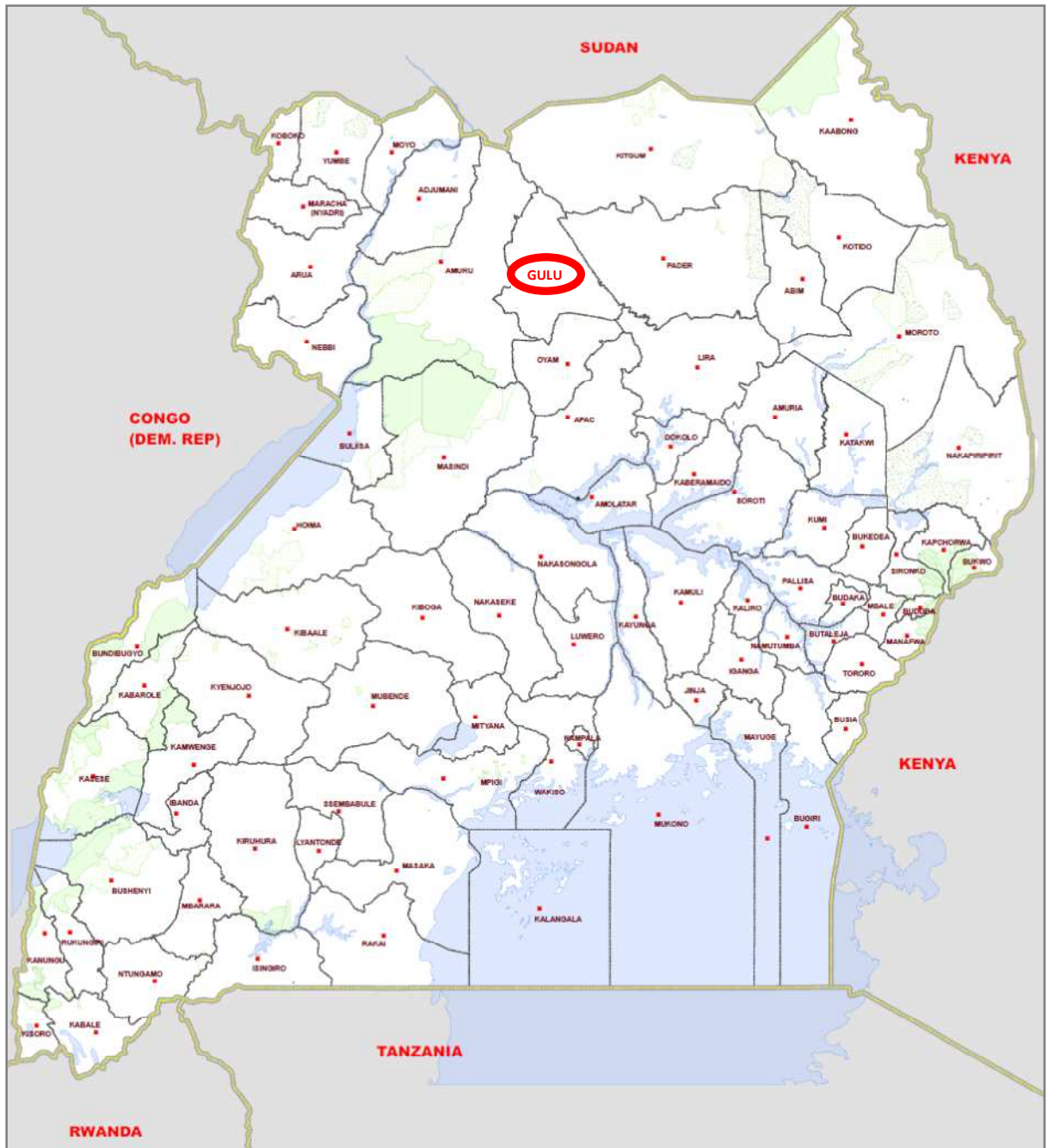
The location where the pilot testing and case-study for this thesis took place is identified in a series of maps (FIGURES A2.1 – A2.4).

FIGURE A2.1 Map of Africa showing the location of Uganda



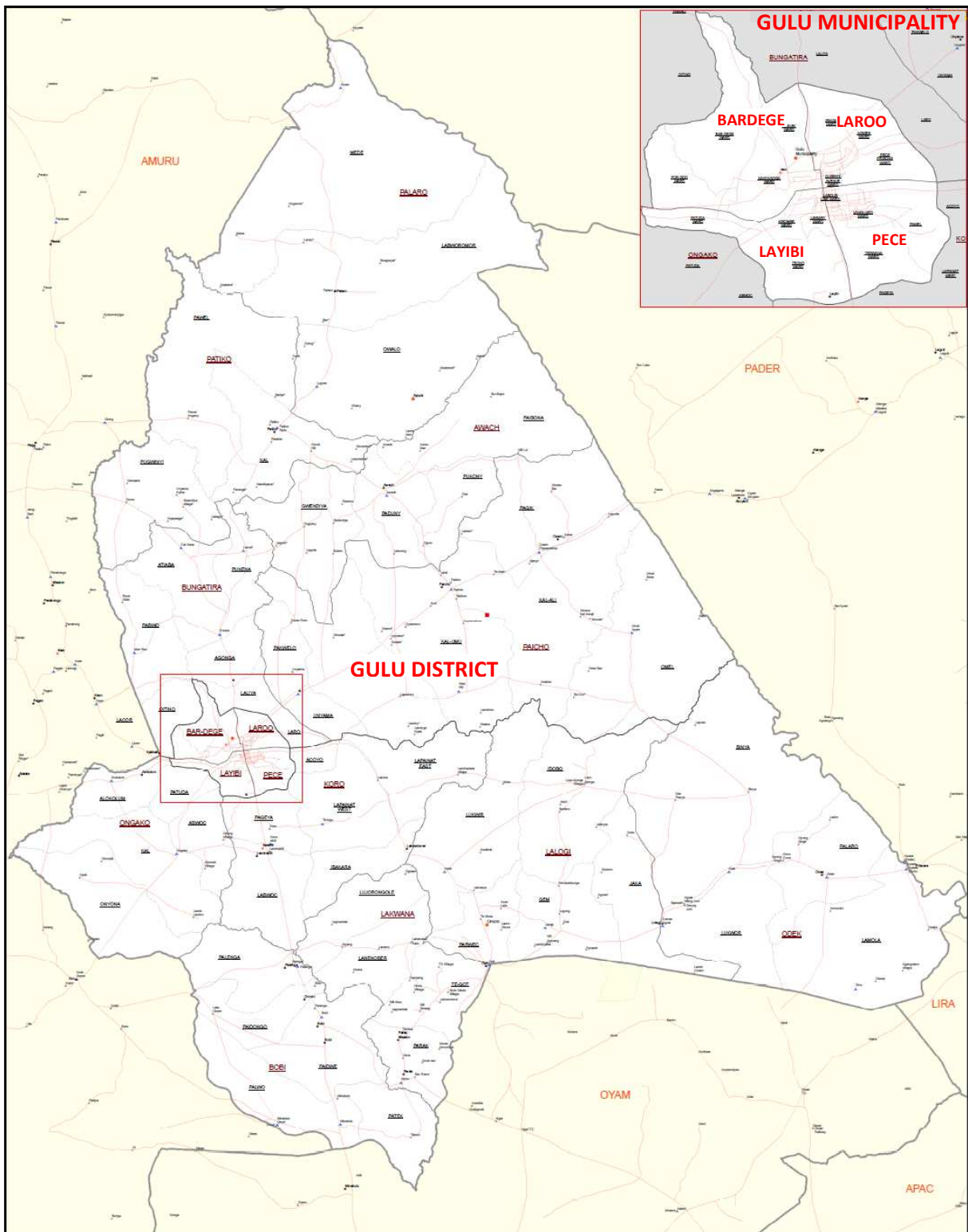
Source: Modified from Tele Atlas (2011) ^[5]

FIGURE A2.2: Map of Uganda showing the location of Gulu district



Source: Modified from UNOCHA (2009) [6]

FIGURE A2.3: Map of Gulu district showing the location of Gulu municipality and its divisions



Source: Modified from UNOCHA (2009) [7]

FIGURE A2.4: Map of the central Gulu municipality showing the division borders and the location of the intervention, pilot testing and 10 target schools



MAP LEGEND

- Gulu Town Centre
- Division Border
- S** Pece Stadium
- P** Pilot School – Laroo Boarding Primary School
- 1** Target School 1 - Christ Church Primary School
- 2** Target School 2 – Gulu Police Primary School
- 3** Target School 3 – Gulu Prison Primary School
- 4** Target School 4 – Gulu Public Primary School
- 5** Target School 5 – Gulu Town Primary School
- 6** Target School 6 – Holy Rosary Primary School
- 7** Target School 7 – Labour Line Primary School
- 8** Target School 8 – Layibi Central Primary School
- 9** Target School 9 – Pece P7 Primary School
- 10** Target School 10 – St Peter’s Primary School

NOTE: Locations shaded in blue were the sites where the registration and intervention activities were conducted.

Source: Modified from Tele Atlas (2011) ^[8]

APPENDIX A3: BMI-for-Age and Height-for-Age Measurement Protocol

Body height was measured according to the following protocol (FIGURE A3.1):

- A 2-metre tailor's measuring tape was attached to a vertical wall next to a flat hard surface. Electrical tape was used to secure the attachment so that the 0cm end of the measuring tape was flush with the junction between the wall and the floor.
- The subject removed their shoes and any hair ornaments. They stood tall, looking straight ahead with their feet together, heels on the ground and their back to the wall in line with the measuring tape.
- The assessor placed a flat 30cm ruler horizontally on top of the subject's head to observe the measurement on the tape measure to the nearest 1cm. The result was written directly onto the subject's recording sheet by the assessor.

FIGURE A3.1: Photo of the body height measurement in Gulu



*Photo taken by Justin Richards with consent of the subjects
– Gulu 2010*

Body weight was measured according to the following protocol (FIGURE A3.2):

- A set of manual bathroom weighing scales was placed on a flat hard surface and calibrated to 0kg.
- The subjects emptied their pockets, took off their shoes, removed any other heavy outer clothing (e.g. jacket) and then stood still with two feet facing forward in the middle of the weighing scales.
- The assessors stood directly in front of the scales and observed the measurement to the nearest 1kg. The result was written directly onto the subject's recording sheet by the assessor.

FIGURE A3.2: Photo of the body weight measurement in Gulu



*Photo taken by Justin Richards with consent of the subjects
– Gulu 2010*

Self-reported age was assessed according to the following protocol:

- Subjects were asked to write their DOB onto their recording sheet. Their age in whole months (M) at the time of measurement was calculated by the assessors.
- If subjects did not know their DOB, they were asked to write how old they were in years (Y). For the purposes of this study, the assessors calculated M as the number of months in Y and a half years (i.e. $M = [Y \times 12] + 6$).
- If subjects did not know how old they were in years, they were excluded from the analysis in this section of the study.

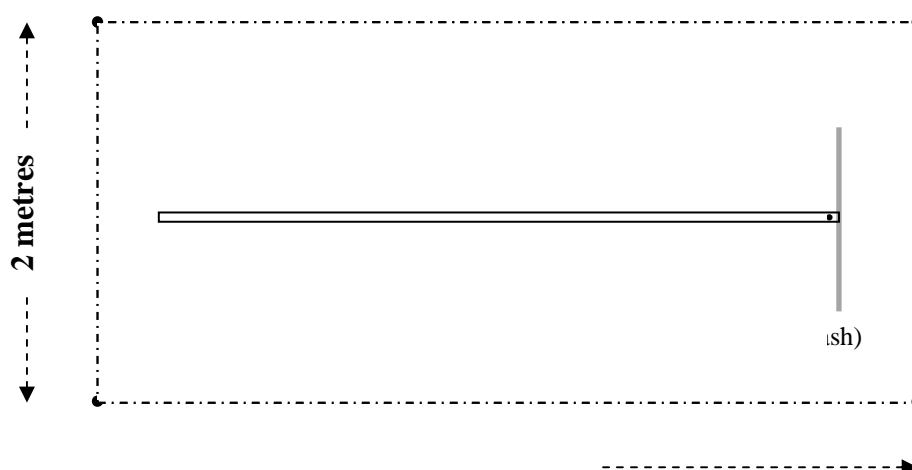
The assessors used the WHO AnthroPlus programme in SPSS to convert height (cm), weight (kg) and age (months) measurements into BFA and HFA z-scores based on 2007 normative values ^[9].

APPENDIX A4: Standing Broad Jump Measurement Protocol

The SBJ was measured according to the following protocol:

- The testing area was a rectangle (4 x 2 metres) on a flat area of dry and dusty ground. A tape measure at least 3 metres long was laid out straight and flat on the ground in the middle of the testing area. It was fixed at either end using a large flat top nail. A local broom was used to brush the loose dirt into a thin and smooth layer within the testing area. Ash was used to make a 1-metre long “starting” line that was perpendicular to the tape measure at the end that showed 0cm. The subjects lined up behind the “starting” line in a random order (FIGURE A4.1).

FIGURE A4.1: Diagram of testing area for the SBJ in Gulu



- The assessors demonstrated the testing procedure to the subjects while delivering the following verbal instructions in both English and Luo:

This test is called the standing broad jump. You are required to jump as far forwards as you can. You will be completing the test one at a time and cannot have a run-up before you jump. When it is your turn you will stand with your toes on the “starting” line and jump forward using both feet. You should land with one foot on either side of the tape measure and make sure you do not fall backwards. You will get three attempts to do the biggest jump you can.

- All subjects were given the opportunity to ask questions.
- The subject at the front of the line was invited to jump when they were ready. The assessor used a flat 30cm ruler perpendicular to the tape measure to identify the subject's foot marking in the loose dirt that was closest to the "starting" line. The assessor wrote the distance to the nearest 1cm directly onto the subject's recording sheet. The subject then went to the back of the line and waited for their next attempt (FIGURE A4.2).

FIGURE A4.2: Photo of the SBJ measurement in Gulu



Photo taken by Justin Richards with consent of the subjects – Gulu 2010

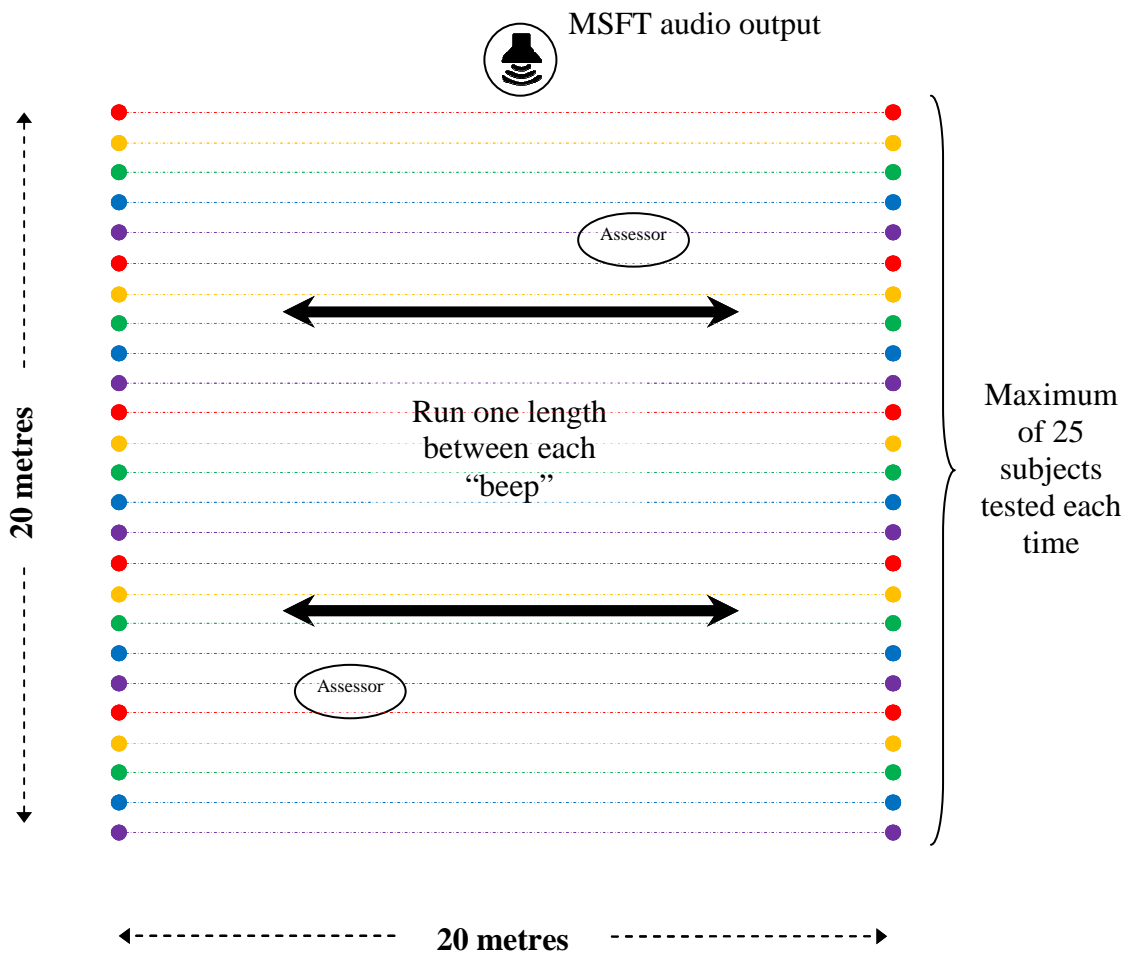
- If a subject fell backwards when landing from the test they were given the opportunity to try again.
- The testing area was re-swept by the assessor to create a fresh smooth and thin layer of loose dirt before each jump.
- The **longest** of the three jumps completed by each subject was used for data analysis.

APPENDIX A5: Multi-Stage Fitness Test Measurement Protocol

The MSFT was measured according to the following protocol:

- The testing area was a square (20 x 20 metres) on a flat area of ground that was measured using a 20-metre length of rope and demarcated with coloured markers. The audio device for the MSFT was set up at one end of the testing area. A randomly selected group of subjects (max = 25 at a time) lined up across one end of the testing area in line with a marker (FIGURE A5.1).

FIGURE A5.1: Diagram of testing area for the MSFT in Gulu



- The assessors demonstrated the testing procedure to these subjects while delivering the following verbal instructions in both English and Luo:

This is called the multi-stage fitness test. You are required to run between the markers in time with the “beeps” that you will hear. Each “beep” is your signal to start running towards the opposite marker. You must run one length between the markers before you hear the next “beep”. You must not start running before the “beep”. At the start of the test there is a long time between each “beep” so you will not have to run very fast. After about one minute you will hear the audio recording progress to the next level. When this happens the time between each “beep” gets shorter so you will have to run a little bit faster. Each minute you will progress to a new level. The aim of the test is to get to as high level as possible so keep trying as hard as you can to keep up with the “beeps”. If you do not make it to the opposite marker before the “beep”, try your best to catch up to the pace before the next “beep” sounds.

- All subjects were given the opportunity to ask questions.
- When all subjects were ready the MSFT audio recording was started.
- The subjects ran the 20 metres between the markers when signalled to do so in time with the “beeps”. The assessors circulated within the testing area to encourage the subjects and ensure they completed the test correctly (FIGURE A5.2).

FIGURE A5.2: Photo of the MSFT test measurement in Gulu



Photo taken by Justin Richards with consent of the subjects – Gulu 2010

- If a subject did not reach the marker in time for a “beep”, they were encouraged to continue and try to catch up to the pace.
- The subjects were stopped by the assessors if they failed to reach the marker (within 2 metres) for two consecutive ends.
- Each subject received a score according to the level and stage at which they were stopped (i.e. level 9, stage 5 was recorded as "9.5"). When a subject was stopped, the assessors wrote their score on a post-it note and gave it to the subject. The subject then exited the testing area and waited for all of the other participants in the group to finish.
- When all of the subjects in the group had completed the test, the scores were transcribed from the post-it notes onto each participant’s recording sheet by the assessors.
- The highest **complete** level achieved by each subject was used for data analysis.

APPENDIX A6: APAI Questionnaire Measurement Protocol

The APAI guided questionnaire was delivered according to the following protocol:

- All adolescents in one class were moved to be seated randomly within the room and assessed simultaneously (FIGURE A6.1).

FIGURE A6.1: Photo of the APAI measurement in Gulu



Photo taken by Justin Richards with consent of the subjects – Gulu 2010

- The assessors guided the class through the explanatory instructions and each item in the questionnaire in both English and Luo. The subjects wrote their responses directly on their recording sheet. Those who were experiencing difficulty were assisted by the class teachers and assessors (FIGURE A6.2).
- After the questionnaire was completed, the recording sheets were collected by the assessors and checked. Questions that were missed or had ambiguous answers were clarified directly with the subject by the assessors.
- The scores from each question were tallied according to the “local syndromes”. Final scores for TT, K, P, ML, DT and AT were used for data analysis.

FIGURE A6.2: The APAI questionnaire for assessing local MH syndromes in Gulu

	Gwet ma inongo ni in bene iwok ki iyee, ii cabit acel ma okato ni. Ka pe otime ii cabit ma okato ni gwet pe. <i>Please tick each item showing how frequently these comments are true for you over the past seven days. If they did not occur during that time tick the 'Never' box.</i>	Pe <i>Never</i>	Manok <i>A little</i>	Madwong <i>A lot</i>	Tutwal <i>Constantly</i>
1	Atamo madwong <i>I have a lot of thoughts</i>				
2	Atiko par <i>I have constant worries</i>				
3	Koma weng rem arema <i>I have pain all over my body</i>				
4	Wiya pe tiyo maber <i>My brain is not functioning well</i>				
5	Atamo ni konya pe <i>I think I am of no use</i>				
6	Atamo ni a dene <i>I think about suicide</i>				
7	Atito peko ma atye kwede <i>I talk constantly about my problems</i>				
8	Abedo kena-kena <i>I sit alone</i>				
9	Ape ki miti me kwan <i>I lose interest in school</i>				
10	Wiya bara <i>I get headaches</i>				
11	Doga pe mito cam <i>I lose my appetite</i>				
12	Awinyo ma lit I cwinya <i>I feel a lot of pain in my heart</i>				
13	Abedo ma nongo ateno tika <i>I sit with my cheek in my palm</i>				
14	Akok ka atye kena <i>I cry when I'm alone</i>				
15	Pe anino I dye wor <i>I do not sleep at night</i>				
16	Ading <i>I am disobedient</i>				
17	Awinyo koyo <i>I feel cold</i>				
18	Atiko butu (I dye ceng) <i>I lie down all the time during the day</i>				
19	Aparo par madwong <i>I have lots of worries</i>				
20	Amito bedo kena <i>I want to be alone</i>				
21	Akejo oyot <i>I am easily annoyed</i>				
22	Amako wiya <i>I hold my head</i>				
23	Tama bedo kama bor I ot kwan <i>I lose concentration in class</i>				
24	Amato kongo <i>I drink alcohol</i>				
25	Ayeto luwota <i>I insult friends</i>				
26	Pe amoto dano <i>I don't greet people</i>				
27	Pe atamo atir <i>I don't think straight</i>				
28	Ajwat kekena <i>I mutter to myself</i>				
29	Pe ageno dano <i>I don't trust</i>				
30	Awinyo ni ape kigin mo wek akonye kede <i>I feel I can do nothing to help myself</i>				
31	Amako ludito matek <i>I cling to elders</i>				
32	Atamo ni anyima peke <i>I think I do not have a future</i>				
33	Atiko ngwec angweca <i>I am constantly running around</i>				
34	Pe amito woo <i>I don't like noise</i>				
35	Atamo ni dano tye ka ryema <i>I think people are chasing me</i>				
36	Cwinya gone oyot <i>I have a fast heart rate</i>				
37	Alworo bedo kena <i>I fear being alone</i>				
38	Cwinya cwer <i>I feel sad</i>				
39	Atamo jami maraco <i>I think of bad things</i>				
40	Agoro <i>I am weak</i>				
41	Pe amito lok <i>I don't feel like talking</i>				
42	Wiya maro wil ngo ma atamo <i>I like to forget</i>				
43	Atiko kok akoka <i>I cry continuously</i>				
44	Pe aparo kedi a kwo nyo a too <i>I do not care whether I live or die</i>				

APPENDIX A7: Accelerometry Measurement Protocol

PA was measured with accelerometers using the following protocol:

- Each accelerometer was attached to an elastic belt in a position that would place it in front of the right hip when being worn. The elastic belts were of variable lengths and the accelerometers were covered in strapping tape to hide their appearance. They were numbered on the outside of this tape to assist the assessors with tracking and final recording (FIGURE A7.1).

FIGURE A7.1: Photo of boy wearing an accelerometer in Gulu



Photo taken by Justin Richards with consent of the subjects – Gulu 2010

- The selected subjects were given a demonstration and the following verbal instructions by the assessors in English and Luo:

You will be given a belt to wear for the next week. This is a special belt that you must wear at all times. Only you can wear the belt – no-one else should wear the belt.

There is a small box on the belt that is very important. This box is not dangerous or harmful and you should keep it safe. There is something inside the box that can tell me how active you are during the week. When I collect the box from you, I will also know how much you have remembered to wear the belt. There will be a prize for the boys who remember to wear the belt and box the most and do not damage them. If you damage, lose or give the belt and box to someone else to wear you will be disqualified.

You should:

- ***Wear the belt and box everyday – including weekends;***
- ***Wear the belt and box all the time – even when you are not in school uniform;***
- ***Wear the belt and box during any sport that you play;***
- ***Only take the belt and box off when you are washing yourself or sleeping.***

You should NOT:

- ***Take the belt and box off for anyone;***
- ***Give the belt and box to anyone else;***
- ***Try to open the box;***
- ***Use the belt and box for any other purpose.***

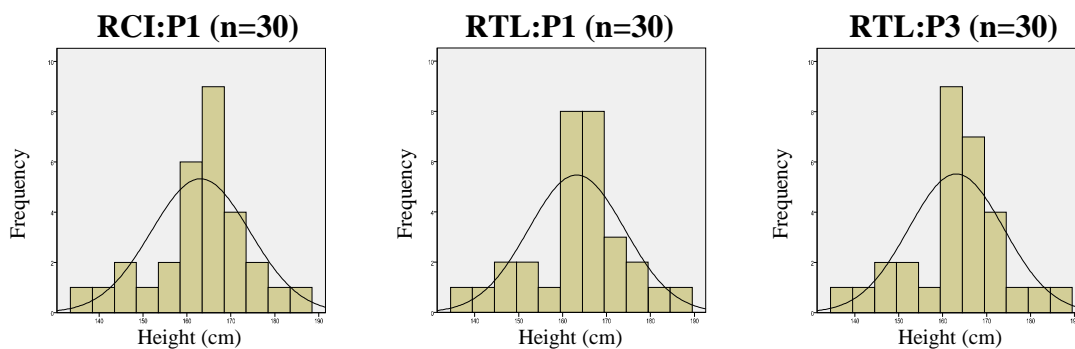
- All subjects were given the opportunity to ask questions and an information sheet outlining these details was also provided for them to take home.
- The subjects were assigned an accelerometer according to the belt size that would best fit them. They were advised on how to wear the accelerometer to optimise comfort and minimise the attention drawn to them. During the week the assessors visited the subjects within the school setting on two occasions prior to recollection. The purpose was to encourage compliance and troubleshoot as necessary.
- At the end of the data collection period, the members of the group who were most compliant with the measurement protocol were each awarded a University of Oxford bookmark.

APPENDIX A8: Reliability of Physical Fitness Outcomes (Assumptions)

The statistical methods used to assess intra-tester and inter-tester reliability of the PF outcome variables assumed that the samples in the pilot study were approximately normally distributed. This was initially assessed visually using a histogram for the crude results of each of the APAI sub-scales (FIGURES A8.1 – A8.4).

FIGURE A8.1: Histograms of the distribution of the crude height measurements in the pilot study in Gulu - stratified by gender, assessor and test time

Boys



Girls

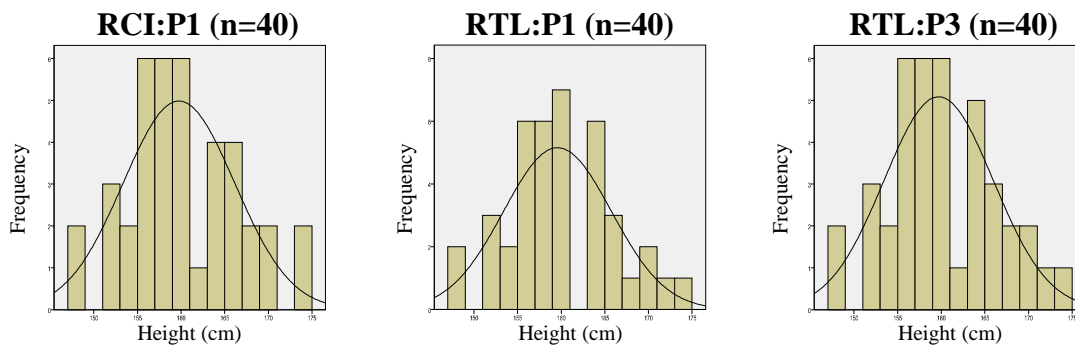
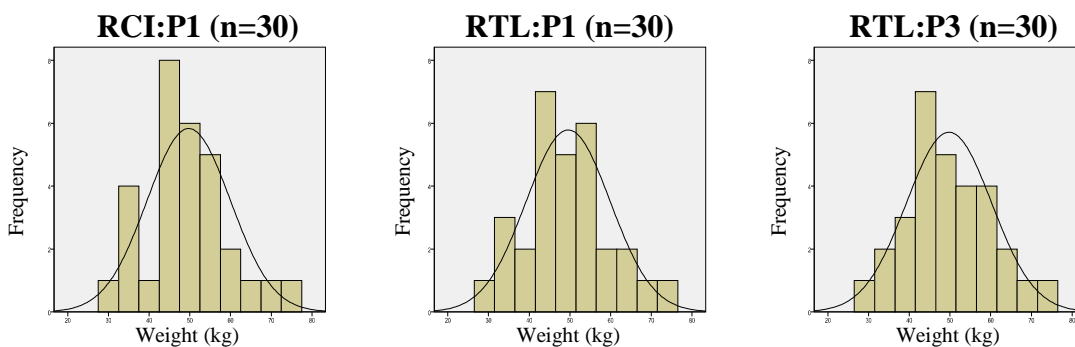


FIGURE A8.2: Histograms of the distribution of the crude weight measurements in the pilot study in Gulu - stratified by gender, assessor and test time

Boys



Girls

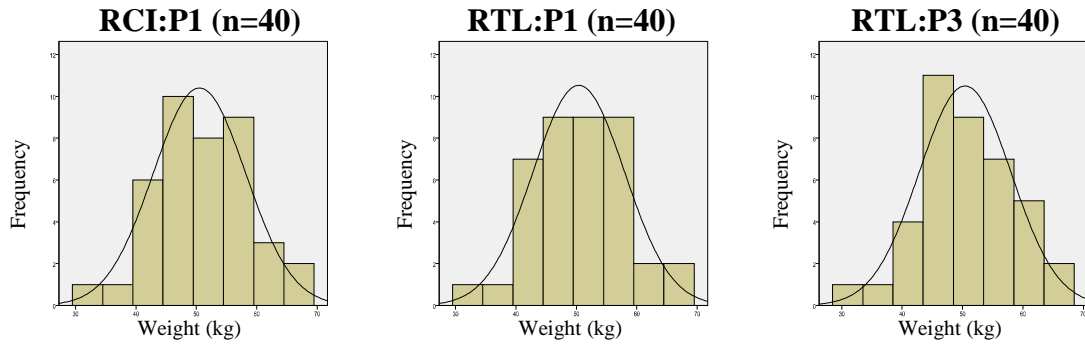
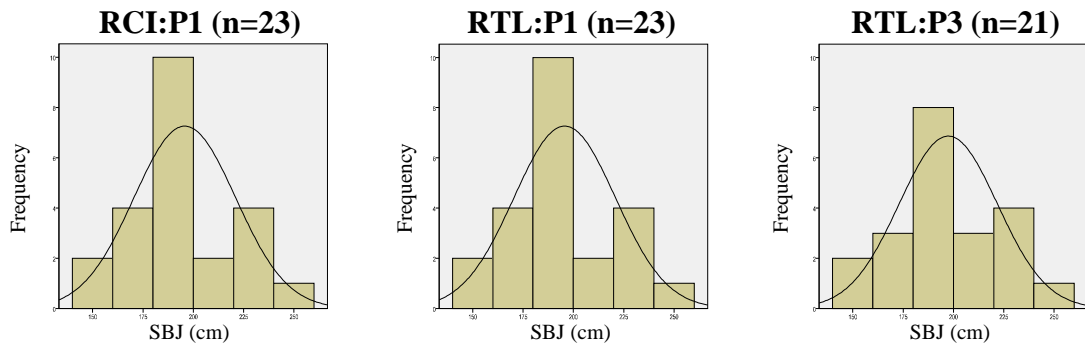


FIGURE A8.3: Histograms of the distribution of the crude SBJ measurements in the pilot study in Gulu - stratified by gender, assessor and test time

Boys



Girls

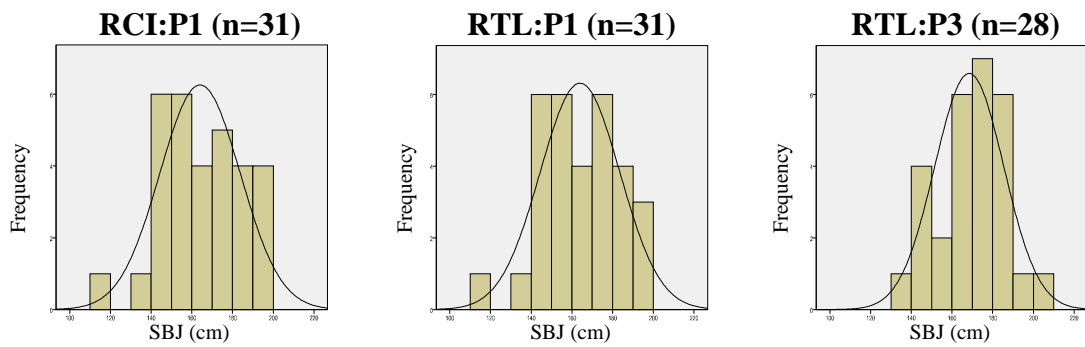
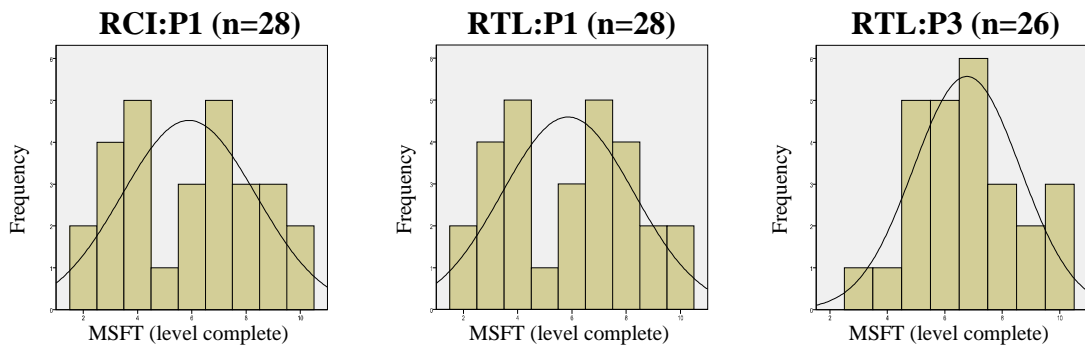
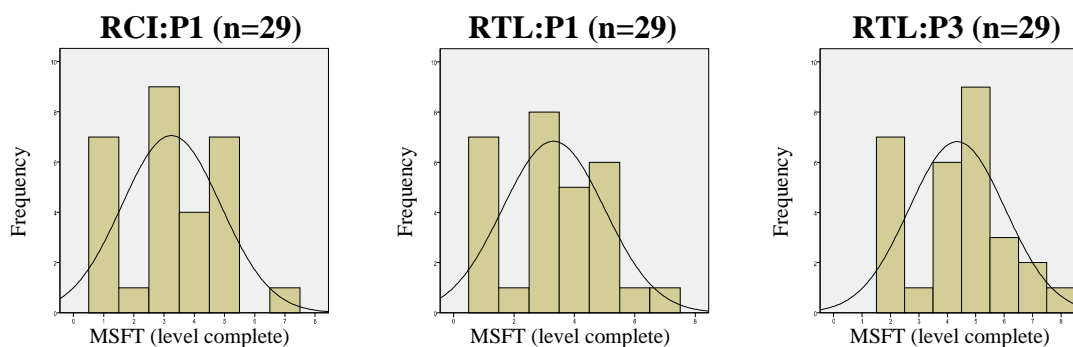


FIGURE A8.4: Histograms of the distribution of the crude MSFT measurements in the pilot study in Gulu - stratified by gender, assessor and test time

Boys



Girls



Inspection of the histograms warranted further assessment of normality and statistical tests of the crude results were also completed. Assessment of the skew and kurtosis of each of the PF outcome variables suggested a normal distribution of the data (skew within $\pm 2.0 \times \text{SES}$, kurtosis within $\pm 2.0 \times \text{SEK}$). These findings were supported by the results of the Shapiro-Wilk test for all of the variables except the girls MSFT for RCI:P1 [$p = 0.010$], RTL:P1 [$p = 0.023$] and RTL:P3 [$p = 0.020$] (TABLES A8.1 & A8.2).

TABLE A8.1: Statistical tests for normality of the crude PF outcomes in the boys pilot study in Gulu - stratified by assessor and test time

Outcome variable	n	Assessor and test time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
Height (cm)	30	RCI:P1	-0.468	0.427	0.842	0.833	0.960	0.304
	30	RTL:P1	-0.423	0.427	0.796	0.833	0.963	0.361
	30	RTL:P3	-0.408	0.427	0.885	0.833	0.962	0.343
Weight (kg)	30	RCI:P1	0.402	0.427	0.375	0.833	0.973	0.611
	30	RTL:P1	0.360	0.427	0.475	0.833	0.981	0.859
	30	RTL:P3	0.397	0.427	0.441	0.833	0.981	0.839
SBJ (cm)	23	RCI:P1	-0.042	0.481	-0.324	0.935	0.969	0.665
	23	RTL:P1	-0.042	0.481	-0.306	0.935	0.970	0.677
	21	RTL:P3	-0.292	0.501	0.125	0.972	0.977	0.871
MSFT (level complete)	28	RCI:P1	0.037	0.441	-1.226	0.858	0.937	0.091
	28	RTL:P1	0.029	0.441	-1.180	0.858	0.939	0.102
	26	RTL:P3	0.163	0.456	-0.453	0.887	0.954	0.293

TABLE A8.2: Statistical tests for normality of the crude PF outcomes in the girls pilot study in Gulu - stratified by assessor and test time

Outcome variable	n	Assessor and test time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
Height (cm)	40	RCI:P1	0.317	0.374	-0.325	0.733	0.977	0.585
	40	RTL:P1	0.306	0.374	-0.245	0.733	0.978	0.608
	40	RTL:P3	0.292	0.374	-0.324	0.733	0.981	0.738
Weight (kg)	40	RCI:P1	-0.013	0.374	-0.140	0.733	0.984	0.821
	40	RTL:P1	0.008	0.374	-0.145	0.733	0.987	0.912
	40	RTL:P3	-0.008	0.374	0.031	0.733	0.978	0.604
SBJ (cm)	31	RCI:P1	-0.289	0.421	-0.276	0.821	0.966	0.407
	31	RTL:P1	-0.321	0.421	-0.235	0.821	0.966	0.408
	28	RTL:P3	-0.255	0.441	-0.352	0.858	0.982	0.901
MSFT (level complete)	29	RCI:P1	0.105	0.434	-0.616	0.845	0.900	0.010
	29	RTL:P1	0.091	0.434	-0.706	0.845	0.915	0.023
	29	RTL:P3	0.075	0.434	-0.585	0.845	0.913	0.020

NOTE: RCI = research chief investigator, RTL = research team leader, P1 = pilot measurement 1, P3 = pilot measurement 3.

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times \text{SE}$ of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12,13]. The assumption of homogeneity of variance (HOV) between RCI:P1 vs. RTL:P1 and RTL:P1 vs. RTL:P3 was assessed for the crude results using the Levene and Brown-Forsythe tests (TABLES A8.3 & A8.4). These statistical tests for HOV were robust to violations of the normality assumption ^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$).

TABLE A8.3: Statistical tests of the HOV for the intra-tester reliability of the crude PF outcomes in the pilot study in Gulu (RTL:P1 vs.RTL:P3) - stratified by gender

Outcome variable	Gender	RTL:P1		RTL:P3		Levene test		Brown-Forsythe test		Summary
		n	Variance	n	Variance	F	p	F	p	
Height (cm)	Boys	30	119.476	30	117.472	0.006	0.938	0.003	0.959	√
	Girls	40	38.302	40	39.344	0.017	0.895	0.021	0.885	√
Weight (kg)	Boys	30	106.944	30	109.633	0.003	0.955	0.006	0.938	√
	Girls	40	57.477	40	57.789	0.002	0.962	0.001	0.981	√
SBJ (cm)	Boys	21	657.590	21	595.157	0.023	0.881	0.013	0.910	√
	Girls	28	291.704	28	287.517	0.336	0.564	0.319	0.574	√
MSFT (level complete)	Boys	26	5.386	26	3.465	2.350	0.132	2.291	0.136	√
	Girls	29	2.865	29	2.877	0.001	0.970	0.000	1.000	√

TABLE A8.4: Statistical tests of the HOV for the inter-tester reliability of the crude PF outcomes in the pilot study in Gulu (RTL:P1 vs. RCI:P1) -stratified by gender

Outcome Variable	Gender	RTL:P1		RCI:P1		Levene test		Brown-Forsythe test		Summary
		n	Variance	n	Variance	F	p	F	p	
Height (cm)	Boys	30	119.476	30	126.024	0.017	0.897	0.007	0.933	√
	Girls	40	38.302	40	40.962	0.058	0.810	0.040	0.842	√
Weight (kg)	Boys	30	106.944	30	105.306	0.007	0.936	0.003	0.954	√
	Girls	40	57.477	40	58.871	0.001	0.971	0.001	0.981	√
SBJ (cm)	Boys	23	637.403	23	639.905	0.000	0.998	0.000	0.985	√
	Girls	31	383.499	31	390.899	0.003	0.960	0.002	0.963	√
MSFT (level complete)	Boys	28	5.905	28	6.099	0.009	0.923	0.012	0.913	√
	Girls	29	2.865	29	2.690	0.057	0.812	0.064	0.801	√

NOTE: Results in ***BOLD ITALICS*** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

NOTE: RTL = research team leader, RCI = research chief investigator, P1 = pilot measurement 1, P3 = pilot measure 3

The assumption of HOV held for all of the dependent variables when comparing the crude results for RTL:P1 vs. RTL:P3 and RTL:P1 vs. RCI:P1. The statistical analysis could proceed with little concern when the HOV assumption was satisfied ^[12,14].

APPENDIX A9: Reliability of Mental Health Outcomes (Assumptions)

The statistical methods used to assess intra-tester and inter-method reliability of the MH outcome variables assumed that the samples in the pilot study were approximately normally distributed. This was initially assessed visually using a histogram for the crude results of each of the APAI sub-scales (FIGURES A9.1 – A9.6).

FIGURE A9.1: Histograms of the distribution of the crude TT scores in the pilot study in Gulu - stratified by assessor and method

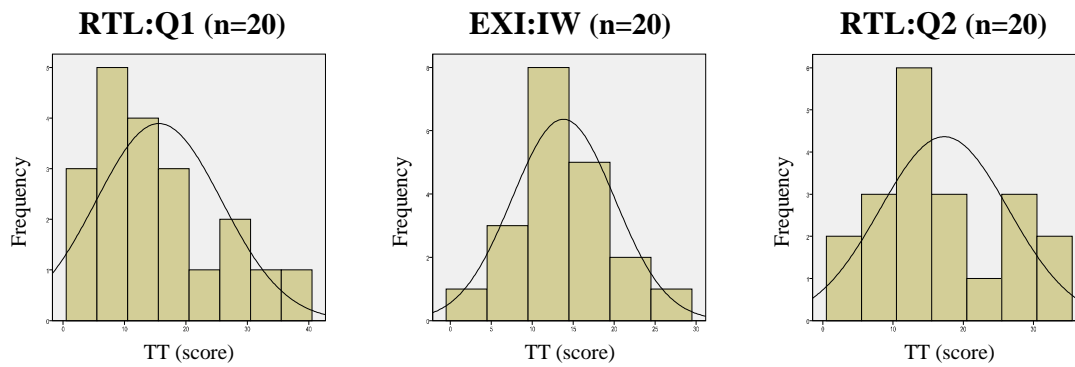


FIGURE A9.2: Histograms of the distribution of the crude K scores in the pilot study in Gulu - stratified by assessor and method

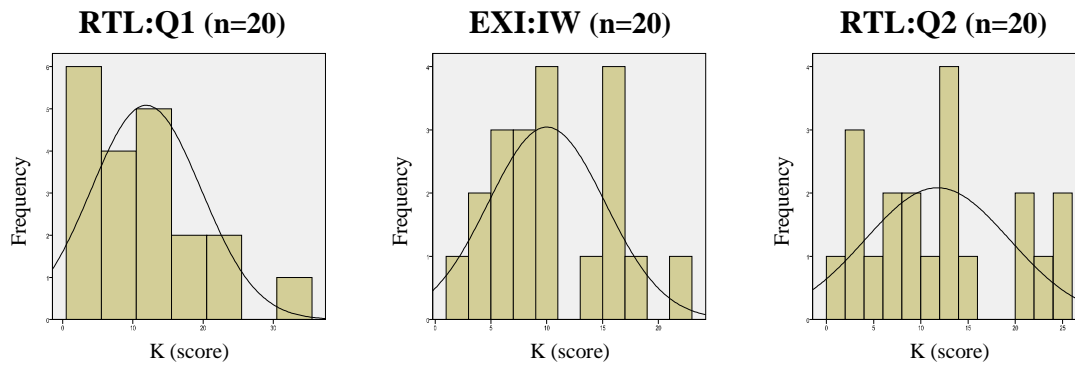


FIGURE A9.3: Histograms of the distribution of the crude P scores in the pilot study in Gulu - stratified by assessor and method

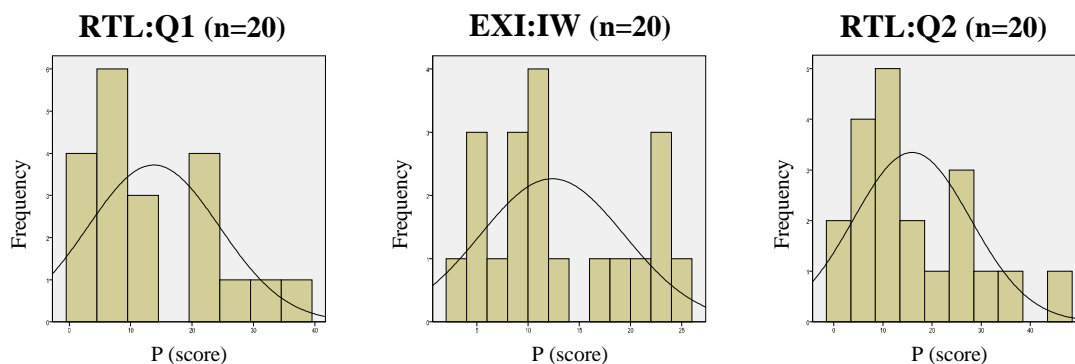


FIGURE A9.4: Histograms of the distribution of the crude ML scores in the pilot study in Gulu - stratified by assessor and method

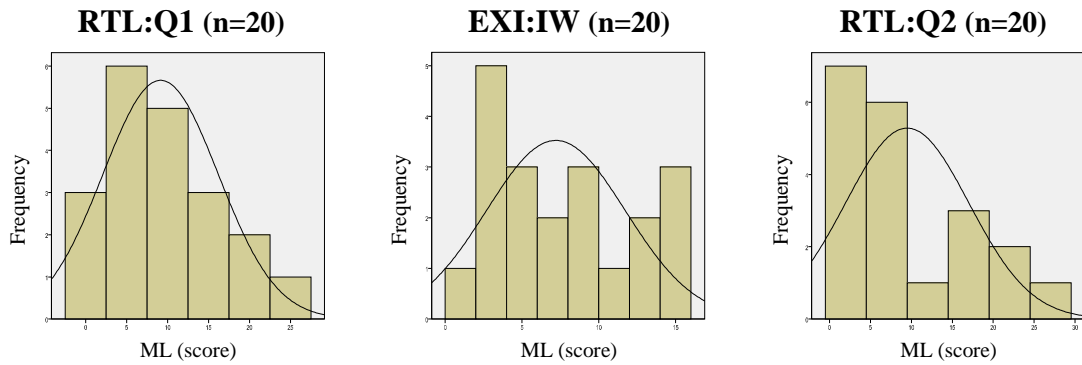


FIGURE A9.5: Histograms of the distribution of the crude DT scores in the pilot study in Gulu - stratified by assessor and method

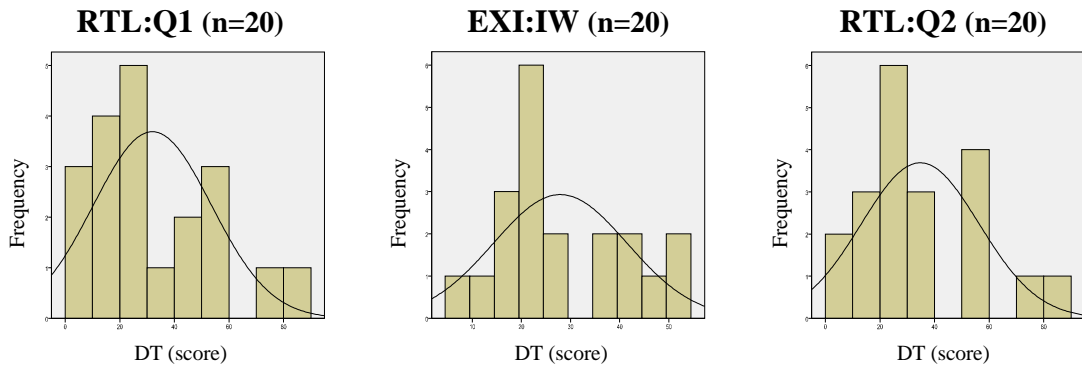
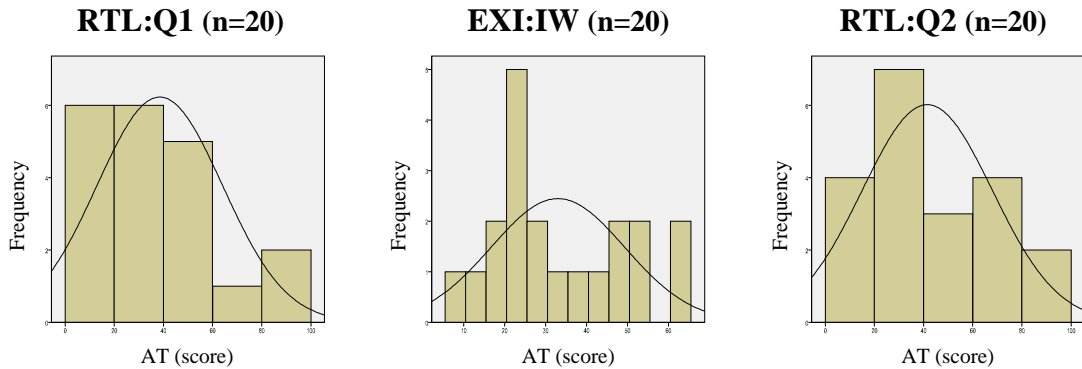


FIGURE A9.6: Histograms of the distribution of the crude AT scores in the pilot study in Gulu - stratified by assessor and method



Inspection of the histograms warranted further assessment of normality and statistical tests of the crude results were also completed. Assessment of the skew and kurtosis of each of the MH outcome variables suggested a normal distribution of the data (skew within $\pm 2.0 \times \text{SES}$, kurtosis within $\pm 2.0 \times \text{SEK}$). These findings were supported by the results of the Shapiro-Wilk test for all of the variables except P in Q1 [$p = 0.026$] and ML in Q2 [$p = 0.014$] (TABLE A9.1).

TABLE A9.1: Statistical tests for normality of the crude MH outcomes in the pilot study in Gulu - stratified by assessor and method

Outcome variable	n	Assessor and method	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
TT (score)	20	RTL:Q1	0.874	0.512	-0.163	0.992	0.910	0.064
	20	EXI:IW	0.263	0.512	-0.092	0.992	0.987	0.993
	20	RTL:Q2	0.389	0.512	-0.706	0.992	0.956	0.463
K (score)	20	RTL:Q1	1.024	0.512	0.849	0.992	0.912	0.068
	20	EXI:IW	0.390	0.512	-0.715	0.992	0.953	0.419
	20	RTL:Q2	0.461	0.512	-0.920	0.992	0.927	0.138
P (score)	20	RTL:Q1	0.800	0.512	-0.540	0.992	0.890	0.026
	20	EXI:IW	0.451	0.512	-1.193	0.992	0.905	0.050
	20	RTL:Q2	1.001	0.512	0.595	0.992	0.911	0.066
ML (score)	20	RTL:Q1	0.670	0.512	-0.302	0.992	0.939	0.231
	20	EXI:IW	0.449	0.512	-1.208	0.992	0.908	0.059
	20	RTL:Q2	0.906	0.512	-0.207	0.992	0.874	0.014
DT (score)	20	RTL:Q1	0.983	0.512	0.339	0.992	0.905	0.050
	20	EXI:IW	0.504	0.512	-0.863	0.992	0.923	0.111
	20	RTL:Q2	0.673	0.512	-0.433	0.992	0.934	0.186
AT (score)	20	RTL:Q1	0.933	0.512	0.239	0.992	0.917	0.085
	20	EXI:IW	0.416	0.512	-0.987	0.992	0.933	0.174
	20	RTL:Q2	0.695	0.512	-0.541	0.992	0.921	0.103

NOTE: RTL = research team leader, EXI = expert APAI interviewer, Q1 = APAI questionnaire 1, IW = APAI interview, Q2 = APAI questionnaire 2.

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times \text{SE}$ of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

Although several of the histograms indicated a non-Gaussian distribution of the sampled data, the assumption of normality was supported by most of the statistical tests that were applied. The assumption of a normal distribution was also supported by previous results obtained for the APAI by Betancourt et al (2009). The limited sample size in this pilot study may have contributed to the ambiguity of the normality tests. Consequently, for the intra-tester and inter-method reliability analyses of the APAI in questionnaire form, the data were treated as a sample derived from a normally distributed population.

The statistical analyses performed, also assumed that the groups being compared had approximately equal variance ^[12,13]. The assumption of HOV between the crude results for Q1 vs. Q2 and Q1 vs. IW was assessed using the Levene and Brown-Forsythe tests (TABLES A9.2 & A9.3). These statistical tests for HOV were robust to violations of the normality assumption ^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$).

TABLE A9.2: Statistical tests of the HOV for the intra-tester reliability of the crude MH outcomes in the pilot study in Gulu (RTL:Q1 vs. RTL:Q2)

Outcome variable	RTL:Q1		RTL:Q2		Levene test		Brown-Forsythe test		Summary
	n	Variance	n	Variance	F	p	F	p	
TT (score)	20	105.095	20	83.589	0.165	0.687	0.122	0.729	√
K (score)	20	61.568	20	58.618	0.011	0.919	0.010	0.919	√
P (score)	20	114.934	20	142.261	0.090	0.766	0.071	0.792	√
ML (score)	20	49.608	20	57.000	0.093	0.762	0.010	0.919	√
DT (score)	20	467.524	20	467.566	0.001	0.982	0.000	0.983	√
AT (score)	20	655.421	20	701.737	0.052	0.821	0.016	0.900	√

NOTE: Results in ***BOLD ITALICS*** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

NOTE: RTL = research team leader, Q1 = APAI questionnaire 1, Q2 = APAI questionnaire 2.

TABLE A9.3: Statistical tests of the HOV for the inter-method reliability of the crude MH outcomes in the pilot study in Gulu (RTL:Q1 vs. EXI:IW)

Outcome variable	RTL:Q1		EXI:IW		Levene test		Brown-Forsythe test		Summary
	n	Variance	n	Variance	F	p	F	p	
TT (score)	20	105.095	20	39.326	4.590	0.039	3.258	0.079	√
K (score)	20	61.568	20	27.474	1.824	0.185	1.879	0.178	√
P (score)	20	114.934	20	49.713	4.792	0.035	1.997	0.166	√
ML (score)	20	49.608	20	20.484	3.831	0.058	2.951	0.094	√
DT (score)	20	467.524	20	185.253	3.869	0.057	2.457	0.125	√
AT (score)	20	655.421	20	266.421	3.155	0.084	2.352	0.133	√

NOTE: Results in ***BOLD ITALICS*** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

NOTE: RTL = research team leader, EXI = expert APAI interviewer, Q1 = APAI questionnaire 1, Q2 = APAI questionnaire 2, IW = APAI interview.

The assumption of HOV held for all of the dependent variables when comparing the crude results for RTL:Q1 vs. RTL:Q2 and RTL:Q1 vs. EXI:IW. The statistical analysis that assumed approximately equal variances between these groups could proceed with little concern ^[12,14].

APPENDIX A10: Reliability of Mental Health Outcomes (Results)

The statistical methods used to assess inter-method reliability assumed that the mean difference between the methods was approximately normally distributed. Bland et al (1986) advocated checking if the normality was improved by applying a Ln transformation to the crude data^[15]. The distribution of the mean difference between the methods for both the crude and Ln transformed data was initially assessed visually using a histogram for each of the APAI sub-scales (FIGURES A10.1 – A10.6).

FIGURE A10.1: Histogram of the difference between methods for TT scores in the pilot study in Gulu using the crude and Ln transformed data (Q1 vs. IW)

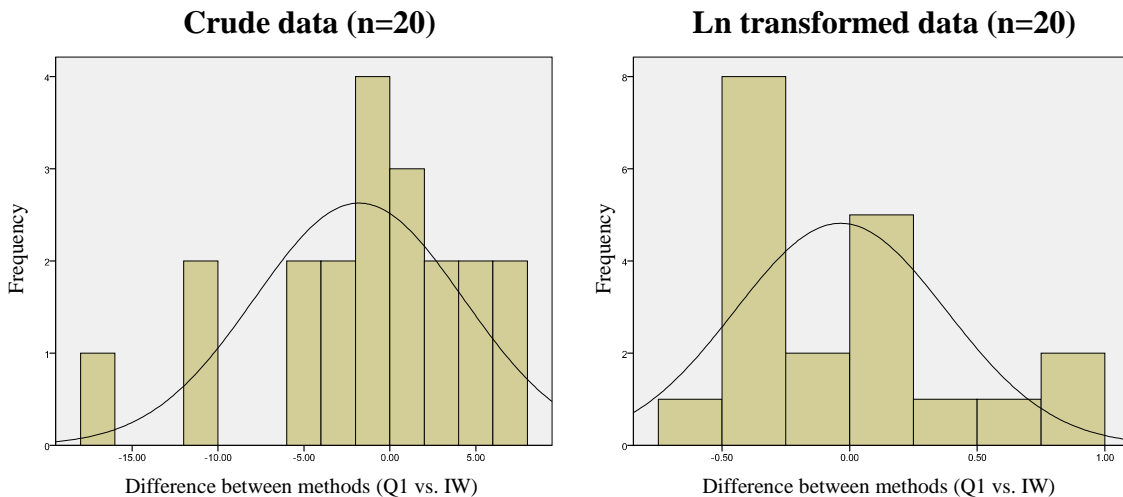


FIGURE A10.2: Histogram of the difference between methods for K scores in the pilot study in Gulu using the crude and Ln transformed data (Q1 vs. IW)

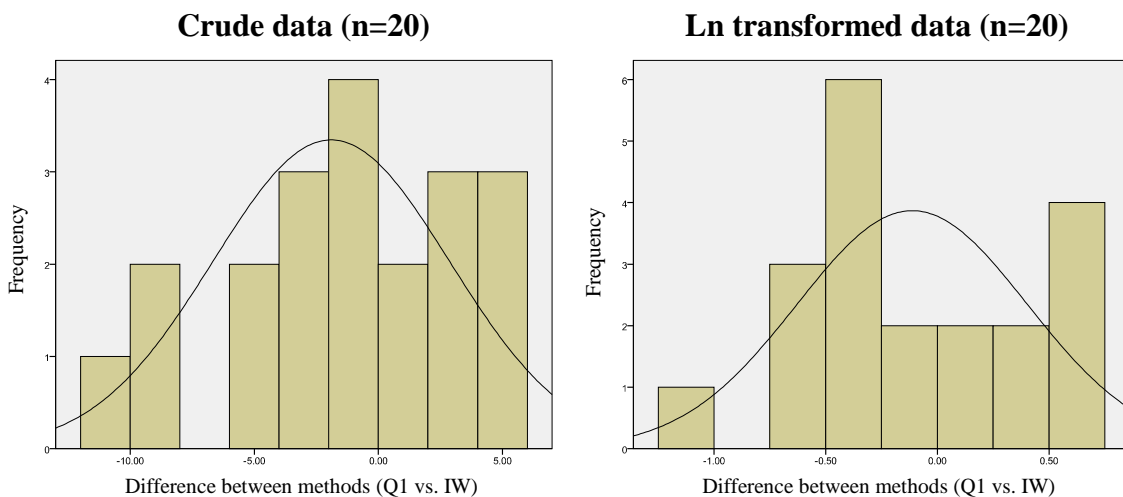


FIGURE A10.3: Histogram of the difference between methods for P scores in the pilot study in Gulu using the crude and Ln transformed data (Q1 vs. IW)

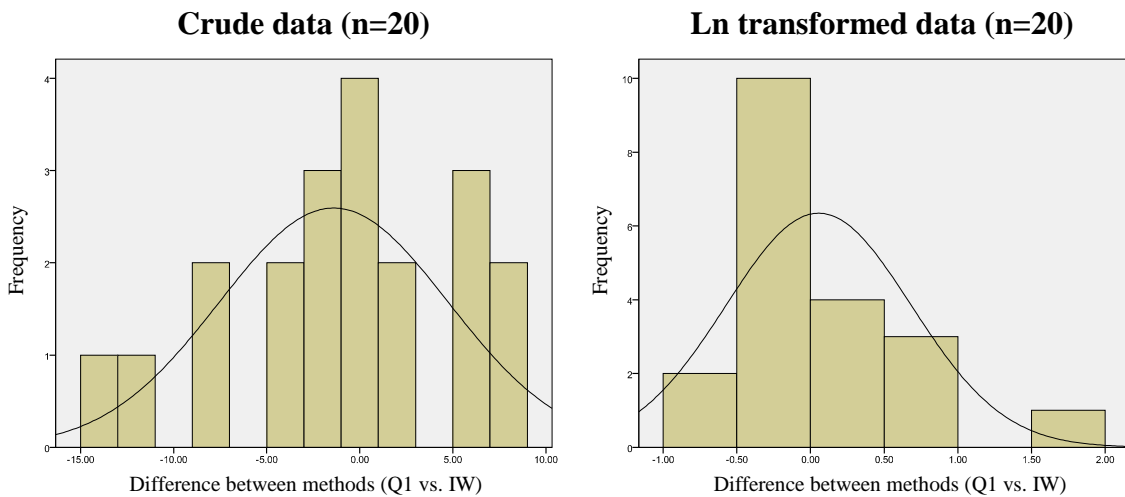


FIGURE A10.4: Histogram of the difference between methods for ML scores in the pilot study in Gulu using the crude and Ln transformed data (Q1 vs. IW)

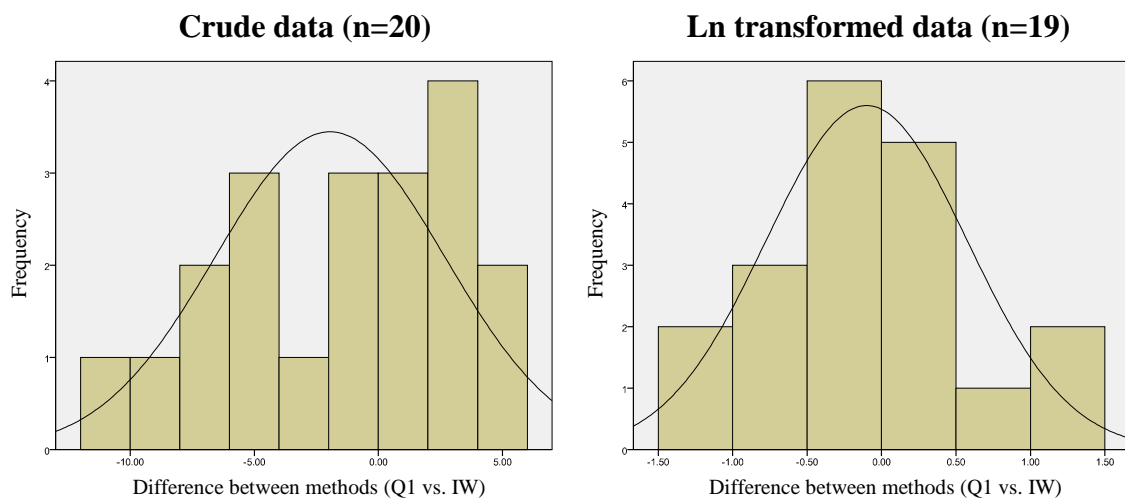


FIGURE A10.5: Histogram of the difference between methods for DT scores in the pilot study in Gulu using the crude and Ln transformed data (Q1 vs. IW)

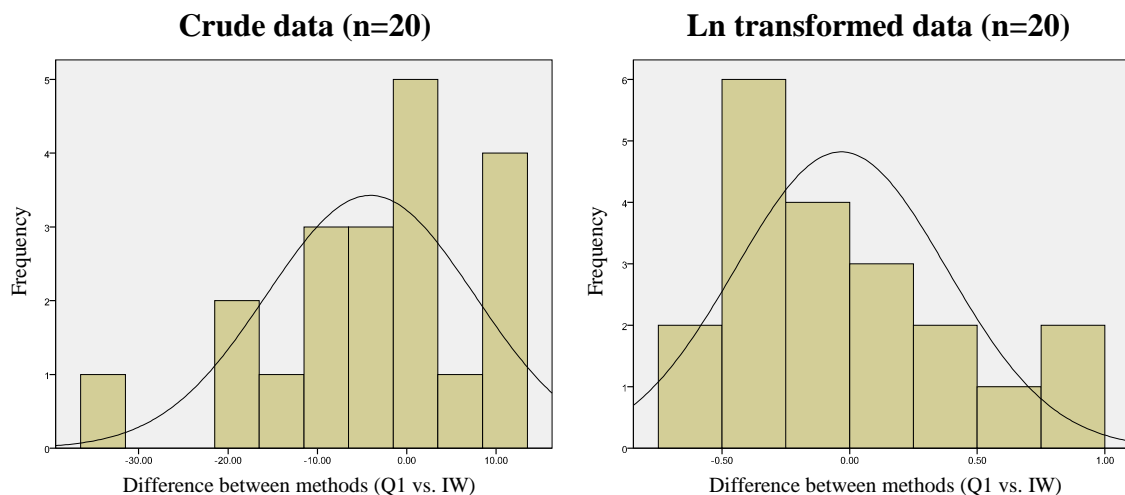
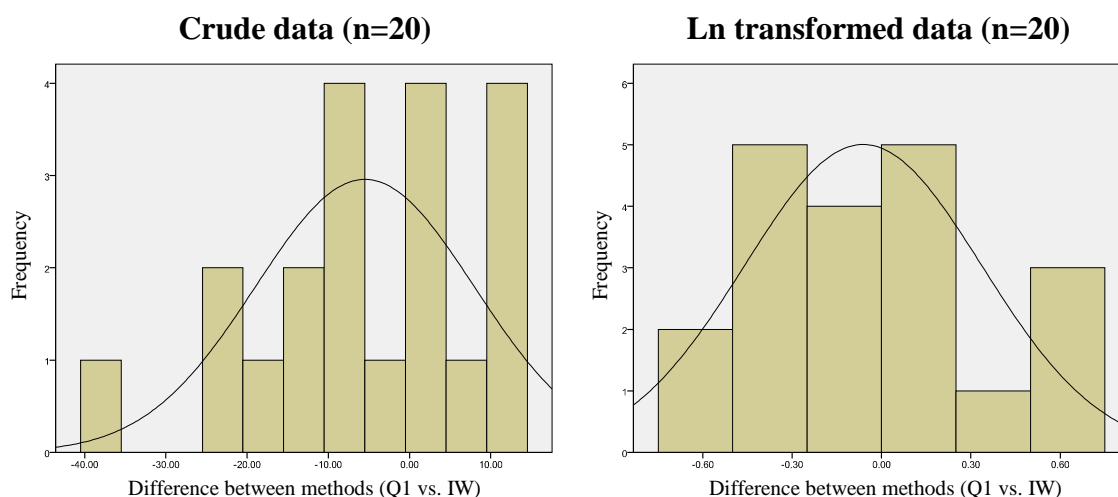


FIGURE A10.6: Histogram of the difference between methods for AT scores in the pilot study in Gulu using the crude and Ln transformed data (Q1 vs. IW)



Inspection of the histograms warranted further assessment of normality and statistical tests were also completed for the crude and Ln transformed data (TABLE A10.1).

TABLE A10.1: Statistical tests for normality of the difference between Q1 vs. IW for the crude and Ln transformed data of the MH outcomes in the pilot study in Gulu

Outcome variable	n	Difference between...	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
TT (score)	20	Q1 vs. IW	-0.890	0.512	0.800	0.992	0.939	0.225
	20	Q1(Ln) vs. IW(Ln)	0.927	0.512	0.082	0.992	0.911	0.066
K (score)	20	Q1 vs. IW	-0.380	0.512	-0.701	0.992	0.951	0.379
	20	Q1(Ln) vs. IW(Ln)	0.201	0.512	-0.709	0.992	0.944	0.280
P (score)	20	Q1 vs. IW	-0.451	0.512	-0.176	0.992	0.956	0.468
	20	Q1(Ln) vs. IW(Ln)	0.997	0.512	0.464	0.992	0.911	0.066
ML (score)	20	Q1 vs. IW	-0.499	0.512	-0.904	0.992	0.926	0.130
	19	Q1(Ln) vs. IW(Ln)	0.164	0.524	-0.495	1.014	0.964	0.653
DT (score)	20	Q1 vs. IW	-0.862	0.512	0.829	0.992	0.939	0.228
	20	Q1(Ln) vs. IW(Ln)	0.771	0.512	-0.211	0.992	0.921	0.106
AT (score)	20	Q1 vs. IW	-0.727	0.512	0.175	0.992	0.937	0.208
	20	Q1(Ln) vs. IW(Ln)	0.426	0.512	-0.835	0.992	0.943	0.274

NOTE: Q1 = APAI questionnaire 1, IW = APAI interview.

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

The statistical tests for both crude and Ln transformed data suggested a normal distribution for the difference between methods (Q1 vs. IW). However, visual examination of the histograms was less convincing. The effect of a Ln transformation on the skew, kurtosis and Shapiro-Wilk test for the difference between methods varied according to the outcome being considered. Similar variations were seen when observing the effect of the Ln transformation on the histograms. Therefore, the procedures recommended by Bland et al (1986) were completed for both the crude and Ln transformed data ^[15].

APPENDIX A11: Comparison between Stratified Levels for Age, Division and School at M1

The mean and 95% CI for all outcome variables were calculated for each of the stratified levels for age, division and school. The results were calculated separately for each gender (TABLES A11.1 & A11.2).

It is evident that there were different age-, division- and school-level effects on each of the outcome variables in this study (FIGURES A11.1 – A11.20). One-way ANOVA tests of the crude means indicated that there were significant differences between the stratified levels that varied according to the gender and outcome variable being assessed (TABLES A11.3 & A11.4).

TABLE A11.1: Crude mean and 95% CI of all outcome variables for the boys in the Gulu sample at M1- stratified by age, division and school

Independent variable		BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Age in years	11 (n=44)	-0.613 (-0.881↔-0.346) n=44	-0.123 (-0.470↔0.223) n=44	165.48 (160.84↔170.12) n=44	5.68 (5.10↔6.26) n=44	11.89 (10.03↔13.75) n=44	8.93 (7.47↔10.39) n=44	12.86 (11.10↔14.62) n=44	9.82 (8.51↔11.13) n=44	26.14 (22.37↔29.91) n=44	33.82 (29.38↔38.26) n=44
	12 (n=141)	-0.796 (-0.948↔-0.644) n=141	-0.588 (-0.738↔-0.437) n=141	171.59 (169.12↔174.06) n=141	5.40 (5.09↔5.71) n=141	11.09 (10.09↔12.09) n=141	8.16 (7.35↔8.97) n=141	11.26 (10.34↔12.18) n=141	8.93 (8.18↔9.68) n=141	23.45 (21.41↔25.49) n=141	30.16 (27.77↔32.55) n=141
	13 (n=234)	-0.677 (-0.788↔-0.567) n=234	-0.537 (-0.674↔-0.401) n=234	182.47 (179.97↔184.97) n=231	5.91 (5.67↔6.15) n=232	11.07 (10.25↔11.89) n=232	8.46 (7.77↔9.15) n=232	11.02 (10.25↔11.80) n=232	8.35 (7.76↔8.94) n=232	23.60 (21.96↔25.24) n=232	29.61 (27.70↔31.52) n=232
	14 (n=199)	-0.506 (-0.630↔-0.381) n=199	-0.337 (-0.471↔-0.203) n=199	195.11 (192.46↔197.76) n=195	6.65 (6.38↔6.92) n=198	11.94 (11.08↔12.80) n=196	9.16 (8.46↔9.86) n=196	12.21 (11.43↔12.99) n=196	8.76 (8.12↔9.40) n=196	25.81 (24.18↔27.44) n=196	32.12 (30.20↔34.04) n=196
Division	Bardege (n=48)	-0.668 (-0.902↔-0.433) n=48	-0.507 (-0.794↔-0.219) n=48	191.23 (184.78↔197.68) n=47	6.30 (5.82↔6.78) n=47	9.67 (8.27↔11.07) n=48	6.63 (5.48↔7.78) n=48	9.46 (8.08↔10.84) n=48	8.27 (7.05↔9.49) n=48	19.94 (17.31↔22.57) n=48	26.27 (23.08↔29.46) n=48
	Laroo (n=123)	-0.627 (-0.762↔-0.492) n=123	-0.349 (-0.504↔-0.194) n=123	184.07 (180.64↔187.50) n=123	6.30 (5.96↔6.64) n=123	12.76 (11.65↔13.87) n=123	9.68 (8.75↔10.61) n=123	12.04 (11.01↔13.07) n=123	9.01 (8.20↔9.82) n=123	26.44 (24.30↔28.58) n=123	32.82 (30.33↔35.31) n=123
	Layibi (n=164)	-0.678 (-0.810↔-0.546) n=164	-0.470 (-0.633↔-0.307) n=164	186.29 (183.21↔189.37) n=163	6.07 (5.75↔6.39) n=164	11.42 (10.53↔12.31) n=163	8.61 (7.83↔9.39) n=163	11.64 (10.74↔12.54) n=163	8.87 (8.15↔9.59) n=163	24.47 (22.65↔26.29) n=163	31.04 (28.86↔33.22) n=163
	Pece (n=283)	-0.629 (-0.743↔-0.515) n=283	-0.483 (-0.605↔-0.362) n=283	178.70 (176.32↔181.08) n=278	5.81 (5.58↔6.04) n=281	11.11 (10.34↔11.88) n=279	8.56 (7.96↔9.16) n=279	11.72 (10.97↔12.47) n=279	8.58 (8.05↔9.11) n=279	24.34 (22.82↔25.86) n=279	30.63 (28.86↔32.40) n=279

Independent variable		BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
School	CC (n=49)	-0.724 (-0.951↔-0.497) n=49	-0.651 (-0.893↔-0.409) n=49	187.88 (183.63↔192.13) n=48	6.17 (5.73↔6.61) n=48	9.47 (8.54↔10.40) n=49	5.98 (5.13↔6.83) n=49	10.37 (9.27↔11.47) n=49	8.78 (7.79↔9.77) n=49	20.10 (18.42↔21.78) n=49	26.71 (24.78↔28.64) n=49
	GPol (n=110)	-0.955 (-1.150↔-0.761) n=110	-0.382 (-0.545↔-0.219) n=110	171.68 (168.49↔174.87) n=110	5.78 (5.40↔6.16) n=110	12.44 (11.14↔13.74) n=110	9.89 (8.89↔10.89) n=110	13.28 (12.05↔14.51) n=110	10.20 (9.30↔11.10) n=110	27.44 (24.93↔29.95) n=110	34.72 (31.76↔37.68) n=110
	GPrs (n=71)	-0.911 (-1.127↔-0.694) n=71	-0.594 (-0.789↔-0.399) n=71	184.65 (179.78↔189.52) n=71	6.30 (5.86↔6.74) n=71	10.35 (9.17↔11.53) n=71	8.65 (7.55↔9.75) n=71	11.35 (10.14↔12.56) n=71	9.07 (8.09↔10.05) n=71	23.45 (21.00↔25.90) n=71	30.37 (27.48↔33.26) n=71
	GPub (n=72)	-0.475 (-0.645↔-0.305) n=72	-0.198 (-0.441↔0.046) n=72	189.03 (184.17↔193.89) n=72	6.71 (6.31↔7.11) n=72	8.35 (7.15↔9.56) n=72	5.85 (4.89↔6.81) n=72	7.93 (6.77↔9.09) n=72	5.75 (4.77↔6.73) n=72	16.81 (14.48↔19.14) n=72	21.14 (18.27↔24.01) n=72
	GTwn (n=35)	-0.545 (-0.809↔-0.280) n=35	-0.254 (-0.564↔0.057) n=35	186.66 (180.71↔192.61) n=35	7.31 (6.80↔7.82) n=35	13.20 (10.79↔15.61) n=35	10.66 (8.83↔12.49) n=35	12.51 (10.46↔14.56) n=35	9.43 (7.73↔11.13) n=35	28.31 (23.86↔32.76) n=35	34.89 (29.55↔40.23) n=35
	HR (n=79)	-0.403 (-0.586↔-0.219) n=79	-0.454 (-0.696↔-0.212) n=79	181.60 (177.10↔186.10) n=78	4.81 (4.39↔5.23) n=78	12.75 (11.36↔14.14) n=76	9.29 (8.09↔10.49) n=76	12.12 (10.67↔13.57) n=76	9.03 (8.01↔10.05) n=76	26.33 (23.47↔29.19) n=76	32.78 (29.57↔35.99) n=76
	LL (n=33)	-0.623 (-0.959↔-0.288) n=33	-0.629 (-1.004↔-0.254) n=33	180.69 (173.04↔188.34) n=32	6.00 (5.37↔6.63) n=33	8.75 (7.29↔10.21) n=32	7.25 (5.93↔8.57) n=32	10.34 (8.72↔11.96) n=32	7.56 (6.66↔8.46) n=32	20.44 (17.39↔23.49) n=32	26.84 (23.36↔30.32) n=32
	LC (n=66)	-0.376 (-0.577↔-0.176) n=66	-0.372 (-0.695↔-0.049) n=66	193.18 (188.54↔197.82) n=66	6.24 (5.68↔6.80) n=66	13.00 (11.68↔14.32) n=66	10.18 (9.05↔11.31) n=66	13.88 (12.59↔15.17) n=66	9.79 (8.76↔10.82) n=66	29.03 (26.43↔31.63) n=66	36.11 (33.01↔39.21) n=66
	PP (n=64)	-0.554 (-0.775↔-0.333) n=64	-0.656 (-0.889↔-0.423) n=64	176.80 (171.88↔181.72) n=60	5.65 (5.21↔6.09) n=63	10.76 (8.96↔12.56) n=63	8.06 (6.55↔9.57) n=63	10.02 (8.31↔11.73) n=63	7.41 (6.23↔8.59) n=63	22.52 (18.86↔26.18) n=63	28.21 (23.97↔32.45) n=63
	SP (n=39)	-0.697 (-0.888↔-0.506) n=39	-0.478 (-0.797↔-0.160) n=39	184.90 (177.49↔192.31) n=39	6.18 (5.71↔6.65) n=39	14.92 (12.94↔16.90) n=39	10.10 (8.63↔11.57) n=39	13.33 (11.61↔15.05) n=39	9.33 (7.97↔10.69) n=39	28.97 (25.35↔32.59) n=39	35.05 (30.96↔39.14) n=39

TABLE A11.2: Crude mean and 95% CI of all outcome variables for the girls in the Gulu sample at M1- stratified by age, division and school

Independent variable		BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Age in years	11 (n=92)	-0.524 (-0.691↔-0.356) n=91	-0.219 (-0.393↔-0.045) n=91	158.13 (155.04↔161.22) n=91	3.38 (3.05↔3.71) n=90	12.88 (11.49↔14.27) n=91	10.84 (9.69↔12.00) n=91	13.81 (12.52↔15.10) n=91	10.63 (9.71↔11.55) n=91	28.23 (25.60↔30.86) n=91	35.89 (32.76↔39.02) n=91
	12 (n=177)	-0.500 (-0.620↔-0.380) n=177	-0.034 (-0.177↔0.109) n=177	160.05 (157.73↔162.37) n=175	3.46 (3.22↔3.70) n=174	13.78 (12.76↔14.80) n=177	10.80 (9.91↔11.69) n=177	13.40 (12.34↔14.46) n=177	9.59 (8.89↔10.29) n=177	29.04 (26.92↔31.16) n=177	36.33 (33.85↔38.81) n=177
	13 (n=293)	-0.162 (-0.263↔-0.060) n=293	-0.067 (-0.162↔0.028) n=293	165.02 (162.93↔167.11) n=292	3.82 (3.62↔4.02) n=291	15.28 (14.49↔16.07) n=293	11.61 (10.98↔12.24) n=293	14.82 (14.07↔15.57) n=293	10.11 (9.58↔10.64) n=293	31.86 (30.32↔33.40) n=293	39.29 (37.51↔41.07) n=293
	14 (n=284)	0.011 (-0.096↔0.118) n=284	-0.021 (-0.130↔0.088) n=284	166.50 (164.38↔168.62) n=278	3.46 (3.25↔3.67) n=276	16.80 (15.95↔17.65) n=280	12.74 (12.03↔13.45) n=280	15.66 (14.86↔16.46) n=280	10.38 (9.83↔10.93) n=280	34.66 (33.00↔36.32) n=280	42.26 (40.34↔44.18) n=280
Division	Bardege (n=57)	-0.164 (-0.406↔0.078) n=57	-0.071 (-0.321↔0.179) n=57	167.21 (162.56↔171.86) n=57	3.68 (3.26↔4.10) n=57	14.16 (12.72↔15.60) n=57	10.44 (9.00↔11.88) n=57	13.33 (11.93↔14.73) n=57	9.84 (8.65↔11.03) n=57	29.05 (26.28↔31.82) n=57	36.12 (32.86↔39.38) n=57
	Laroo (n=144)	-0.305 (-0.460↔-0.149) n=143	0.092 (-0.046↔0.231) n=143	163.90 (160.87↔166.93) n=142	3.64 (3.32↔3.96) n=137	16.76 (15.50↔18.02) n=143	12.88 (11.83↔13.93) n=143	16.37 (15.13↔17.61) n=143	11.04 (10.23↔11.85) n=143	34.97 (32.50↔37.44) n=143	43.01 (40.20↔45.82) n=143
	Layibi (n=209)	-0.067 (-0.183↔0.050) n=209	-0.157 (-0.276↔-0.038) n=209	163.37 (161.05↔165.69) n=207	3.71 (3.48↔3.94) n=206	15.76 (14.80↔16.72) n=208	12.12 (11.31↔12.93) n=208	15.15 (14.23↔16.07) n=208	10.47 (9.82↔11.12) n=208	33.13 (31.21↔35.05) n=208	40.70 (38.49↔42.91) n=208
	Pece (n=436)	-0.260 (-0.344↔-0.177) n=436	-0.064 (-0.150↔0.022) n=436	163.37 (161.72↔165.02) n=430	3.48 (3.32↔3.64) n=431	14.57 (13.90↔15.24) n=433	11.34 (10.81↔11.87) n=433	14.09 (13.47↔14.71) n=433	9.74 (9.32↔10.16) n=433	30.49 (29.19↔31.79) n=433	37.80 (36.29↔39.31) n=433

Independent variable		BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
School	CC (n=60)	-0.355 (-0.578↔-0.131) n=60	-0.057 (-0.310↔0.197) n=60	167.78 (162.94↔172.62) n=59	5.51 (5.07↔5.95) n=59	11.72 (10.41↔13.03) n=60	8.38 (7.51↔9.25) n=60	13.42 (12.19↔14.65) n=60	9.47 (8.55↔10.39) n=60	26.37 (23.99↔28.75) n=60	33.95 (31.23↔36.67) n=60
	GPol (n=154)	-0.375 (-0.503↔-0.246) n=154	-0.077 (0.208↔0.055) n=154	260.76 (157.95↔163.57) n=154	3.13 (2.87↔3.39) n=152	15.56 (14.49↔16.63) n=154	12.42 (11.56↔13.28) n=154	15.55 (14.51↔16.59) n=154	11.31 (10.65↔11.97) n=154	33.06 (30.95↔35.17) n=154	41.09 (38.67↔43.51) n=154
	GPrs (n=98)	0.014 (-0.152↔0.180) n=98	-0.150 (-0.305↔0.004) n=98	162.69 (159.04↔166.34) n=97	3.52 (3.19↔3.85) n=97	16.67 (15.28↔18.06) n=98	12.82 (11.59↔14.05) n=98	15.35 (13.94↔16.76) n=98	10.76 (9.65↔11.87) n=98	34.51 (31.66↔37.36) n=98	41.85 (38.51↔45.19) n=98
	GPub (n=104)	-0.041 (-0.187↔0.104) n=104	-0.041 (-0.218↔0.136) n=104	162.94 (159.65↔166.23) n=104	3.52 (3.21↔3.83) n=102	13.47 (12.22↔14.72) n=104	9.62 (8.64↔10.60) n=104	12.61 (11.42↔13.80) n=104	8.62 (7.81↔9.43) n=104	27.10 (24.71↔29.49) n=104	33.50 (30.73↔36.27) n=104
	GTwn (n=44)	-0.583 (-0.834↔-0.332) n=44	0.186 (-0.058↔0.429) n=44	170.14 (165.70↔174.58) n=43	4.93 (4.43↔5.43) n=43	17.66 (15.12↔20.20) n=44	13.59 (11.60↔15.58) n=44	16.73 (14.34↔19.12) n=44	11.34 (9.76↔12.92) n=44	35.89 (30.97↔40.81) n=44	44.82 (39.21↔50.43) n=44
	HR (n=110)	0.027 (-0.137↔0.192) n=109	-0.006 (-0.160↔0.149) n=109	163.69 (160.53↔166.85) n=108	3.53 (3.19↔3.87) n=109	15.72 (14.43↔17.01) n=108	12.54 (11.41↔13.67) n=108	16.07 (14.81↔17.33) n=108	10.80 (9.92↔11.68) n=108	33.83 (31.21↔36.45) n=108	41.68 (38.64↔44.72) n=108
	LL (n=66)	-0.027 (-0.224↔0.171) n=66	-0.143 (0.357↔0.070) n=66	166.50 (162.45↔170.55) n=66	3.45 (3.09↔3.81) n=65	13.32 (11.33↔15.31) n=66	10.71 (9.28↔12.14) n=66	13.33 (11.43↔15.23) n=66	8.44 (7.46↔9.42) n=66	28.39 (24.54↔32.24) n=66	35.17 (30.80↔39.54) n=66
	LC (n=93)	-0.449 (-0.674↔-0.223) n=93	-0.184 (-0.397↔0.029) n=93	168.85 (165.77↔171.93) n=91	4.45 (4.11↔4.79) n=91	17.01 (15.60↔18.42) n=91	13.04 (11.88↔14.20) n=91	15.45 (14.13↔16.77) n=91	10.97 (10.11↔11.83) n=91	34.93 (32.22↔37.64) n=91	43.02 (39.88↔46.16) n=91
	PP (n=82)	0.046 (-0.140↔0.232) n=82	-0.039 (-0.253↔0.174) n=82	158.49 (154.88↔162.10) n=79	3.40 (3.04↔3.76) n=81	14.01 (12.41↔15.61) n=81	10.67 (9.38↔11.96) n=81	12.43 (11.03↔13.83) n=81	8.26 (7.26↔9.26) n=81	28.51 (25.39↔31.63) n=81	35.09 (31.43↔38.75) n=81
	SP (n=35)	-0.580 (-0.923↔-0.238) n=35	0.138 (-0.168↔0.444) n=35	160.57 (153.07↔168.07) n=35	2.81 (2.26↔3.36) n=32	17.74 (15.59↔19.89) n=35	13.89 (11.75↔16.03) n=35	16.43 (14.23↔18.63) n=35	11.00 (9.32↔12.68) n=35	36.51 (32.09↔40.93) n=35	44.06 (39.02↔49.10) n=35

FIGURE A11.1: Differences between the crude mean and 95% CI of the stratified levels for BFA in the boys Gulu sample at M1

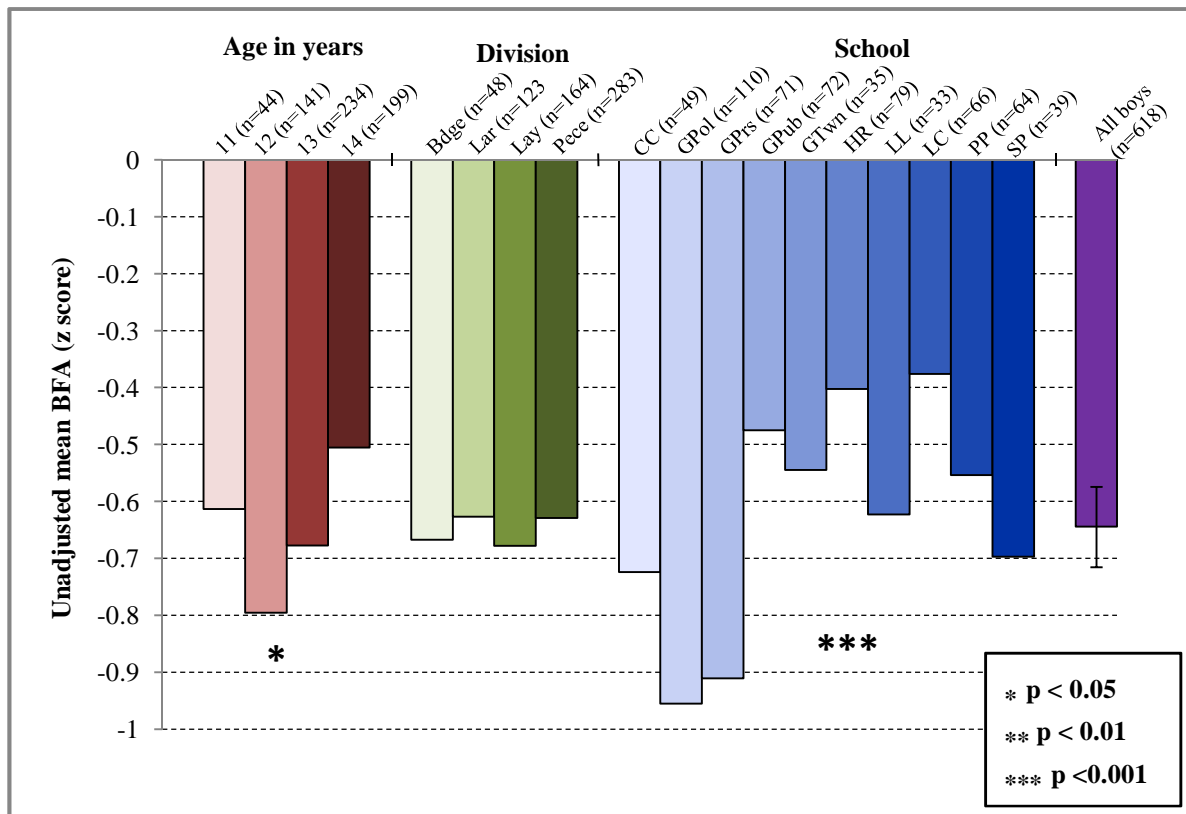


FIGURE A11.2: Differences between the crude mean and 95% CI of the stratified levels for BFA in the girls Gulu sample at M1

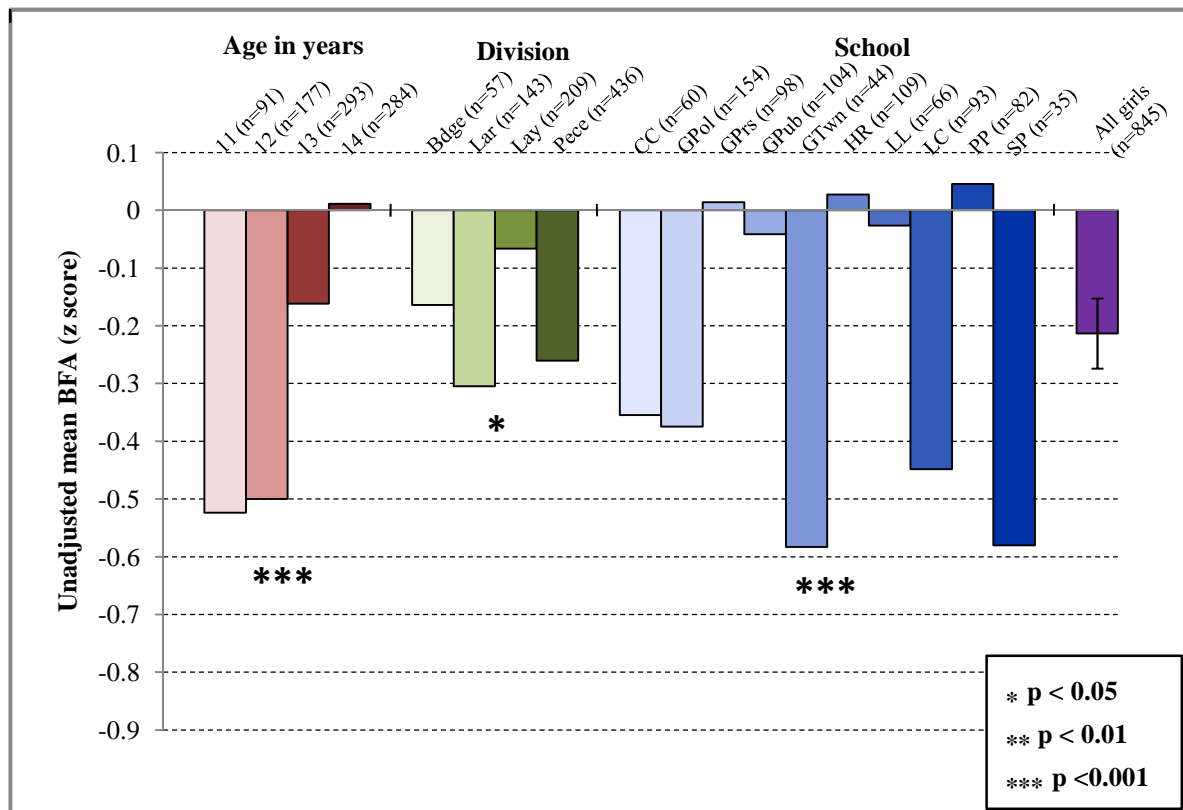


FIGURE A11.3: Differences between the crude mean and 95% CI of the stratified levels for HFA in the boys Gulu sample at M1

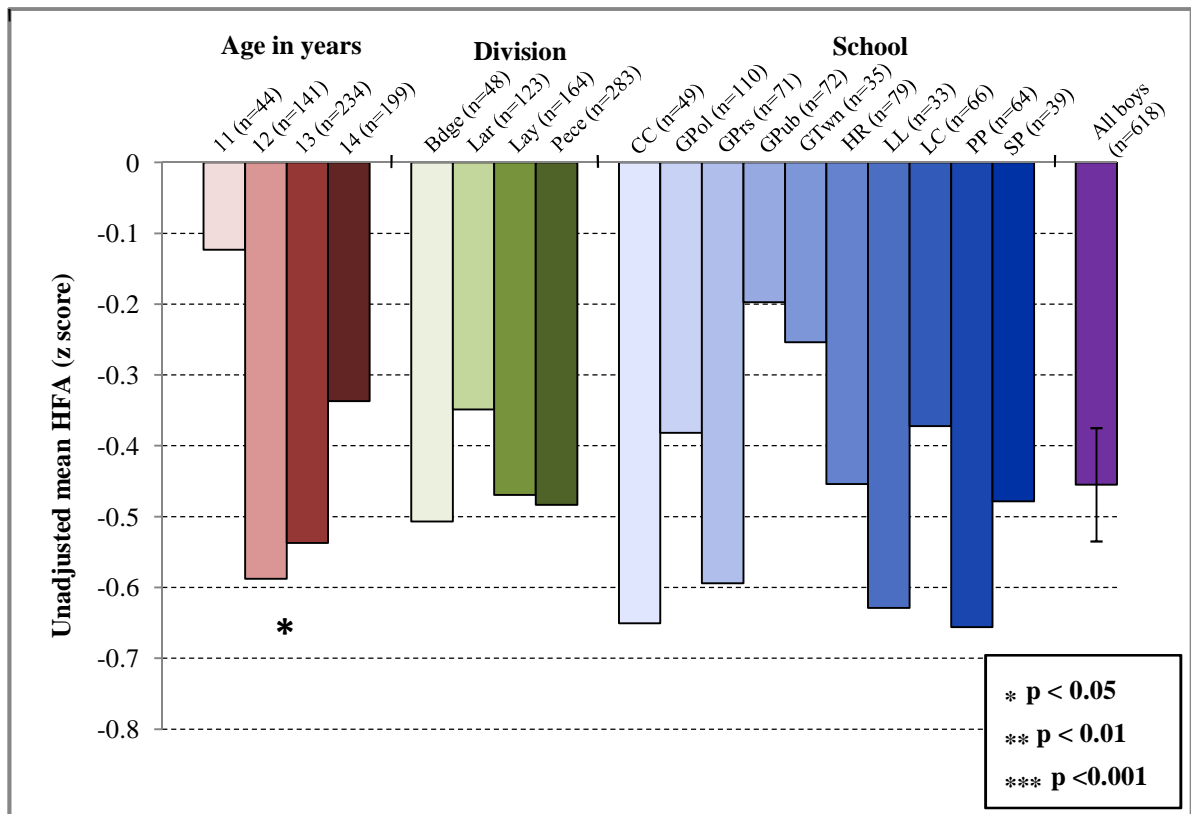


FIGURE A11.4: Differences between the crude mean and 95% CI of the stratified levels for HFA in the girls Gulu sample at M1

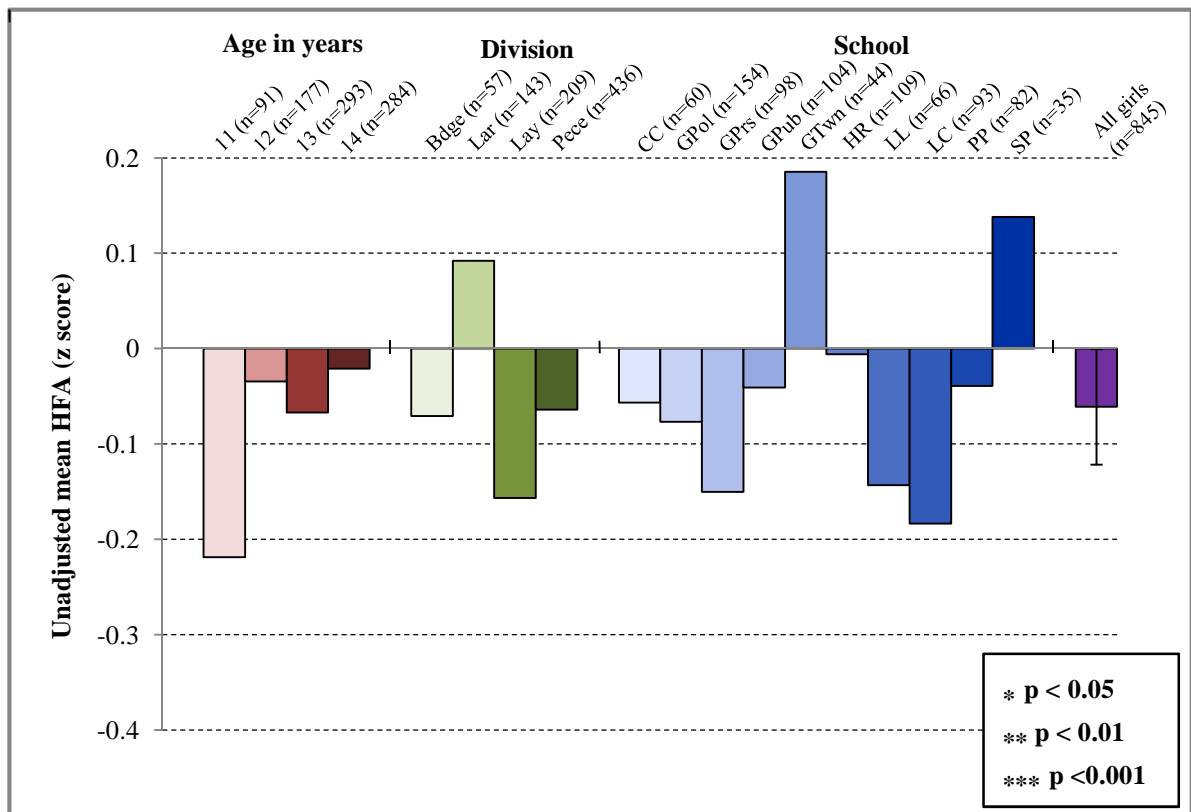


FIGURE A11.5: Differences between the crude mean and 95% CI of the stratified levels for the SBJ in the boys Gulu sample at M1

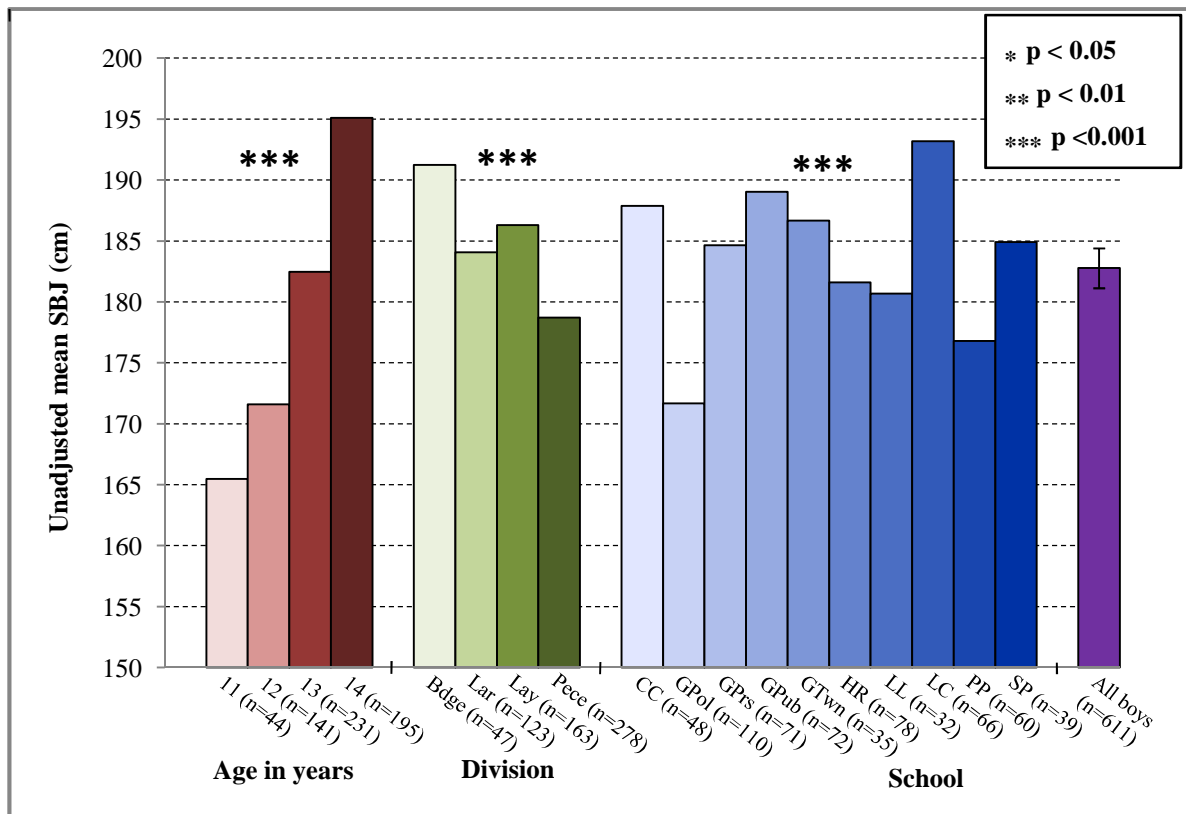


FIGURE A11.6: Differences between the crude mean and 95% CI of the stratified levels for the SBJ in the girls Gulu sample at M1

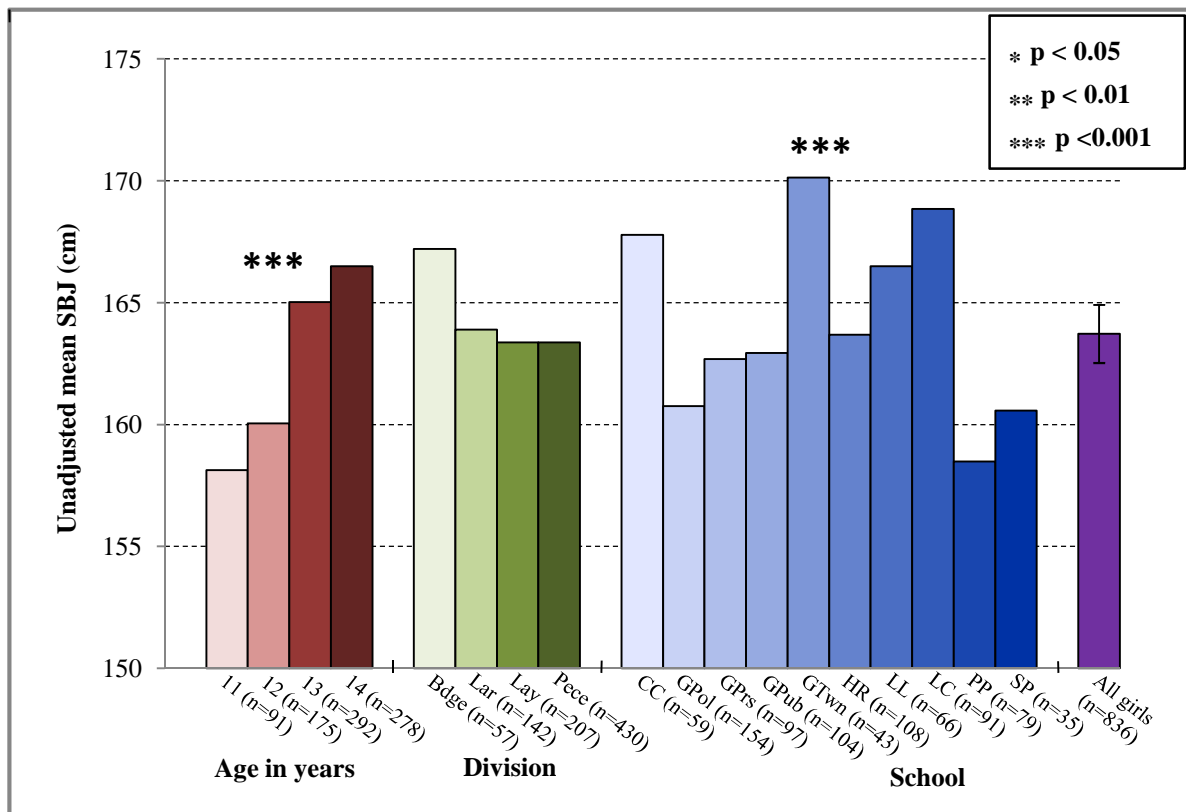


FIGURE A11.7: Differences between the crude mean and 95% CI of the stratified levels for the MSFT in the boys Gulu sample at M1

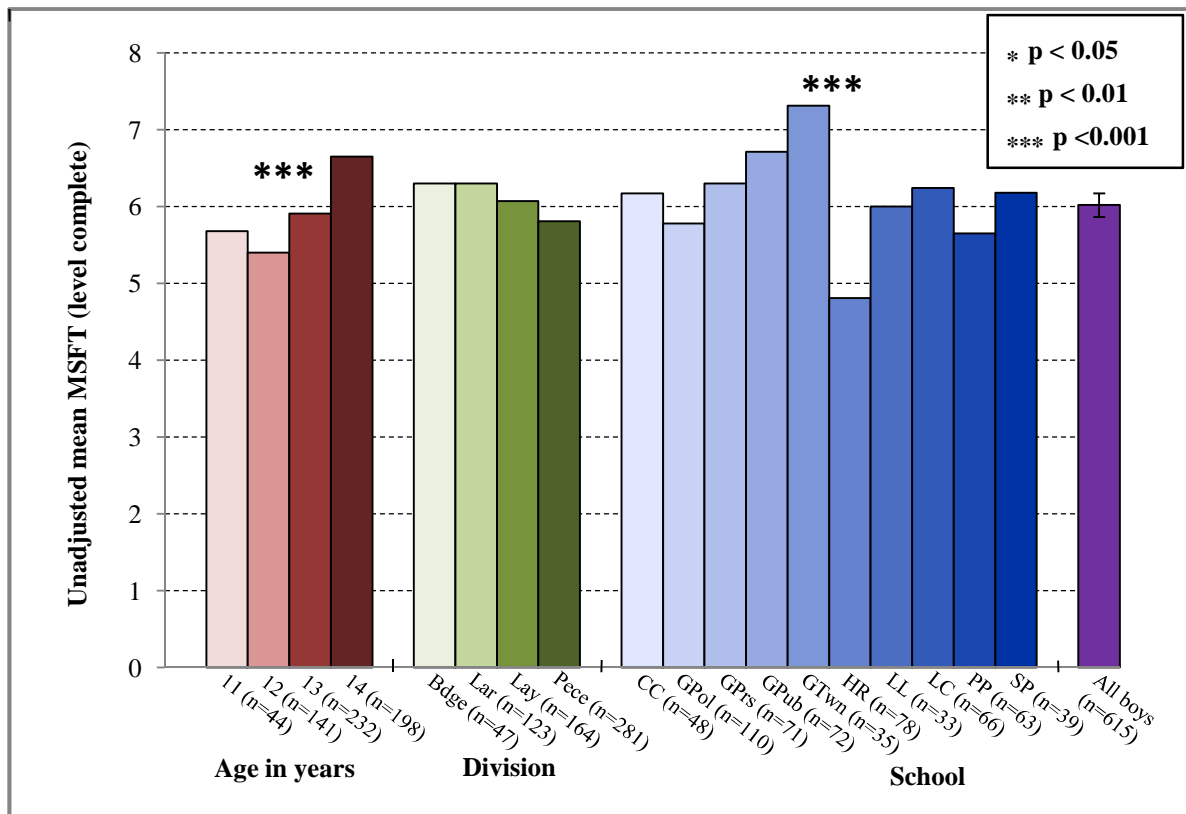


FIGURE A11.8: Differences between the crude mean and 95% CI of the stratified levels for the MSFT in the girls Gulu sample at M1

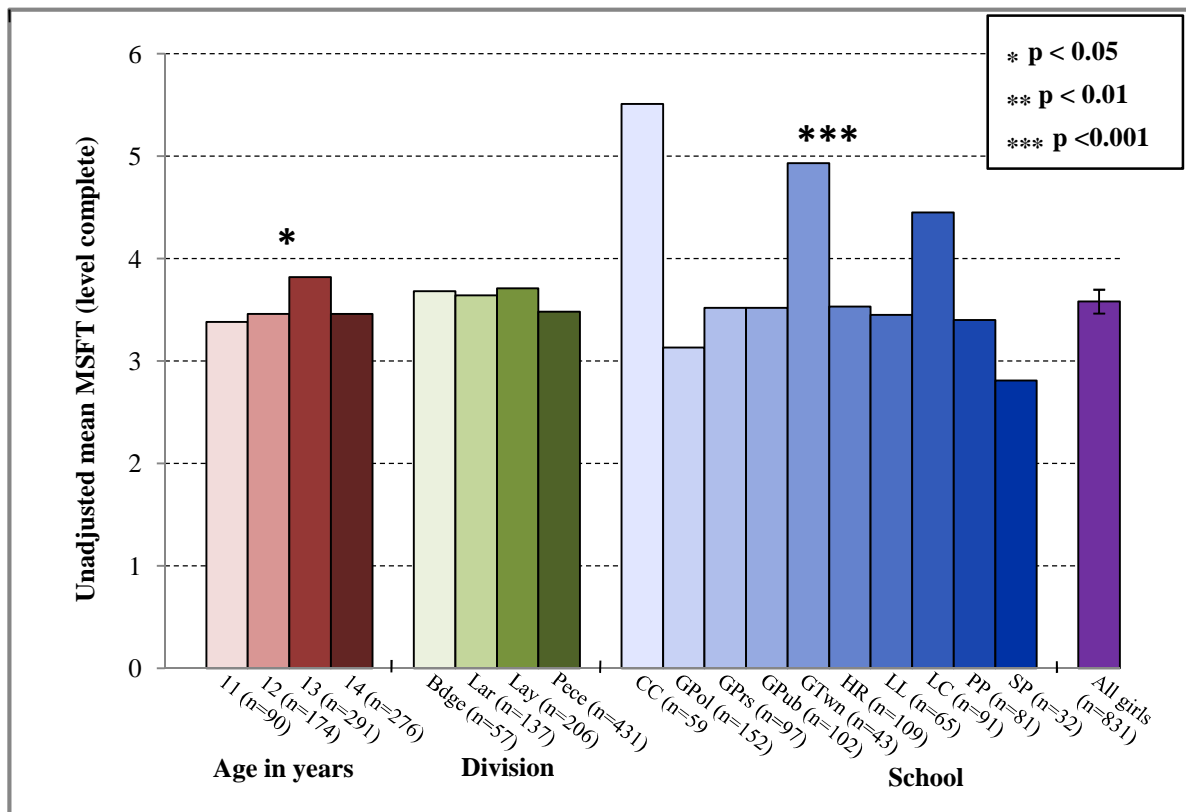


FIGURE A11.9: Differences between the crude mean and 95% CI of the stratified levels for TT in the boys Gulu sample at M1

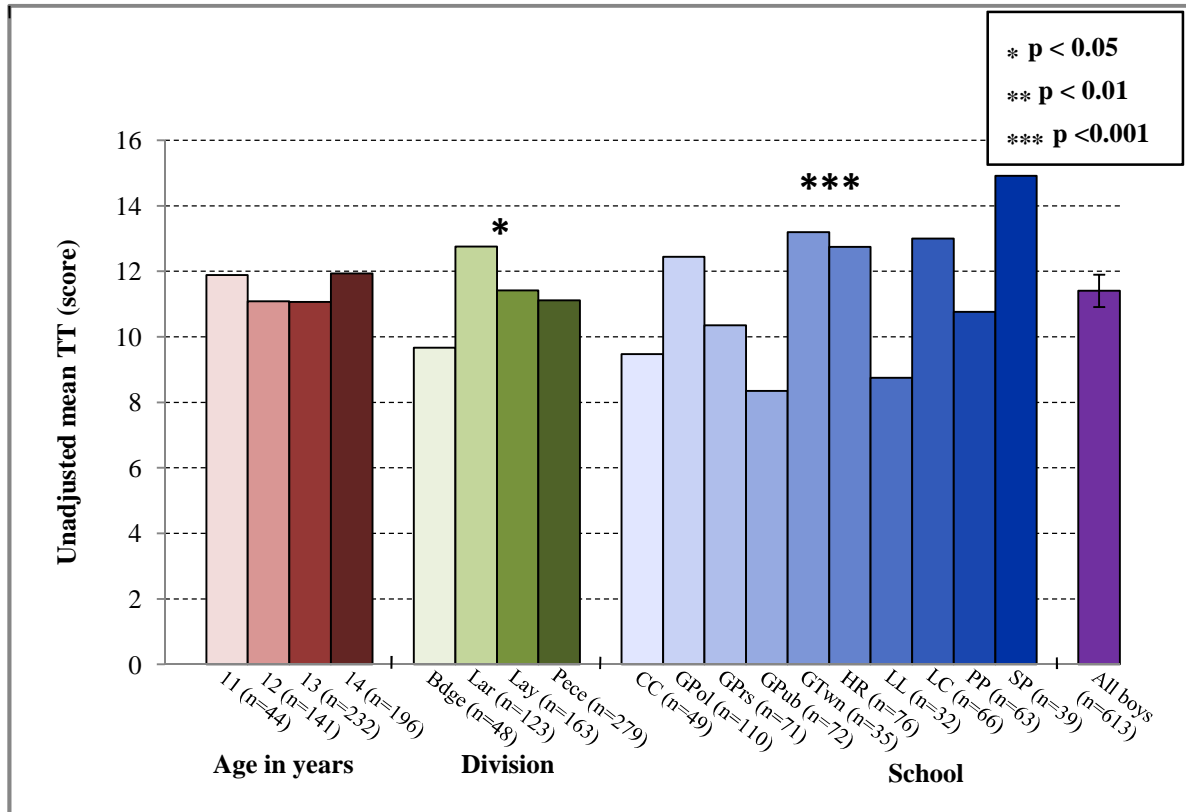


FIGURE A11.10: Differences between the crude mean and 95% CI of the stratified levels for TT in the girls Gulu sample at M1

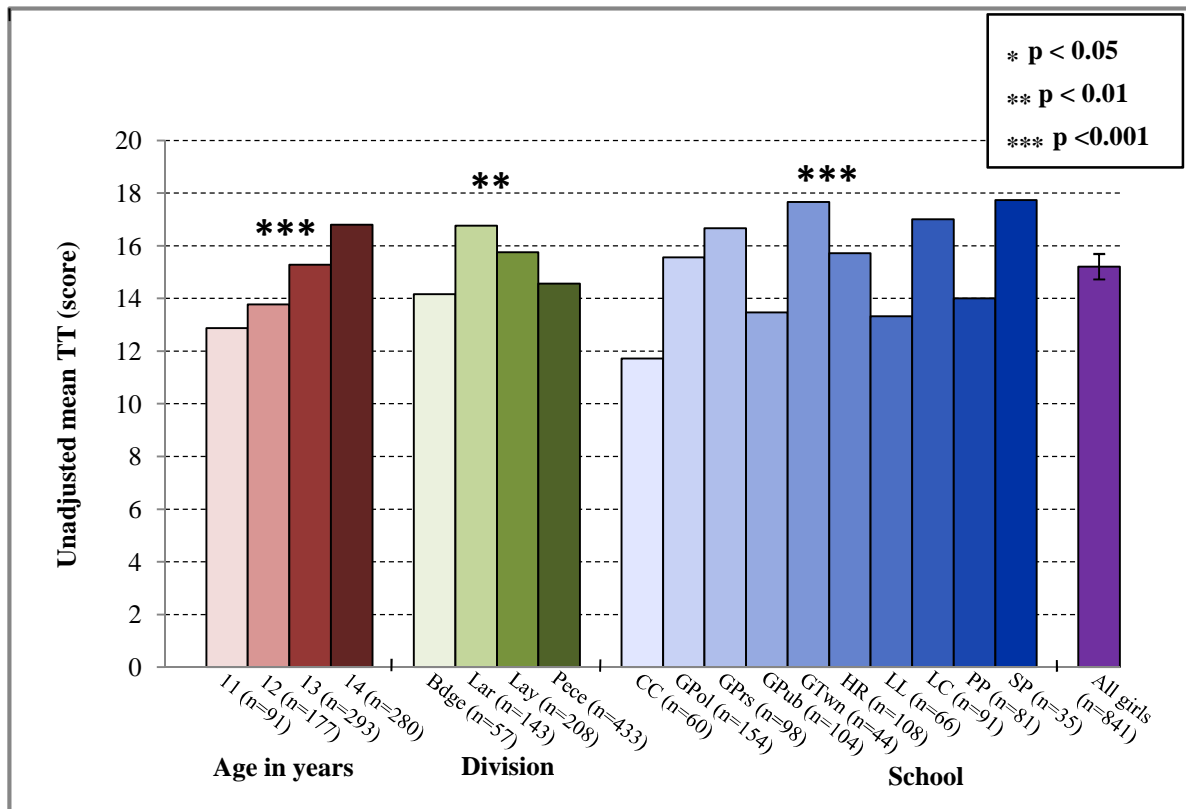


FIGURE A11.11: Differences between the crude mean and 95% CI of the stratified levels for K in the boys Gulu sample at M1

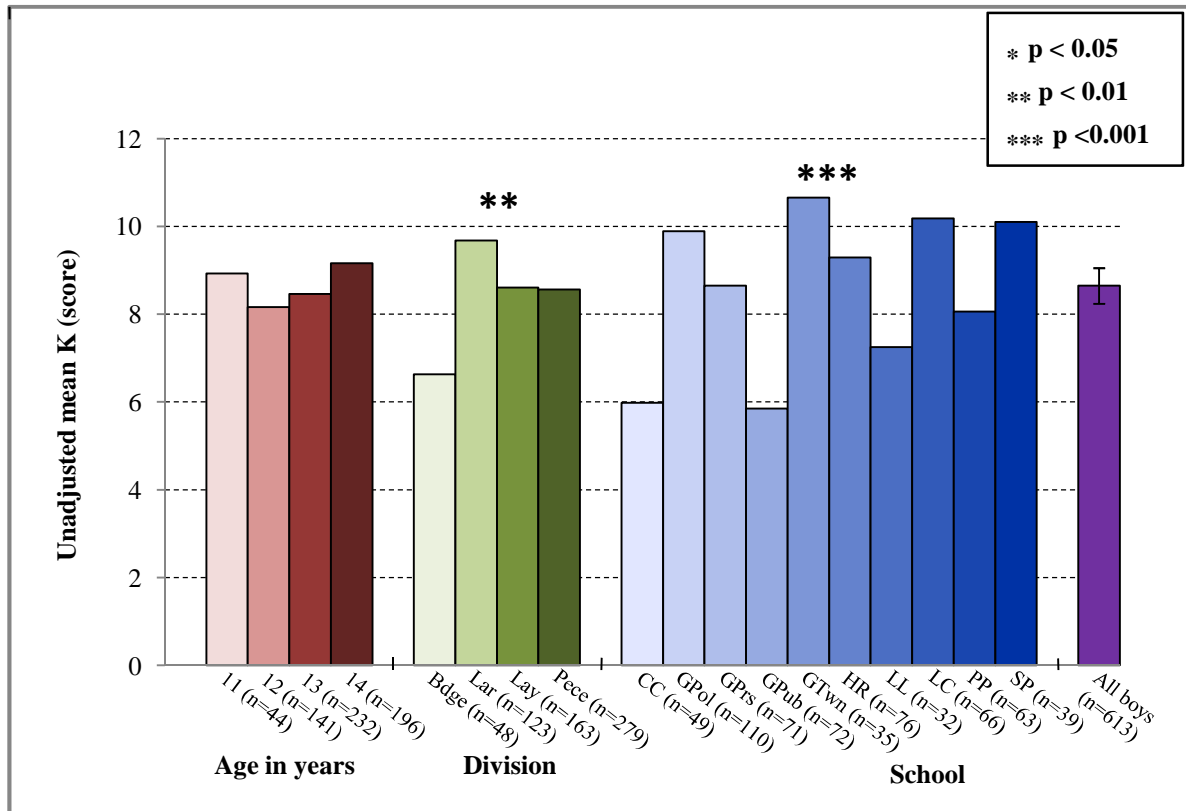


FIGURE A11.12: Differences between the crude mean and 95% CI of the stratified levels for K in the girls Gulu sample at M1

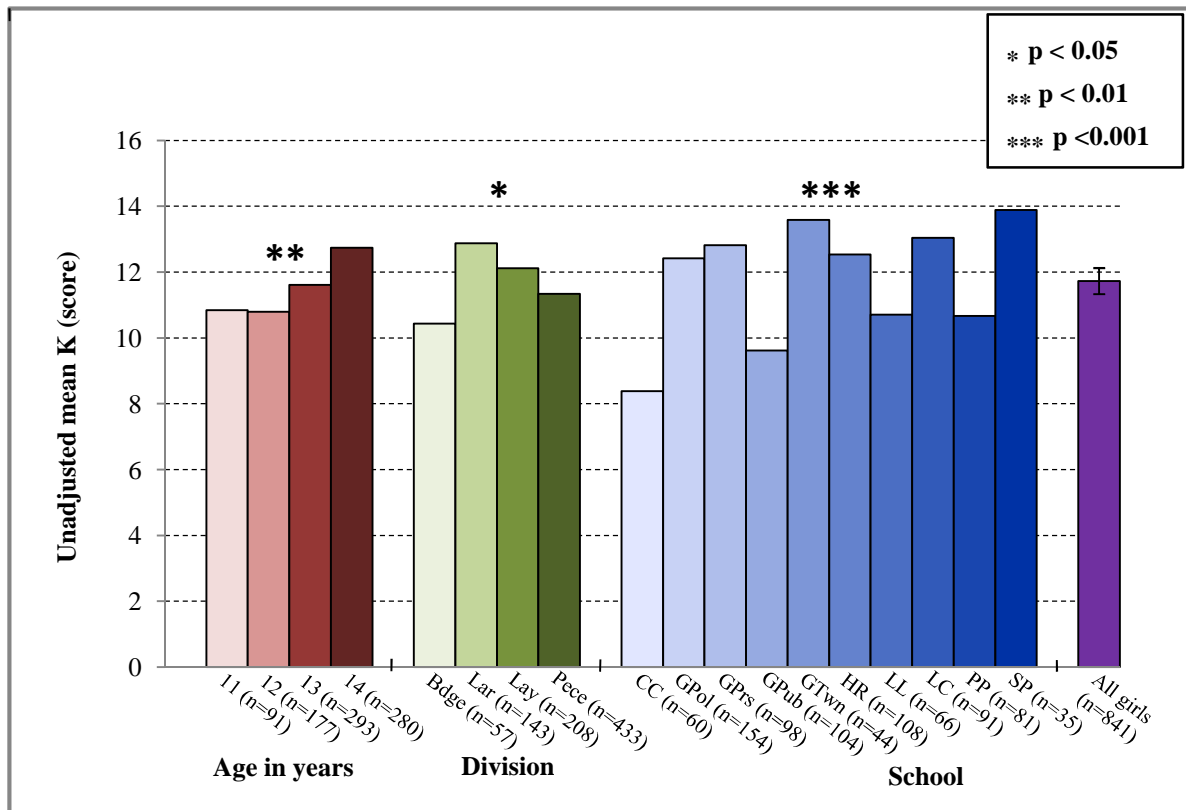


FIGURE A11.13: Differences between the crude mean and 95% CI of the stratified levels for P in the boys Gulu sample at M1

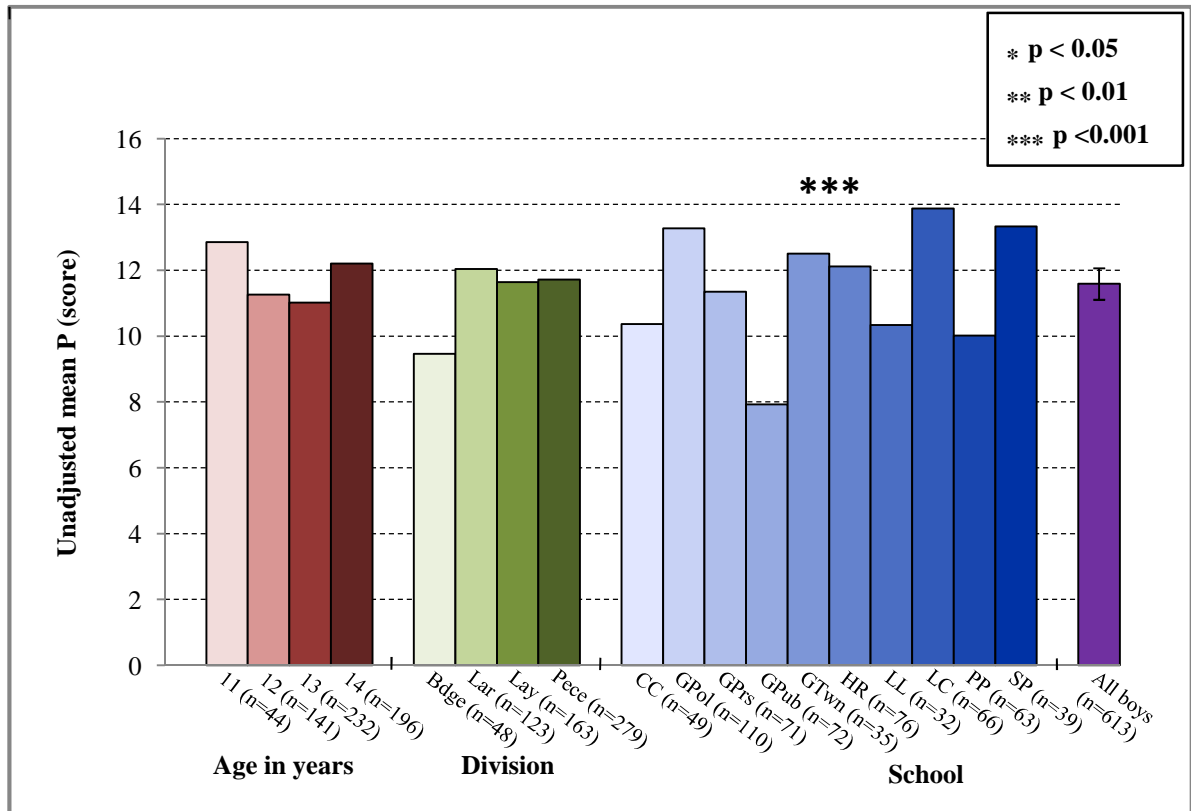


FIGURE A11.14: Differences between the crude mean and 95% CI of the stratified levels for P in the girls Gulu sample at M1

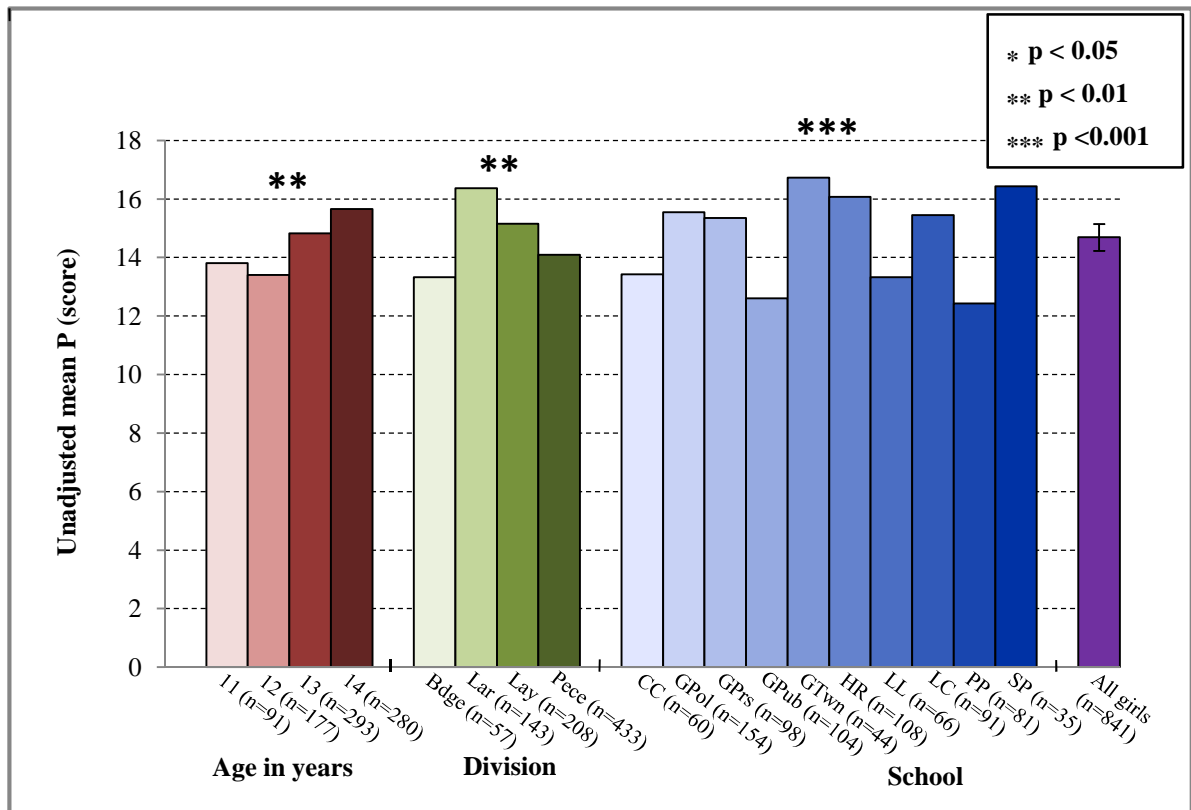


FIGURE A11.15: Differences between the crude mean and 95% CI of the stratified levels for ML in the boys Gulu sample at M1

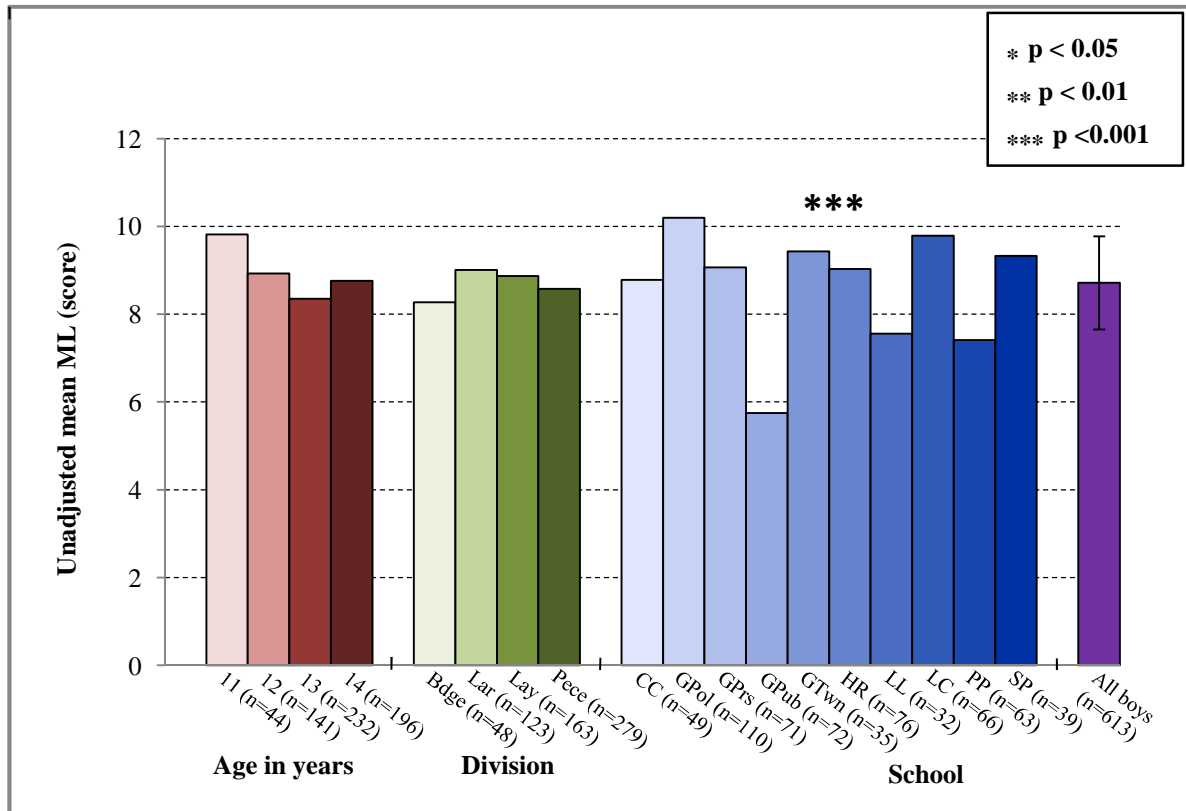


FIGURE A11.16: Differences between the crude mean and 95% CI of the stratified levels for ML in the girls Gulu sample at M1

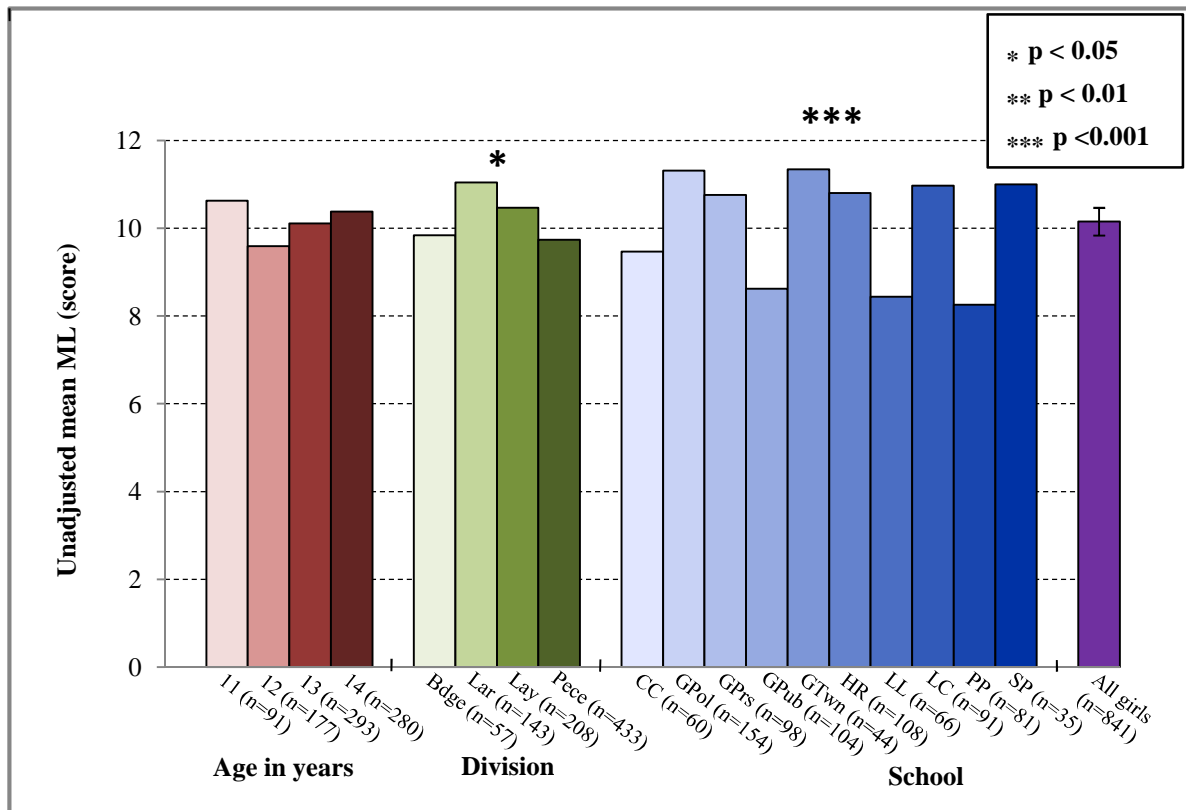


FIGURE A11.17: Differences between the crude mean and 95% CI of the stratified levels for DT in the boys Gulu sample at M1

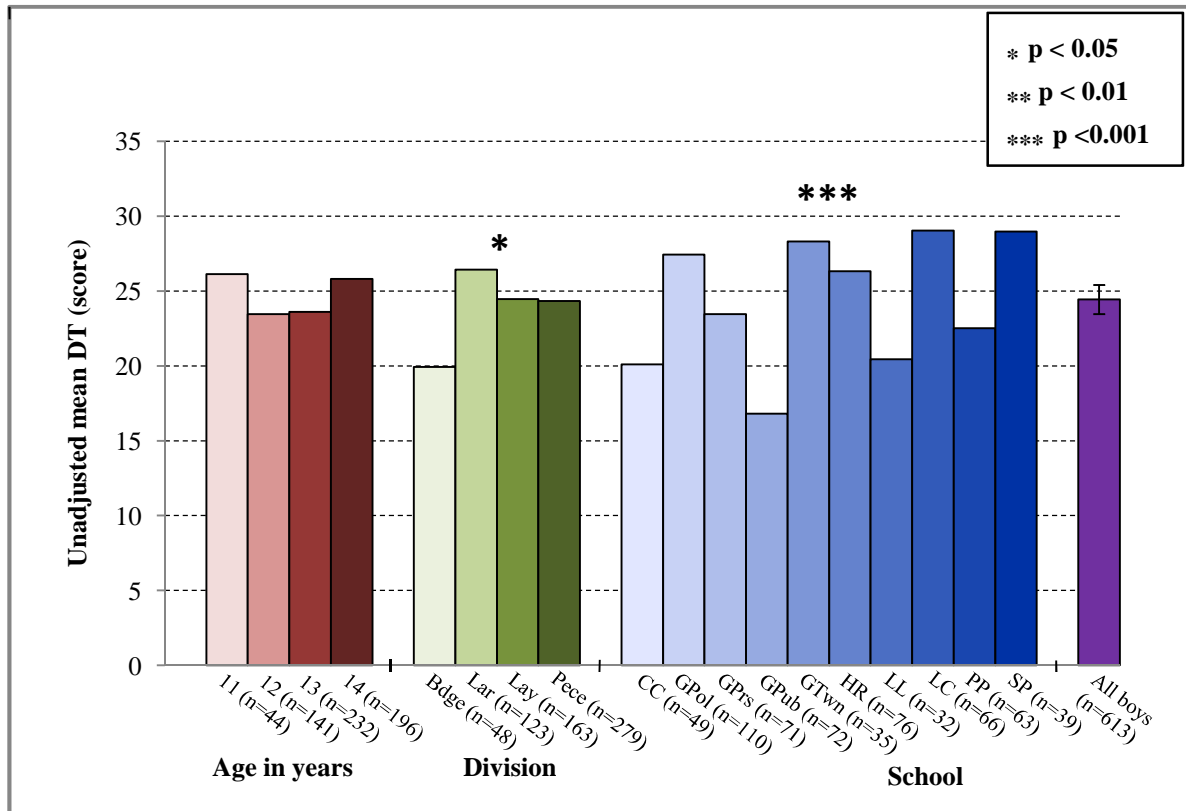


FIGURE A11.18: Differences between the crude mean and 95% CI of the stratified levels for DT in the girls Gulu sample at M1

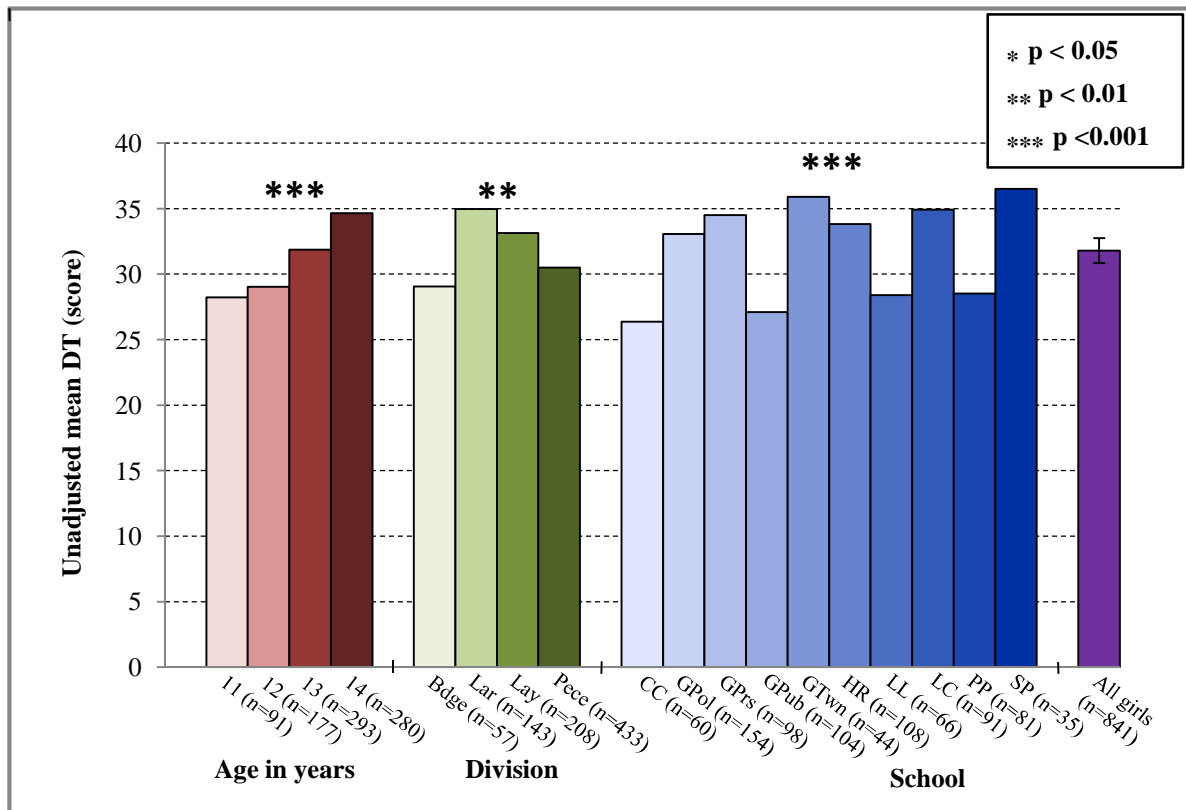


FIGURE A11.19: Differences between the crude mean and 95% CI of the stratified levels for AT in the boys Gulu sample at M1

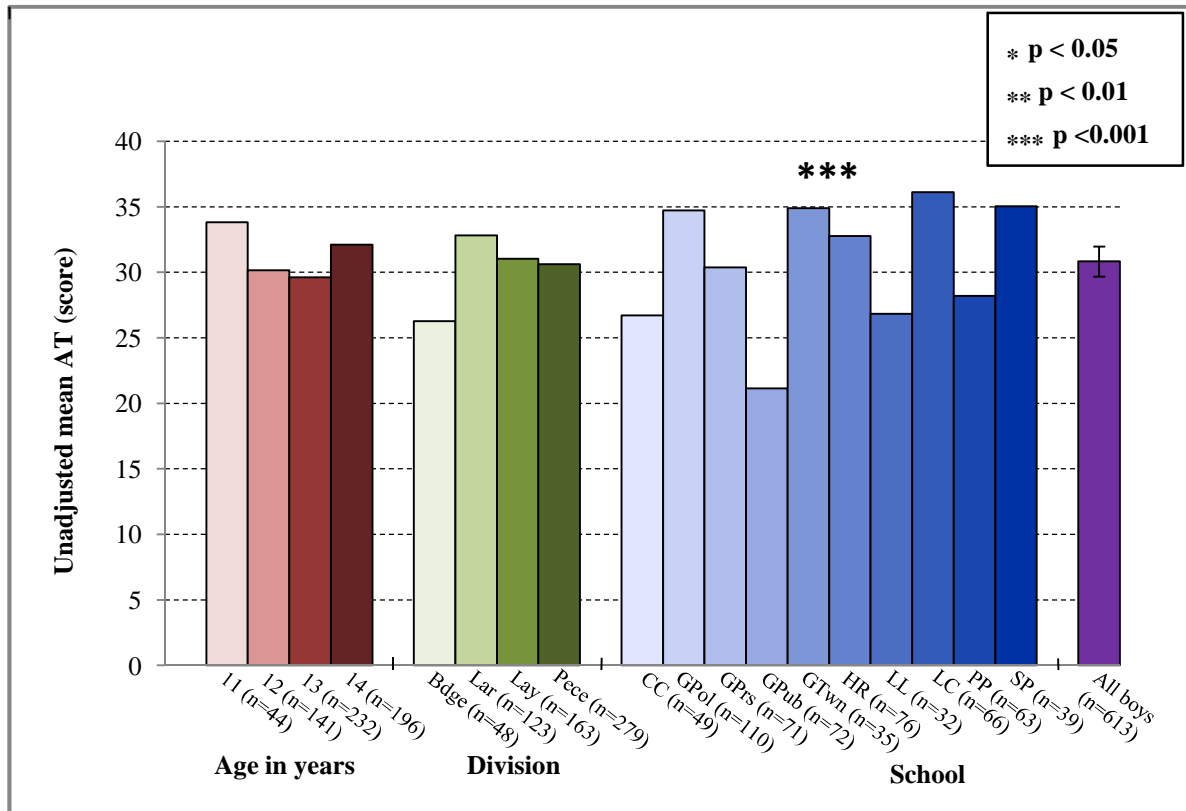


FIGURE A11.20: Differences between the crude mean and 95% CI of the stratified levels for AT in the girls Gulu sample at M1

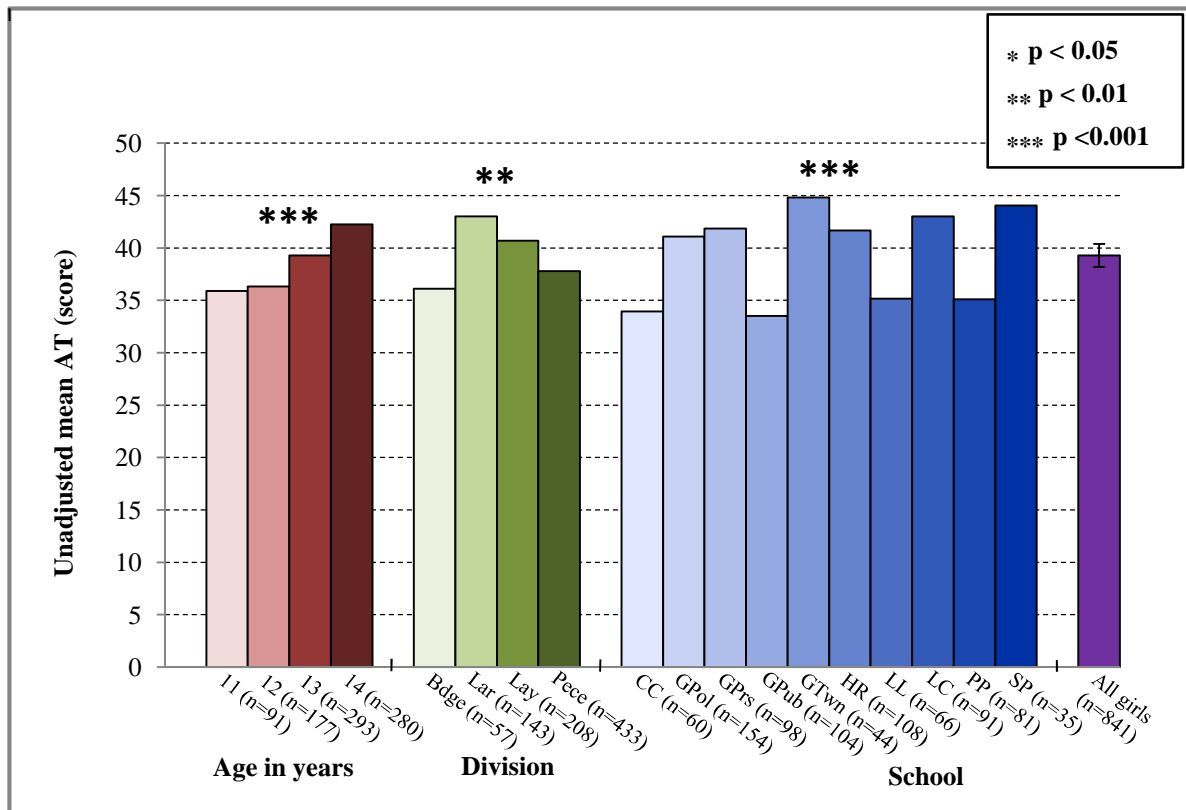


TABLE A11.3: One-way ANOVA comparing levels of the crude PF outcomes in the Gulu sample at M1 - stratified by age, division and school

Stratified independent variable	Gender	BFA (z score)	HFA (z score)	SBJ (cm)	MSFT (level complete)
Age in years	Boys	F=3.101 *	F=3.830 *	F=62.009 ***	F=12.671 ***
	Girls	F=16.688 ***	F=1.181	F=8.709 ***	F=3.069 *
Division	Boys	F=0.131	F=0.574	F=8.340 ***	F=2.260
	Girls	F=2.818 *	F=2.191	F=0.839	F=1.033
School	Boys	F=4.124 ***	F=1.533	F=8.163 ***	F=7.326 ***
	Girls	F=5.957 ***	F=0.983	F=3.578 ***	F=8.233 ***

TABLE A11.4: One-way ANOVA comparing levels of the crude MH outcomes in the Gulu sample at M1 - stratified by age, division and school

Stratified independent variable	Gender	TT (score)	K (score)	P (score)	ML (score)	DT (score)	AT (score)
Age in years	Boys	F=0.914	F=1.245	F=2.187	F=1.465	F=1.748	F=1.809
	Girls	F=10.571 ***	F=5.063 **	F=4.636 **	F=1.387	F=8.334 ***	F=6.558 ***
Division	Boys	F=3.444 *	F=4.253 **	F=2.285	F=0.463	F=3.248 *	F=2.417
	Girls	F=4.263 **	F=3.745 *	F=5.174 **	F=3.297 *	F=5.111 **	F=5.059 **
School	Boys	F=6.865 ***	F=7.276 ***	F=6.879 ***	F=6.633 ***	F=7.973 ***	F=7.990 ***
	Girls	F=5.580 ***	F=6.982 ***	F=4.450 ***	F=6.419 ***	F=5.653 ***	F=5.736 ***

NOTE: Means compared using a model III sum of squares analysis (unbalanced groups with no empty cells for the selected fixed factors).

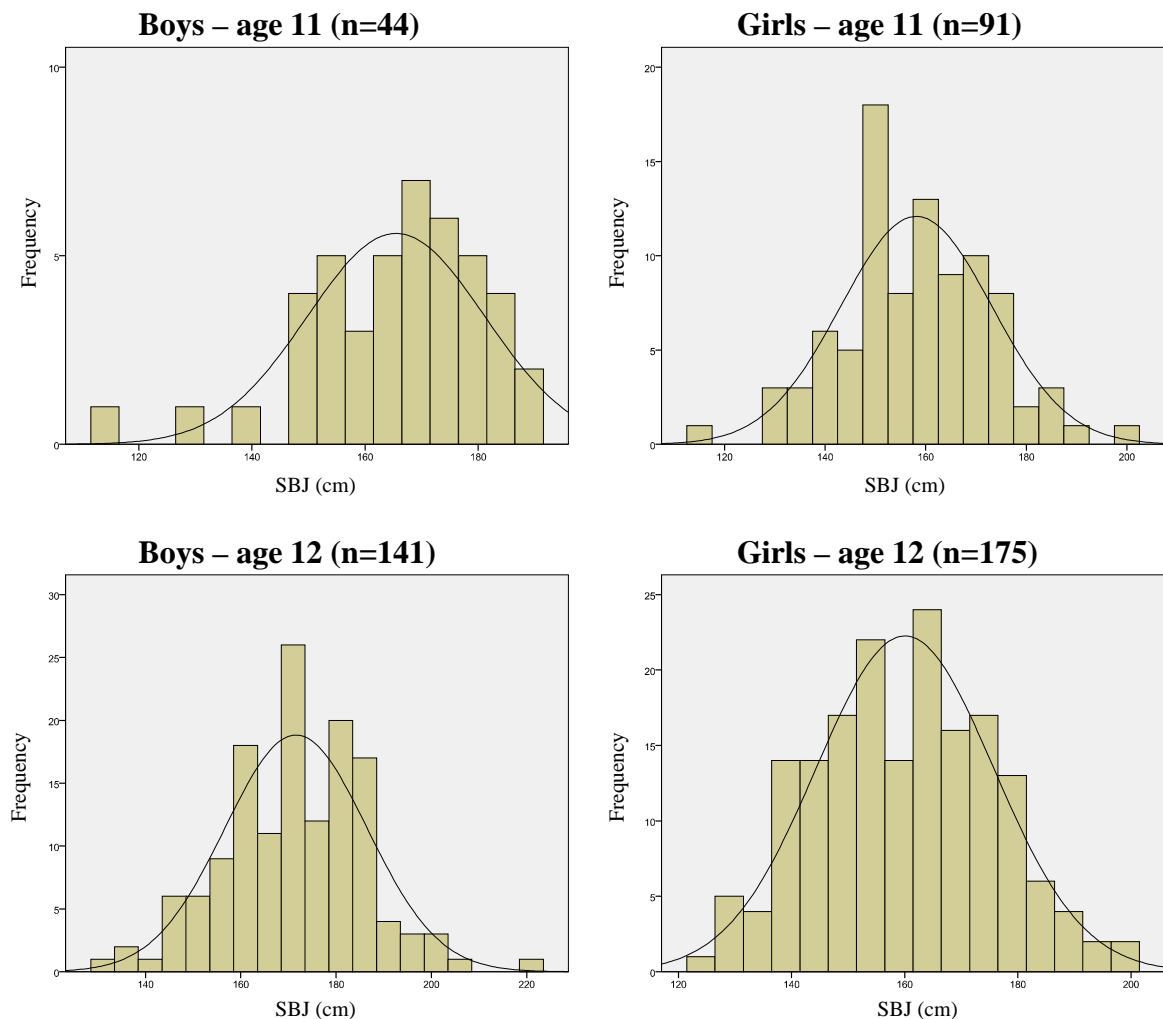
NOTE: * Denotes a significant difference between the stratified groups (* p<0.05, ** p<0.01, *** p<0.001).

APPENDIX A12: Comparison to Normative Data at M1 (Assumptions)

The method used to compare BFA and HFA to global normative values was descriptive (i.e. the proportion of the sample in each category according to the 2007 WHO standards) and did not require the testing of assumptions for the statistical analyses.

The statistical methods used to compare the baseline results to global normative values for the SBJ and MSFT assumed an approximately normally distributed sample for each age group. This was assessed visually using a histogram of the crude results for the eligible sample stratified according to age for each of the dependent variables (FIGURES A12.1 & A12.2).

FIGURE A12.1: Histograms of the distribution of the crude SBJ measurements in the Gulu sample at M1 - stratified by gender and age



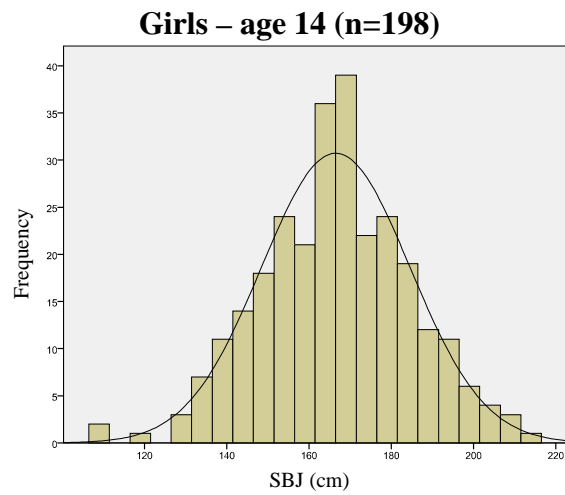
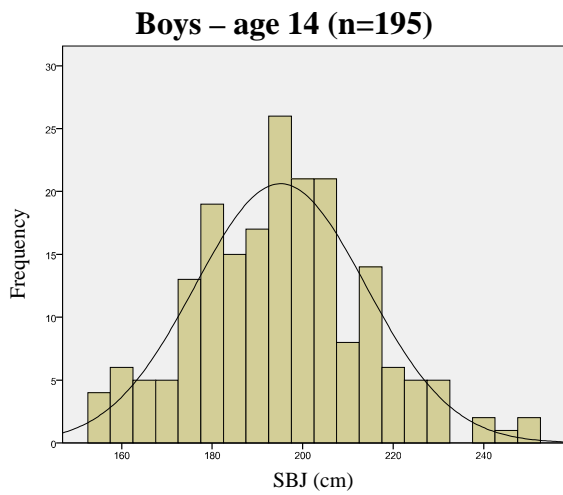
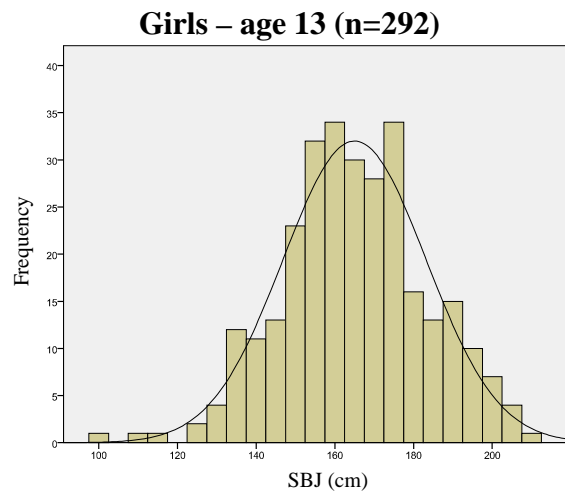
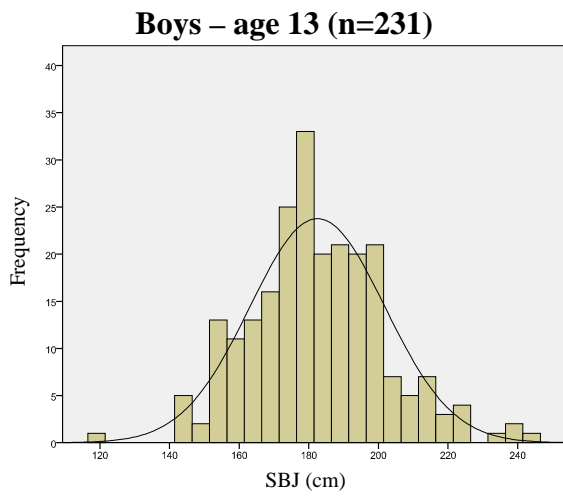
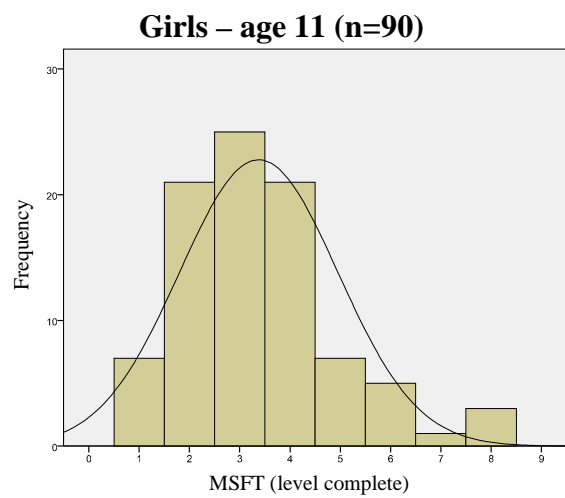
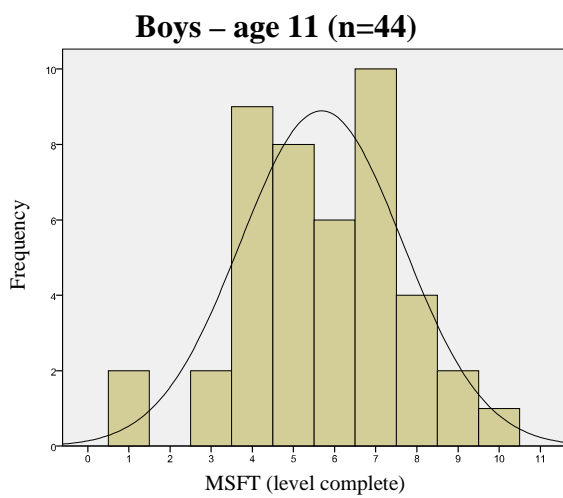
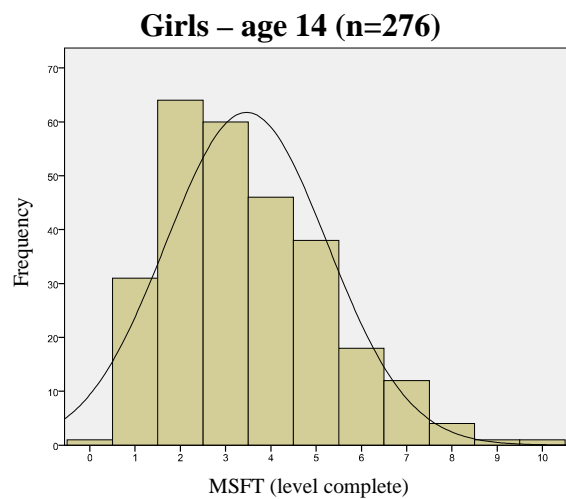
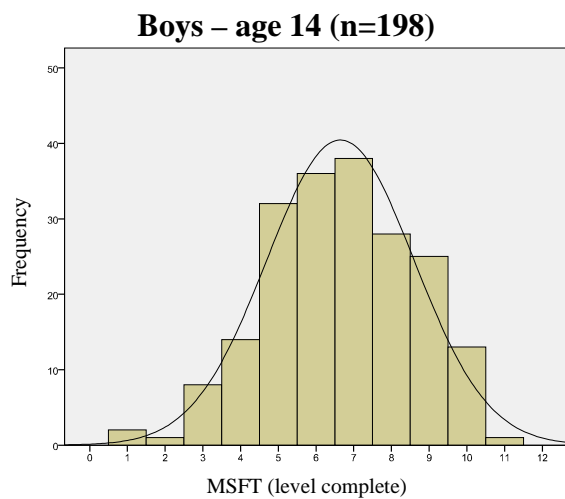
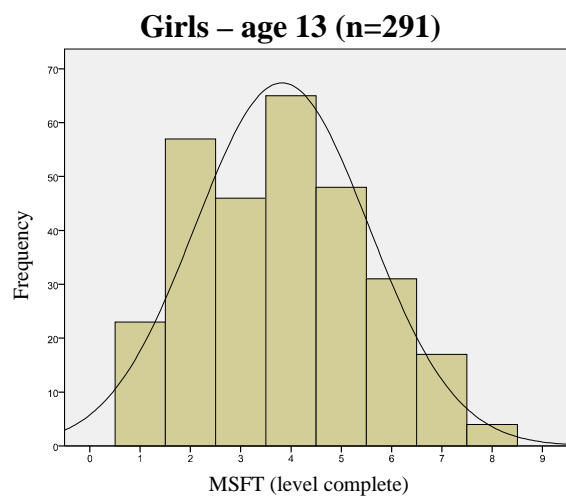
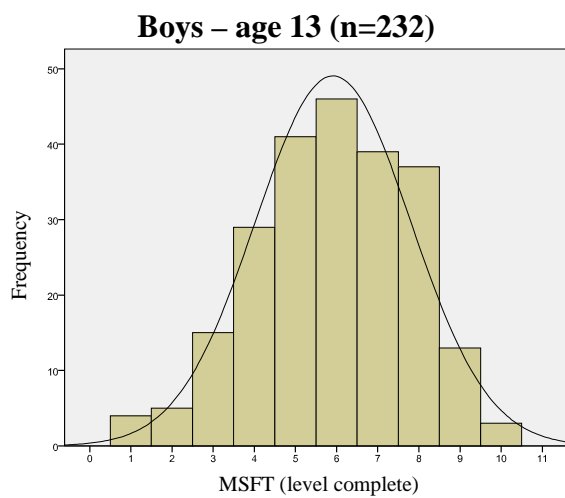
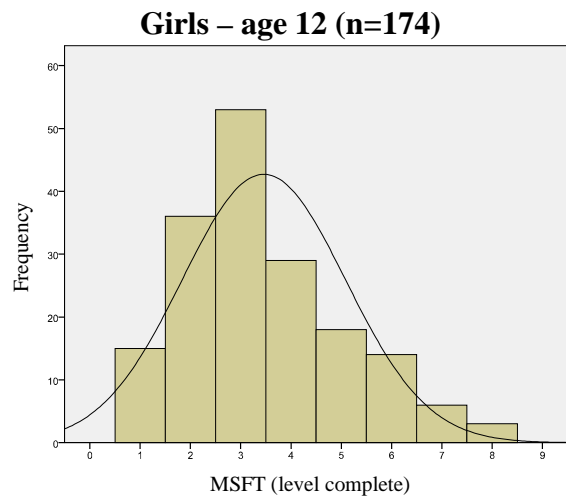
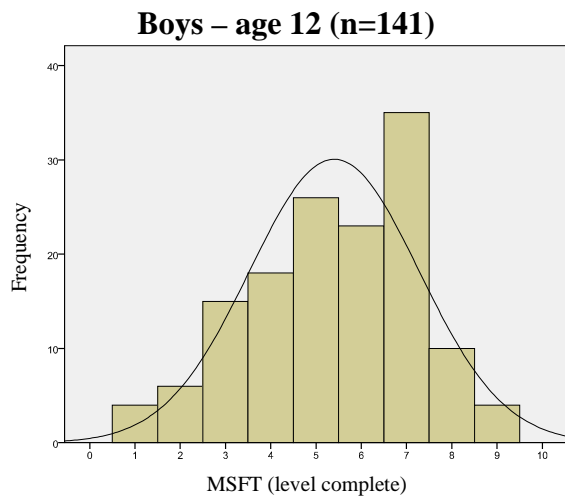


FIGURE A12.2: Histograms of the distribution of the crude MSFT measurements in the Gulu sample at M1 - stratified by gender and age





Although the distribution of all of the dependent variables appeared approximately normal, there were several deviations that warranted further assessment. Statistical tests for normality of the crude results were also completed (TABLE A12.1)

TABLE A12.1: Statistical tests for normality of the crude SBJ and MSFT results in the Gulu sample at M1 - stratified by gender and age

Gender	Outcome variable	Age	n	Skew		Kurtosis		Shapiro Wilk	
				Test stat	SES	Test stat	SEK	Test stat	p
Boys	SBJ (cm)	11	44	<i>-1.105</i>	<i>0.357</i>	1.843	0.702	<i>0.934</i>	<i>0.014</i>
		12	141	-0.012	0.204	0.415	0.406	0.993	0.713
		13	231	0.268	0.160	0.542	0.319	0.989	0.085
		14	195	0.260	0.174	0.163	0.346	0.991	0.227
	MSFT (level complete)	11	44	-0.198	0.357	0.079	0.702	0.963	0.165
		12	141	-0.366	0.204	-0.487	0.406	<i>0.952</i>	<i><0.001</i>
		13	232	-0.265	0.160	-0.313	0.318	<i>0.967</i>	<i><0.001</i>
		14	198	-0.200	0.173	-0.273	0.344	<i>0.969</i>	<i><0.001</i>
Girls	SBJ (cm)	11	91	0.052	0.253	0.287	0.500	0.994	0.962
		12	175	0.086	0.184	-0.486	0.365	0.993	0.540
		13	292	-0.151	0.143	0.193	0.284	0.993	0.214
		14	278	-0.083	0.146	0.138	0.291	0.996	0.675
	MSFT (level complete)	11	90	<i>0.934</i>	<i>0.254</i>	<i>1.041</i>	<i>0.503</i>	<i>0.907</i>	<i><0.001</i>
		12	174	<i>0.689</i>	<i>0.184</i>	0.048	0.366	<i>0.925</i>	<i><0.001</i>
		13	291	0.238	0.143	<i>-0.693</i>	<i>0.285</i>	<i>0.949</i>	<i><0.001</i>
		14	276	<i>0.697</i>	<i>0.147</i>	0.229	0.292	<i>0.935</i>	<i><0.001</i>

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12].

The assumption of HOV for the SBJ and MSFT was assessed using Hartley's test. This involved calculating the ratio of the largest sample variance to the smallest sample variance to produce an F_{\max} score for the crude results of each dependent variable (TABLE A12.2). The F_{\max} score was then compared to tabulated critical values to provide an estimate for HOV ^[16]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$). Hartley's test was used because it was not possible to complete the Levene or Brown-Forsythe tests due to limited access to the normative data.

TABLE A12.2: Statistical tests of the HOV between the crude results of the Gulu sample at M1 and the normative data for the SBJ and MSFT - stratified by gender and age

Outcome variable	Gender	Age	Norm data		Gulu sample		F _{max} score	Summary
			n	Variance	n	Variance		
SBJ (cm)	Boys	11	10045	529	44	246	2.15	<i>C</i>
		12	15313	529	141	223	2.37	<i>C</i>
		13	12052	576	231	376	1.53	<i>C</i>
		14	12415	625	195	356	1.76	<i>C</i>
	Girls	11	9933	484	91	225	2.15	<i>C</i>
		12	16198	529	175	246	2.15	<i>C</i>
		13	12143	529	292	331	1.60	<i>C</i>
		14	12139	529	278	326	1.62	<i>C</i>
MSFT (level complete)	Boys	11	15480	1.445	44	1.094	1.32	√
		12	24544	1.530	141	0.945	1.62	<i>C</i>
		13	27535	1.698	232	0.937	1.81	<i>C</i>
		14	27106	1.750	198	0.984	1.78	<i>C</i>
	Girls	11	15446	1.022	90	0.733	1.39	<i>C</i>
		12	24255	1.084	174	0.787	1.38	<i>C</i>
		13	27110	1.126	291	0.872	1.29	<i>C</i>
		14	24924	1.179	276	1.288	1.09	√

NOTE: Results in **BOLD ITALICS** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

NOTE: In the summary column - √ denotes satisfactory HOV; C indicates a "conservative" statistical test [12,14].

The assumption of HOV held for the MSFT for the 11-year-old boys and 14-year-old girls when comparing crude baseline measures (M1) vs. normative data. The statistical analysis could proceed with little concern when the HOV assumption was satisfied [12,14].

The violation of the HOV assumption for all other samples in this study was offset by the large sample size [12,14]. However, when comparing baseline measures (M1) to the normative data, the unbalanced sample sizes may have significantly compromised the statistical tests. In this case, the group with the biggest n had a larger variance. The analysis was referred to as a “conservative” statistical test and the risk of a type I error was decreased [12,14].

A consistent statistical approach was applied for all outcome variables when comparing crude baseline measures (M1) vs. normative data. The interpretation of any non-significant statistical findings was treated with caution when a “conservative” test was used ^[12,14].

APPENDIX A13: History of Abduction at M1 (Assumptions)

The statistical methods used to assess whether the PF and MH of the adolescents who were abducted during the war were different to their non-abducted peers assumed that both groups were approximately normally distributed. This was initially assessed visually using a histogram of the eligible sample stratified according to history of abduction for the crude results of each of the dependent variables (FIGURES A13.1 – A13.10).

FIGURE A13.1: Histograms of the distribution of the crude BFA results in the Gulu sample at M1 - stratified by gender and history of abduction

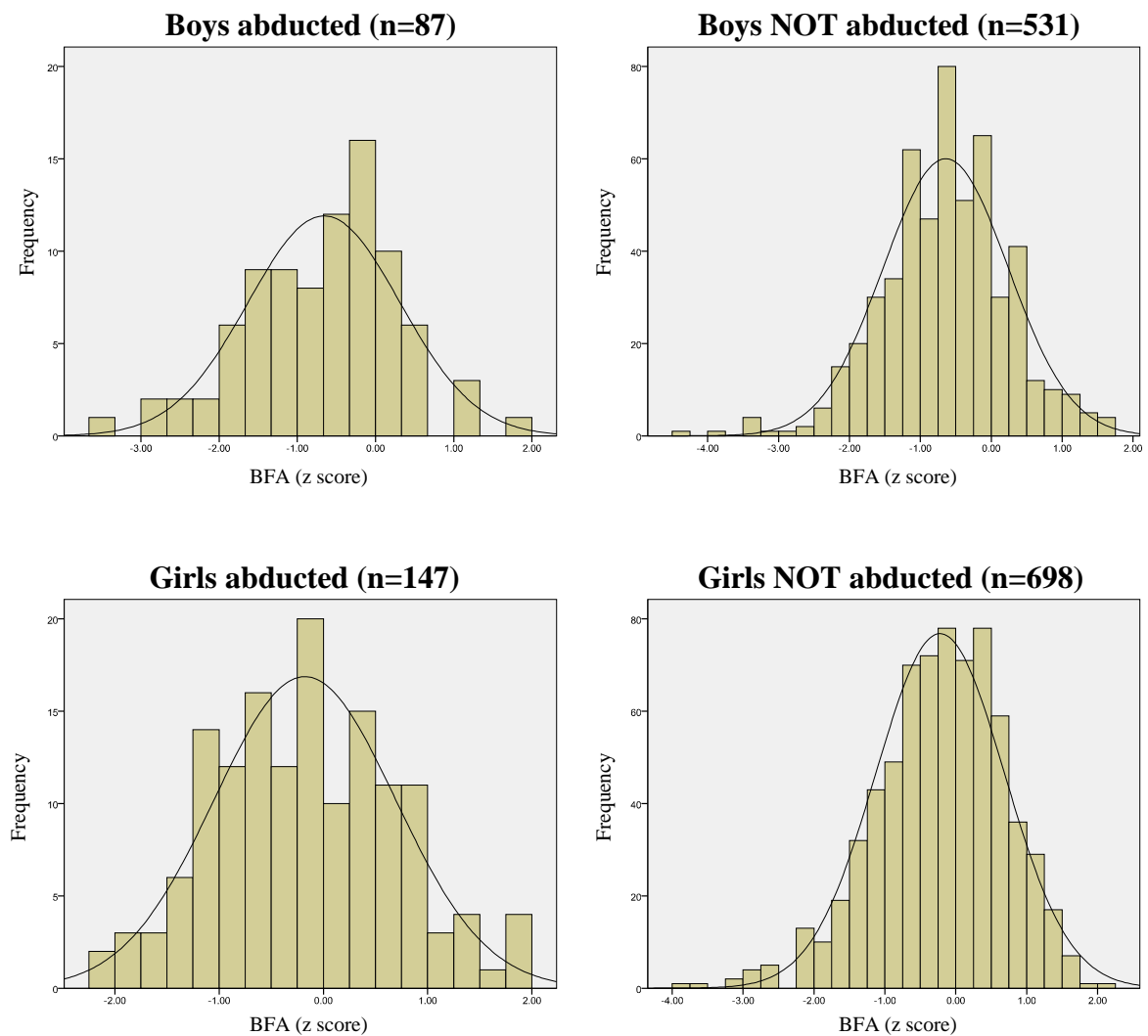


FIGURE A13.2: Histograms of the distribution of the crude HFA results in the Gulu sample at M1 - stratified by gender and history of abduction

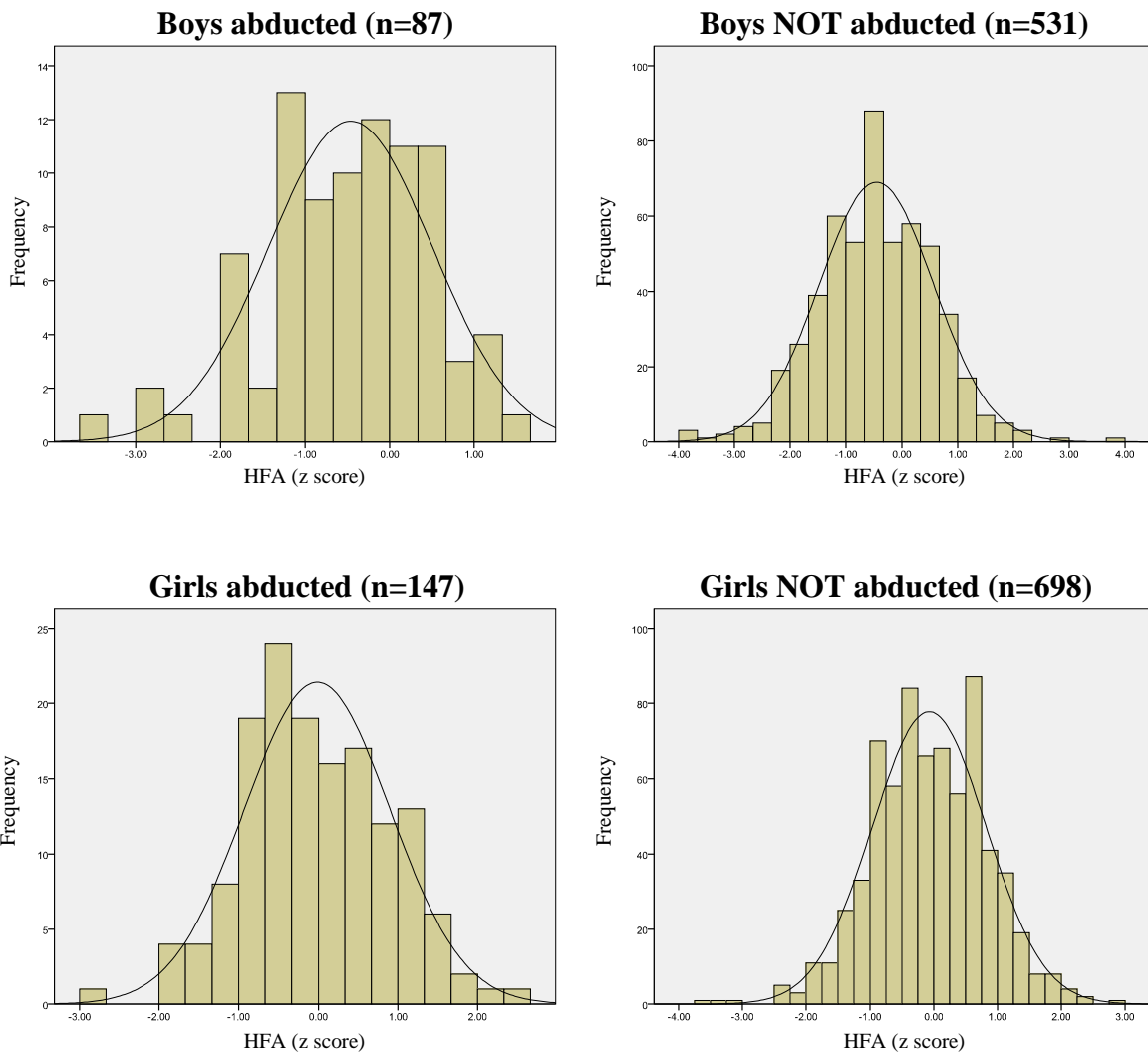
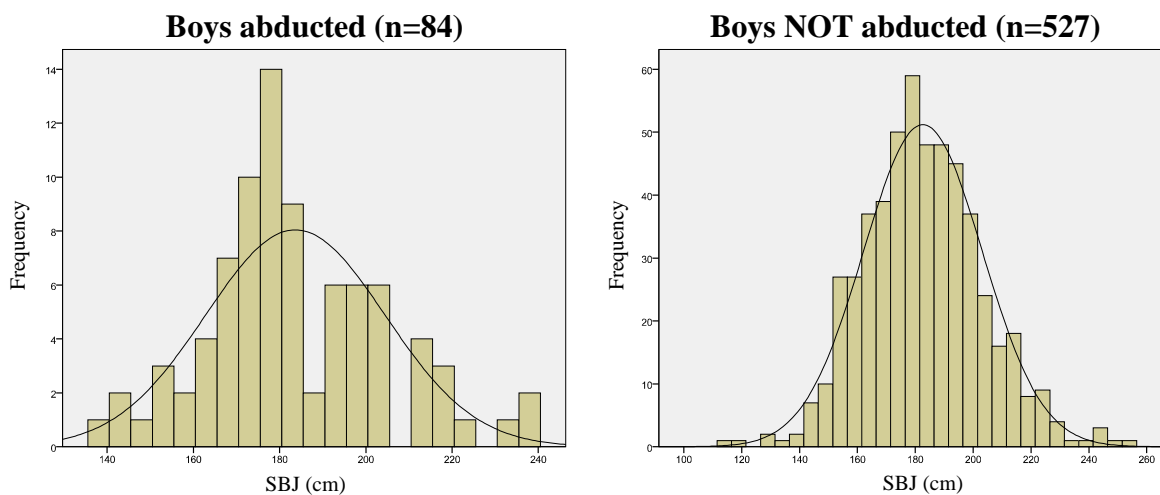


FIGURE A13.3: Histograms of the distribution of the crude SBJ results in the Gulu sample at M1 - stratified by gender and history of abduction



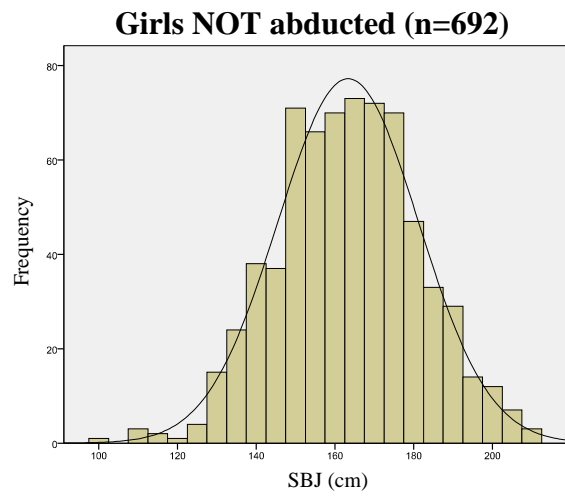
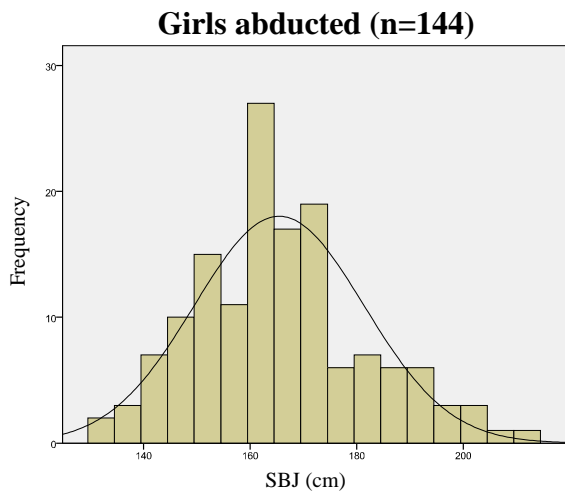


FIGURE A13.4: Histograms of the distribution of the crude MSFT results in the Gulu sample at M1 - stratified by gender and history of abduction

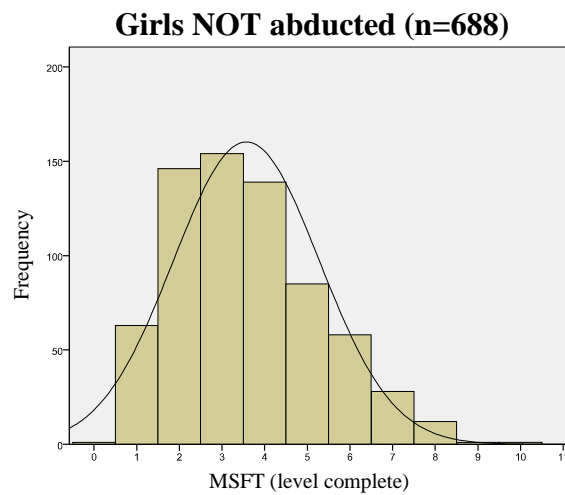
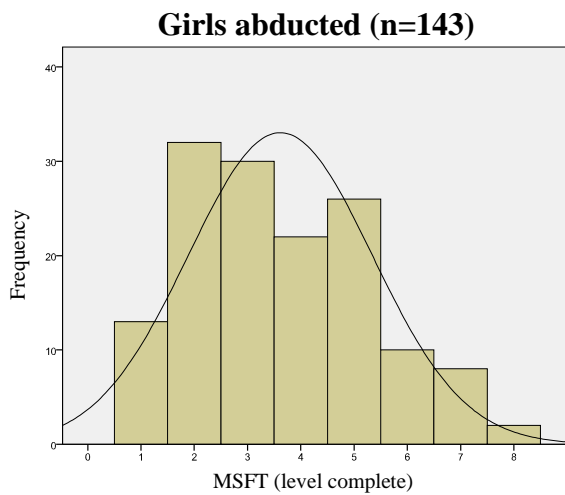
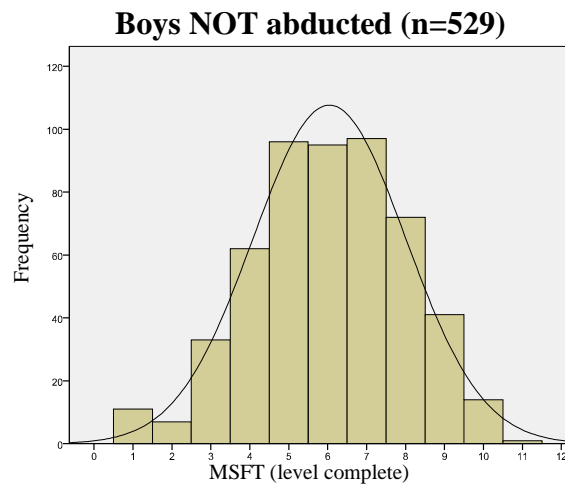
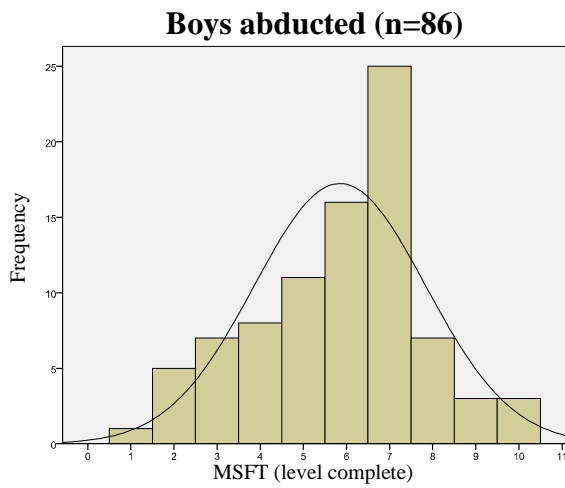


FIGURE A13.5: Histograms of the distribution of the crude TT results in the Gulu sample at M1 - stratified by gender and history of abduction

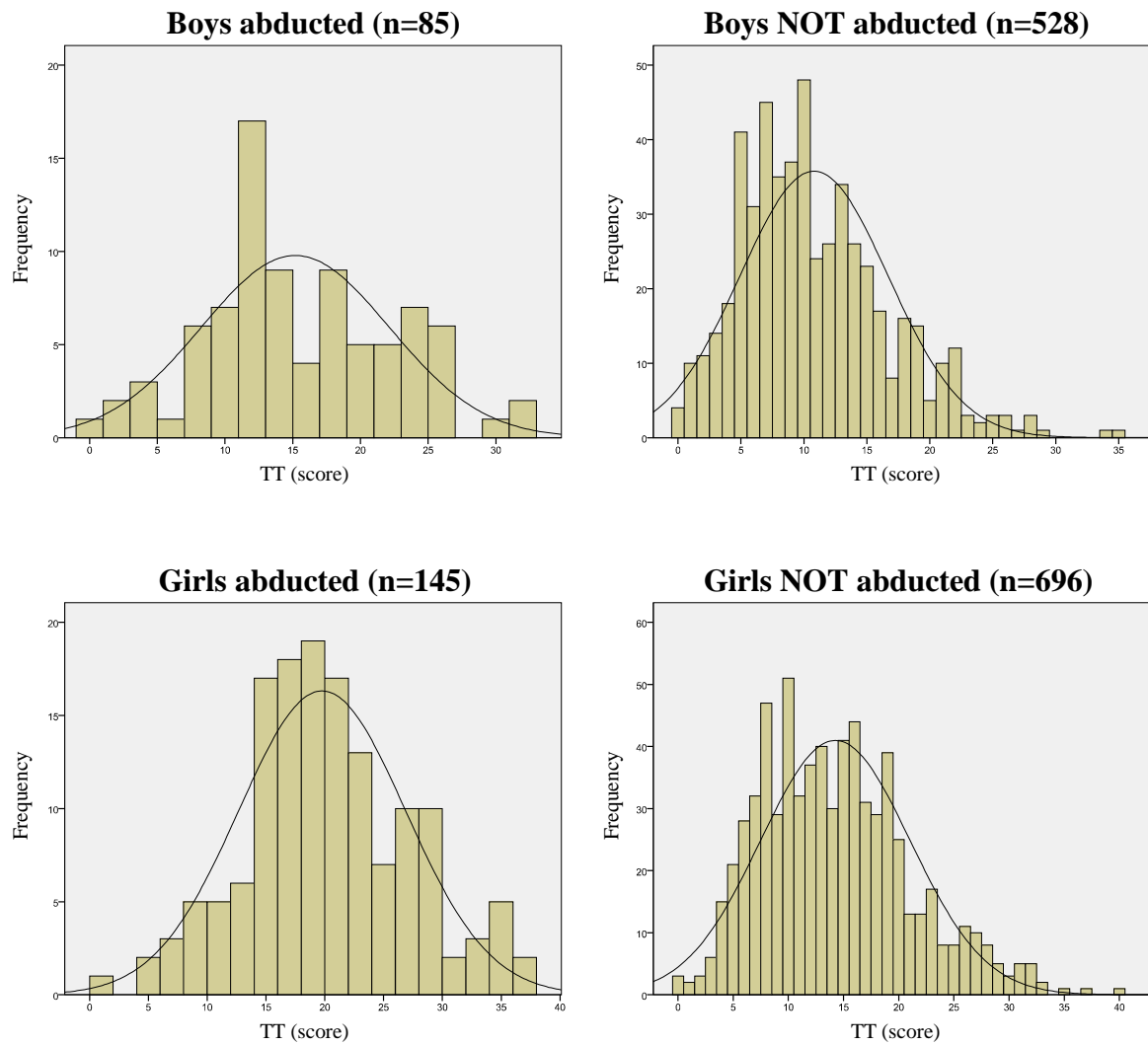
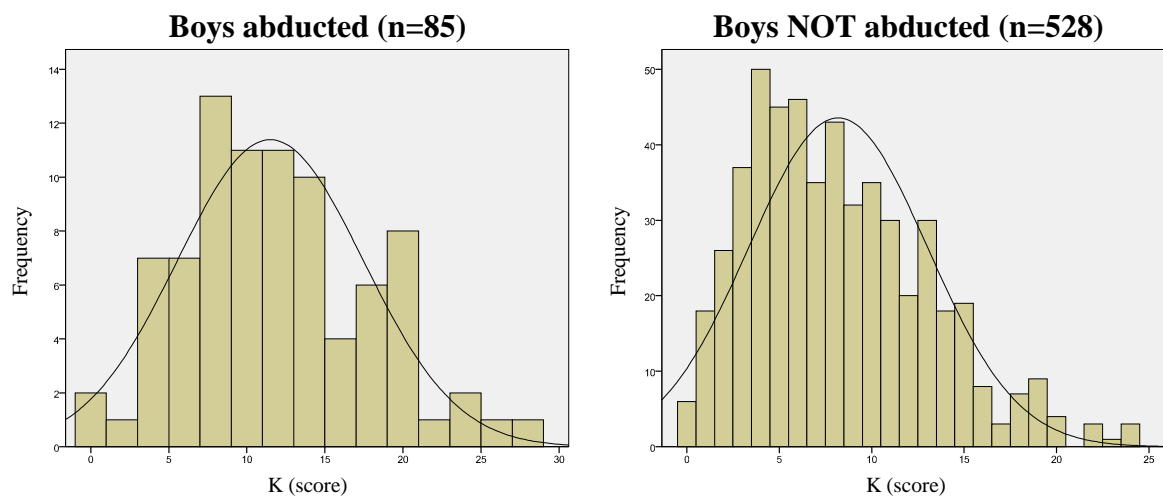


FIGURE A13.6: Histograms of the distribution of the crude K results in the Gulu sample at M1 - stratified by gender and history of abduction



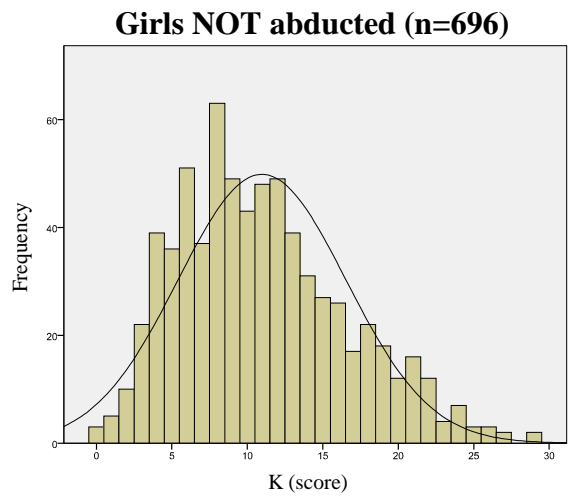
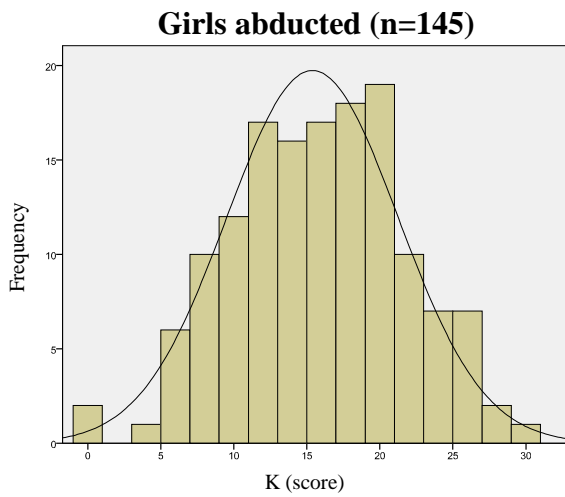


FIGURE A13.7: Histograms of the distribution of the crude P results in the Gulu sample at M1 - stratified by gender and history of abduction

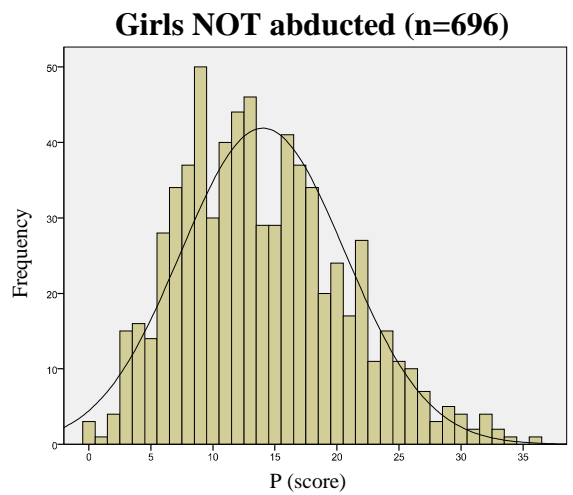
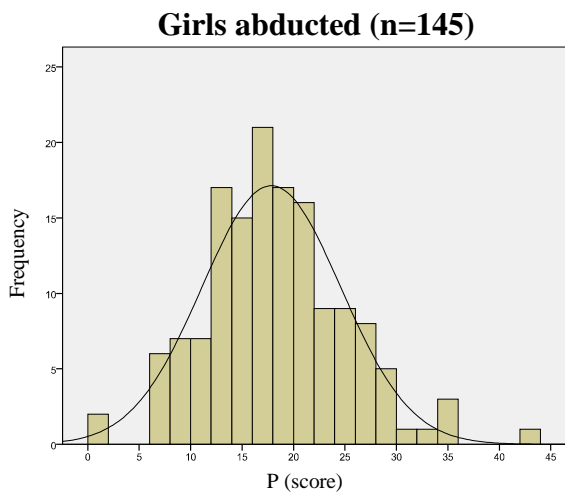
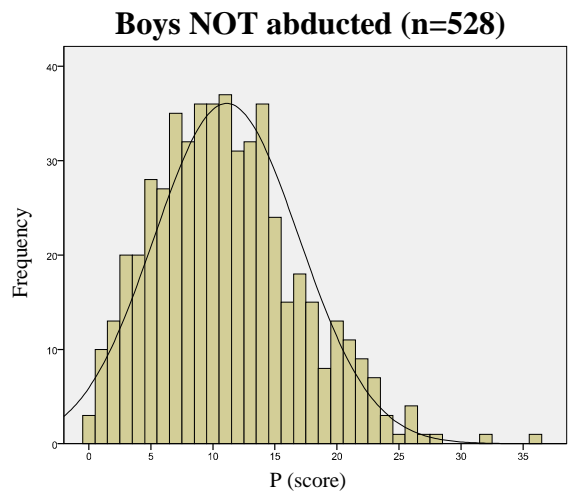
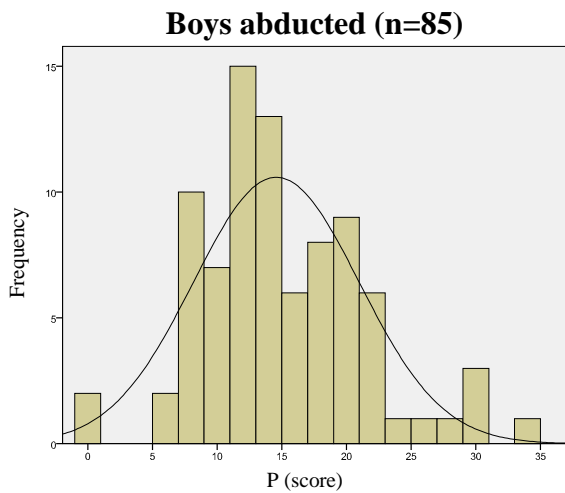


FIGURE A13.8: Histograms of the distribution of the crude ML results in the Gulu sample at M1 - stratified by gender and history of abduction

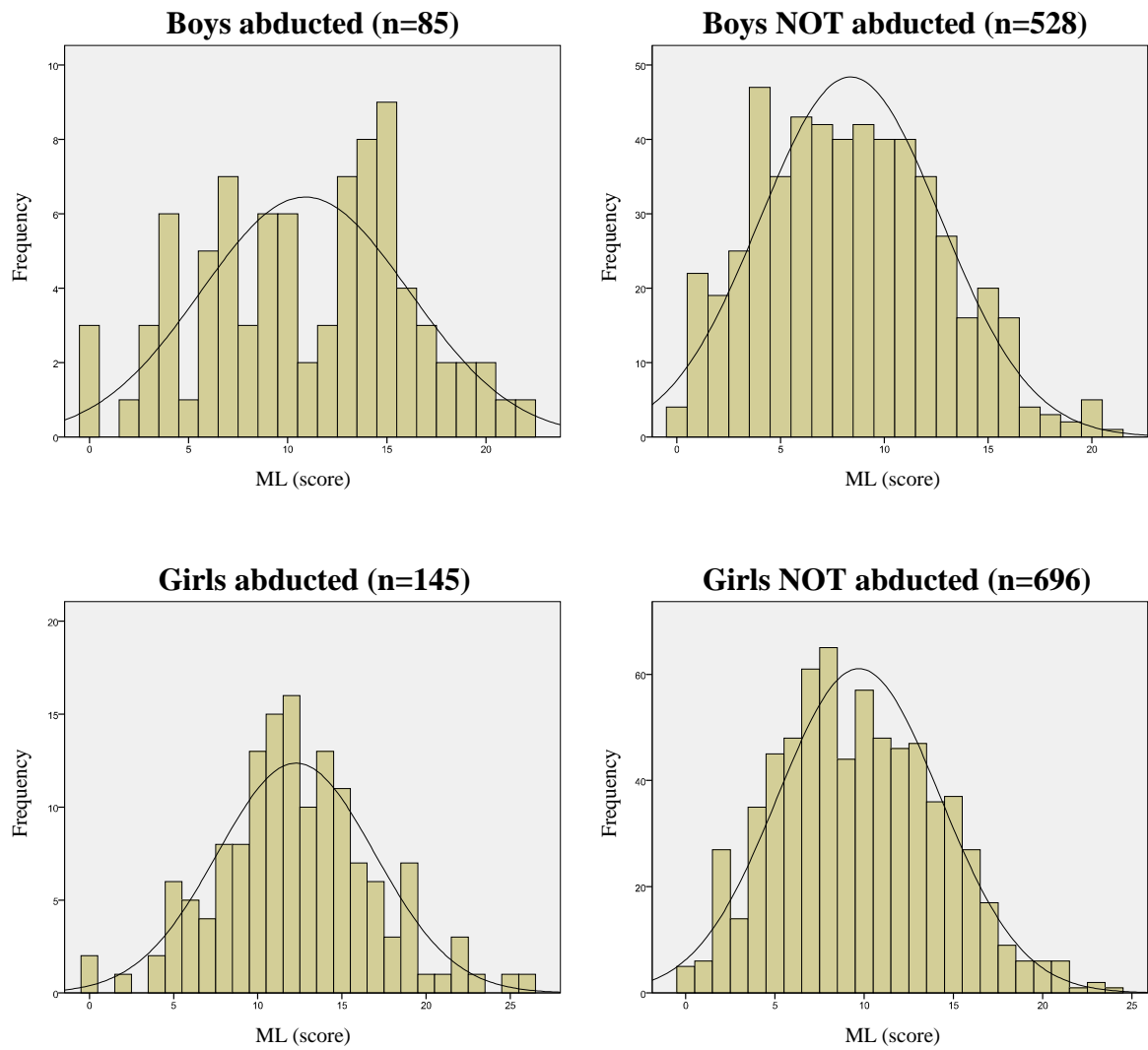
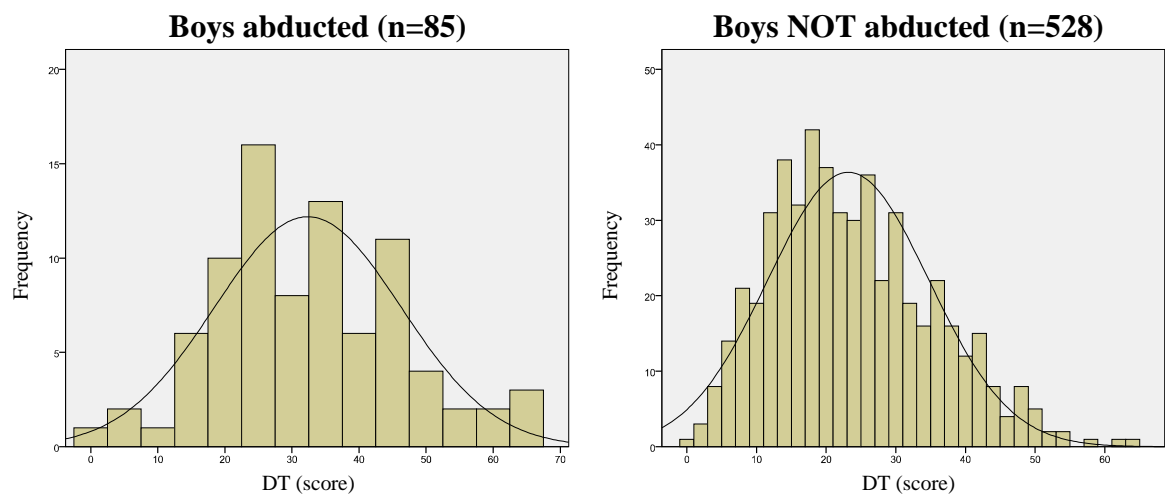


FIGURE A13.9: Histograms of the distribution of the crude DT results in the Gulu sample at M1 - stratified by gender and history of abduction



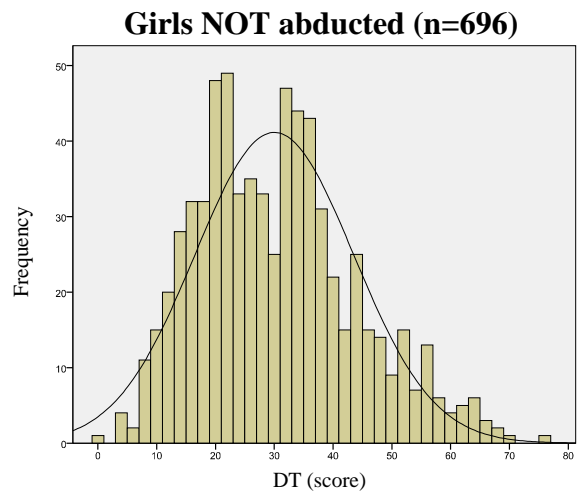
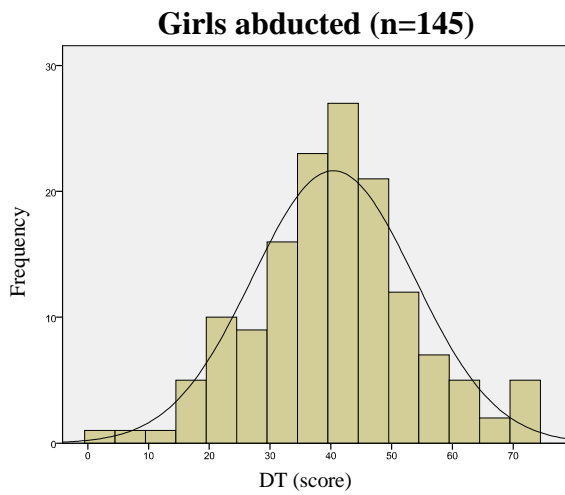
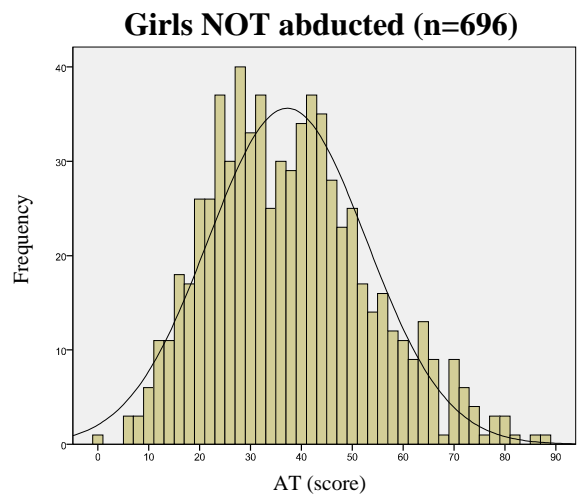
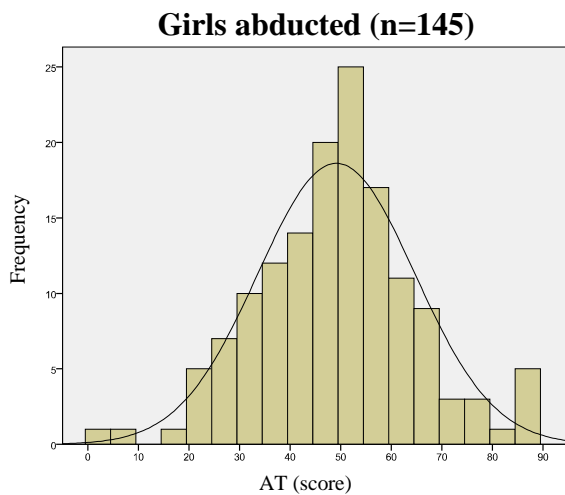
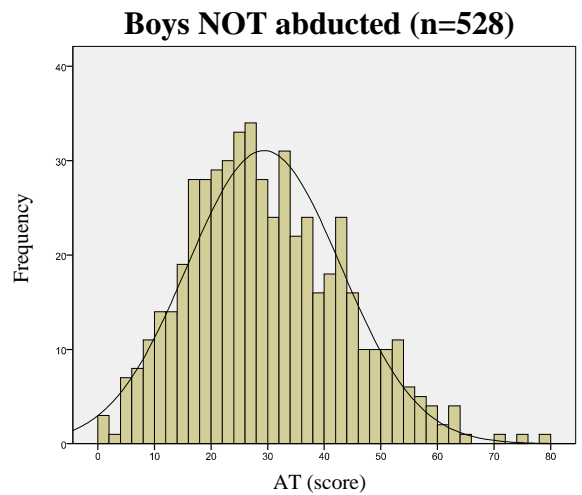
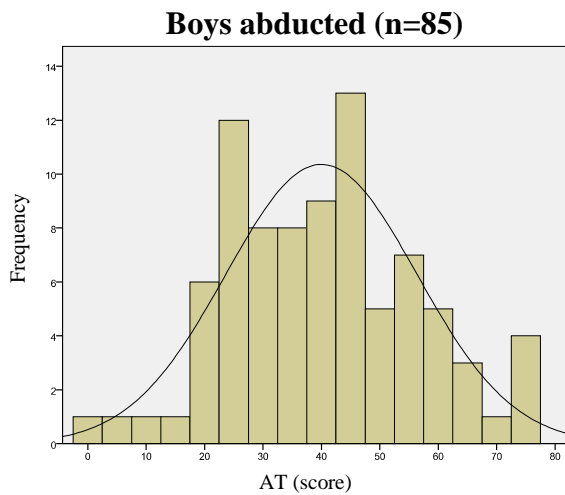


FIGURE A13.10: Histograms of the distribution of the crude AT results in the Gulu sample at M1 - stratified by gender and history of abduction



Inspection of the histograms warranted further assessment of normality and statistical tests for the crude results were also completed (TABLES A13.1 & A13.2).

TABLE A13.1: Statistical tests for normality of the crude results for all of the outcomes for the boys in the Gulu sample at M1 - stratified by history of abduction

Outcome variable	History of abduction	n	Skew		Kurtosis		Shapiro Wilk	
			Stat	SES	Stat	SEK	Stat	p
BFA (z score)	Yes	87	-0.346	0.258	0.179	0.511	0.986	0.444
	No	531	-0.340	0.106	0.985	0.212	0.988	<0.001
HFA (z score)	Yes	87	-0.451	0.258	0.258	0.511	0.983	0.327
	No	531	-0.075	0.106	0.717	0.212	0.993	0.020
SBJ (cm)	Yes	84	0.480	0.263	0.295	0.520	0.973	0.069
	No	527	0.236	0.106	0.433	0.212	0.994	0.029
MSFT (level complete)	Yes	86	-0.308	0.260	-0.225	0.514	0.953	0.003
	No	529	-0.194	0.106	-0.255	0.212	0.971	<0.001
TT (score)	Yes	85	0.201	0.261	-0.397	0.517	0.981	0.234
	No	528	0.790	0.106	0.712	0.212	0.960	<0.001
K (score)	Yes	85	0.418	0.261	-0.227	0.517	0.979	0.169
	No	528	0.697	0.106	0.154	0.212	0.957	<0.001
P (score)	Yes	85	0.547	0.261	0.617	0.517	0.969	0.040
	No	528	0.573	0.106	0.349	0.212	0.975	<0.001
ML (score)	Yes	85	-0.119	0.261	-0.718	0.517	0.977	0.141
	No	528	0.293	0.106	-0.451	0.212	0.980	<0.001
DT (score)	Yes	85	0.330	0.261	-0.026	0.517	0.983	0.312
	No	528	0.524	0.106	-0.095	0.212	0.977	<0.001
AT (score)	Yes	85	0.135	0.261	-0.235	0.517	0.987	0.561
	No	528	0.434	0.106	-0.063	0.212	0.985	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A13.2: Statistical tests for normality of the crude results for all of the outcomes for the girls in the Gulu sample at M1 - stratified by history of abduction

Outcome variable	History of abduction	n	Skew		Kurtosis		Shapiro Wilk	
			Stat	SES	Stat	SEK	Stat	p
BFA (z score)	Yes	147	0.108	0.200	-0.392	0.397	0.993	0.677
	No	698	-0.478	0.093	0.518	0.185	0.987	<0.001
HFA (z score)	Yes	147	0.114	0.200	-0.054	0.397	0.990	0.344
	No	698	-0.140	0.093	0.421	0.185	0.995	0.015
SBJ (cm)	Yes	144	0.457	0.202	0.051	0.401	0.980	0.033
	No	692	-0.061	0.093	0.001	0.186	0.998	0.395
MSFT (level complete)	Yes	143	0.430	0.203	-0.559	0.403	0.938	<0.001
	No	688	0.573	0.093	-0.071	0.186	0.942	<0.001
TT (score)	Yes	145	0.128	0.201	-0.114	0.400	0.988	0.228
	No	696	0.600	0.093	0.162	0.185	0.973	<0.001
K (score)	Yes	145	-0.052	0.201	-0.345	0.400	0.991	0.518
	No	696	0.584	0.093	-0.114	0.185	0.968	<0.001
P (score)	Yes	145	0.438	0.201	0.798	0.400	0.983	0.075
	No	696	0.486	0.093	-0.125	0.185	0.979	<0.001
ML (score)	Yes	145	0.179	0.201	0.434	0.400	0.989	0.279
	No	696	0.284	0.093	-0.322	0.185	0.986	<0.001
DT (score)	Yes	145	0.022	0.201	0.253	0.400	0.990	0.418
	No	696	0.532	0.093	-0.099	0.185	0.976	<0.001
AT (score)	Yes	145	-0.024	0.201	0.477	0.400	0.988	0.248
	No	696	0.455	0.093	-0.142	0.185	0.983	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12].

The assumption of HOV between abducted vs. NOT abducted for the crude results of each of the outcome variables was assessed using the Levene and Brown-Forsythe tests (TABLE A13.3). These statistical tests for HOV were robust to violations of the normality assumption

^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$).

TABLE A13.3: Statistical tests of the HOV between abducted and NOT abducted adolescents for all of the crude outcomes in the Gulu sample at M1 - stratified by gender

Outcome variable	Gender	Abducted		NOT abducted		Levene test		Brown-Forsythe test		Summary
		n	Variance	n	Variance	F	p	F	p	
BFA (z score)	Boys	87	0.943	531	0.778	2.352	0.126	1.776	0.183	√
	Girls	147	0.755	698	0.821	0.067	0.796	0.052	0.820	√
HFA (z score)	Boys	87	0.939	531	1.047	0.178	0.673	0.263	0.608	√
	Girls	147	0.835	698	0.801	0.489	0.485	0.355	0.551	√
SBJ (cm)	Boys	84	433.80	527	421.64	0.013	0.909	0.020	0.888	√
	Girls	144	253.86	692	319.48	3.899	0.049	4.203	0.041	√
MSFT (level complete)	Boys	86	3.965	529	3.845	0.013	0.911	0.015	0.903	√
	Girls	143	2.985	688	2.935	0.351	0.554	0.270	0.603	√
TT (score)	Boys	85	48.11	528	34.68	6.039	0.014	5.305	0.022	√
	Girls	145	50.27	696	45.93	0.105	0.746	0.076	0.782	√
K (score)	Boys	85	35.47	528	23.39	6.551	0.011	6.326	0.012	√
	Girls	145	34.35	696	31.04	0.968	0.326	1.028	0.311	√
P (score)	Boys	85	41.03	528	34.12	0.858	0.355	0.494	0.482	√
	Girls	145	45.62	696	43.94	0.269	0.604	0.207	0.649	√
ML (score)	Boys	85	27.65	528	18.94	8.646	0.003	8.602	0.003	L
	Girls	145	21.88	696	20.68	0.280	0.597	0.371	0.543	√
DT (score)	Boys	85	193.52	528	134.23	4.323	0.038	4.202	0.041	√
	Girls	145	178.51	696	182.12	0.872	0.351	0.842	0.359	√
AT (score)	Boys	85	267.88	528	183.94	5.240	0.022	5.382	0.021	√
	Girls	145	241.49	696	243.12	0.751	0.387	0.708	0.400	√

NOTE: Results in **BOLD ITALICS** suggest a statistically significant heterogeneity of variance ($p < 0.01$).
 NOTE: In the summary column - √ denotes satisfactory HOV; L indicates a "liberal" statistical test ^[12,14].

With the exception of ML in the boys sample, the assumption of HOV held for all of the dependent variables when comparing the outcomes for those abducted vs. NOT abducted. The statistical analysis could proceed with little concern when the HOV assumption was satisfied ^[12,14].

The violation of the HOV assumption for ML in the boys sample was offset by the large sample size ^[12,14]. However, when comparing abducted vs. NOT abducted participants, the unbalanced sample sizes may have significantly compromised the outcomes of the statistical tests. In this case, the group with the largest n had a smaller variance. The analysis was referred to as a “liberal” statistical test and the risk of a type I error was increased ^[12,14].

A consistent statistical approach was applied for all outcome variables when comparing abducted vs. NOT abducted. The interpretation of any significant statistical findings was treated with caution for the “liberal” statistical test of ML in the boys sample ^[12,14].

APPENDIX A14: Relationship between Outcome Variables at M1 (Plots)

The crude results for the PF variables were plotted against each of the MH outcomes from the Gulu sample at M1 (FIGURES A14.1 – A14.24).

FIGURE A14.1: Scatter plot of crude results for BFA vs. TT in the Gulu sample at M1

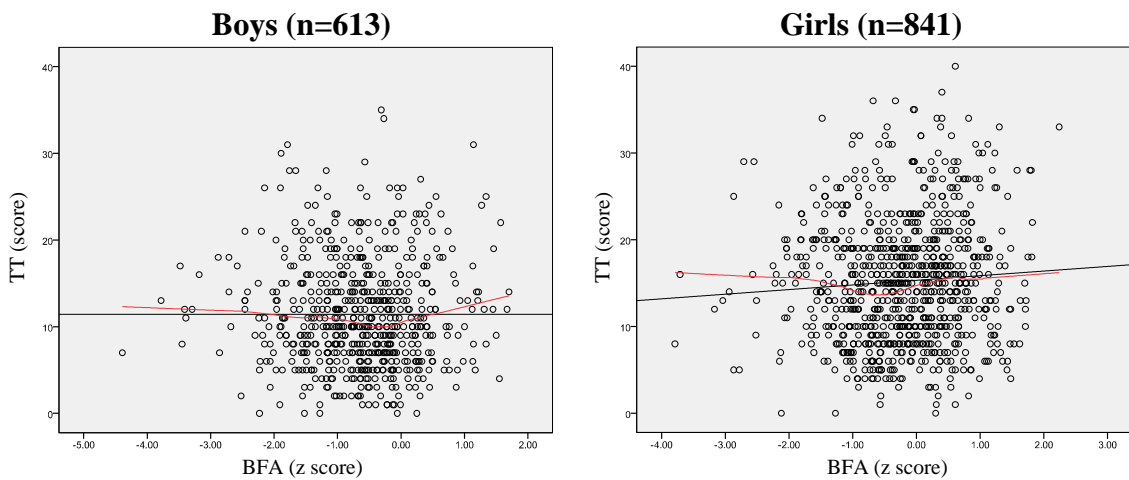


FIGURE A14.2: Scatter plot of crude results for BFA vs. K in the Gulu sample at M1

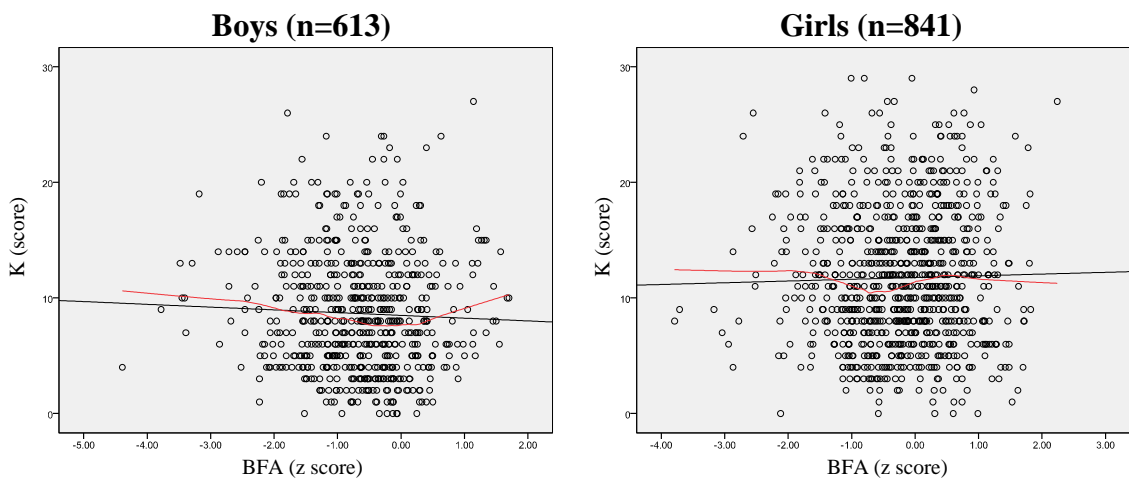


FIGURE A14.3: Scatter plot of crude results for BFA vs. P in the Gulu sample at M1

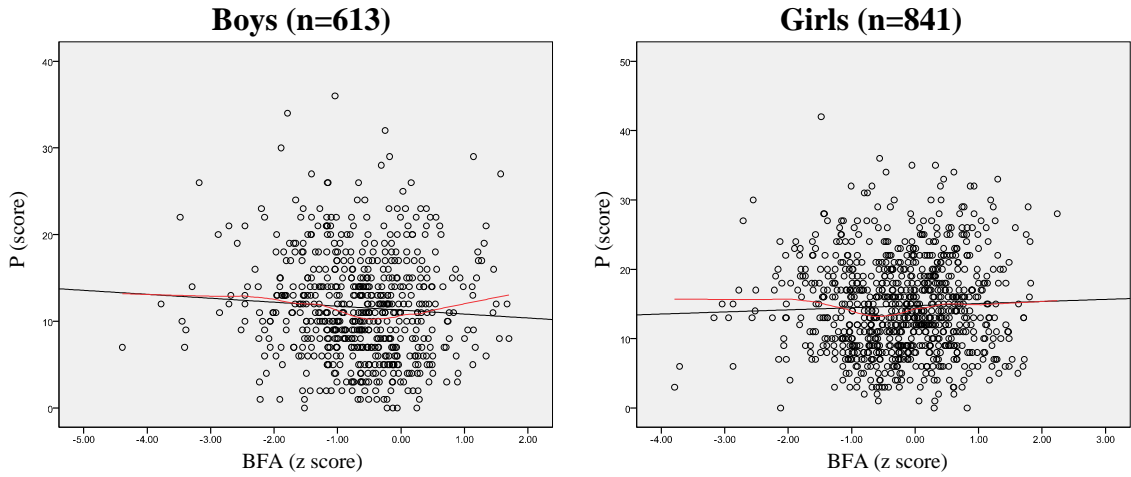


FIGURE A14.4: Scatter plot of crude results for BFA vs. ML in the Gulu sample at M1

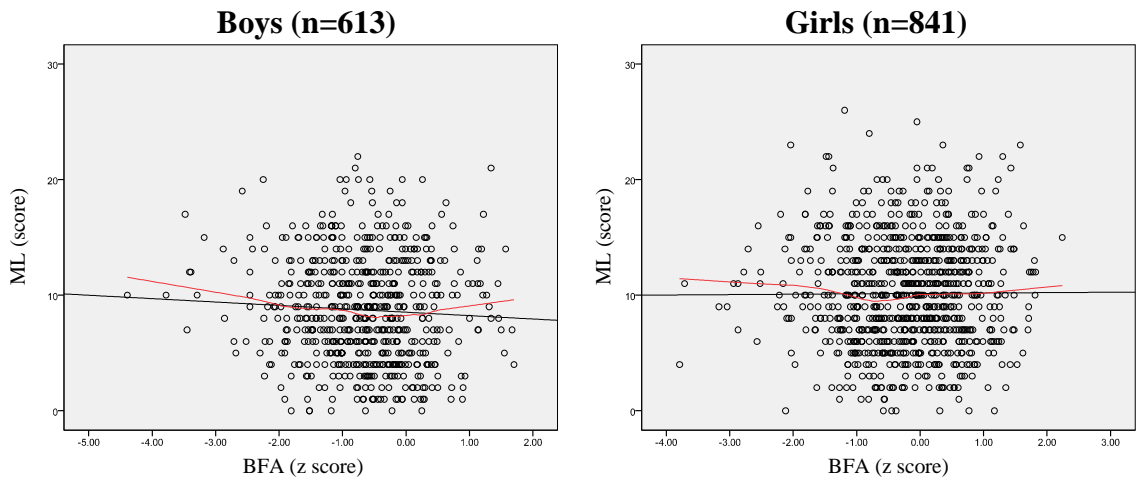


FIGURE A14.5: Scatter plot of crude results for BFA vs. DT in the Gulu sample at M1

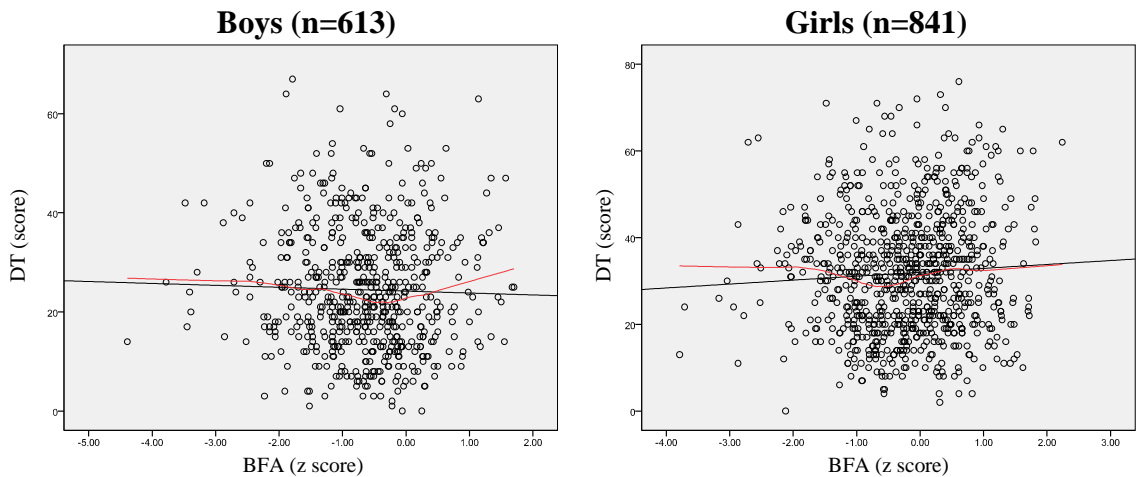


FIGURE A14.6: Scatter plot of crude results for BFA vs. AT in the Gulu sample at M1

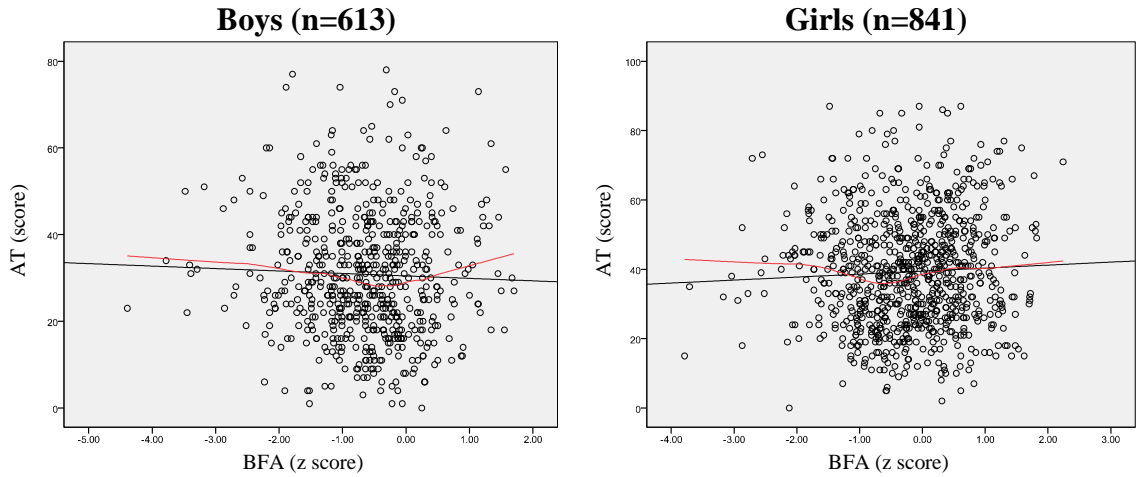


FIGURE A14.7: Scatter plot of crude results for HFA vs. TT in the Gulu sample at M1

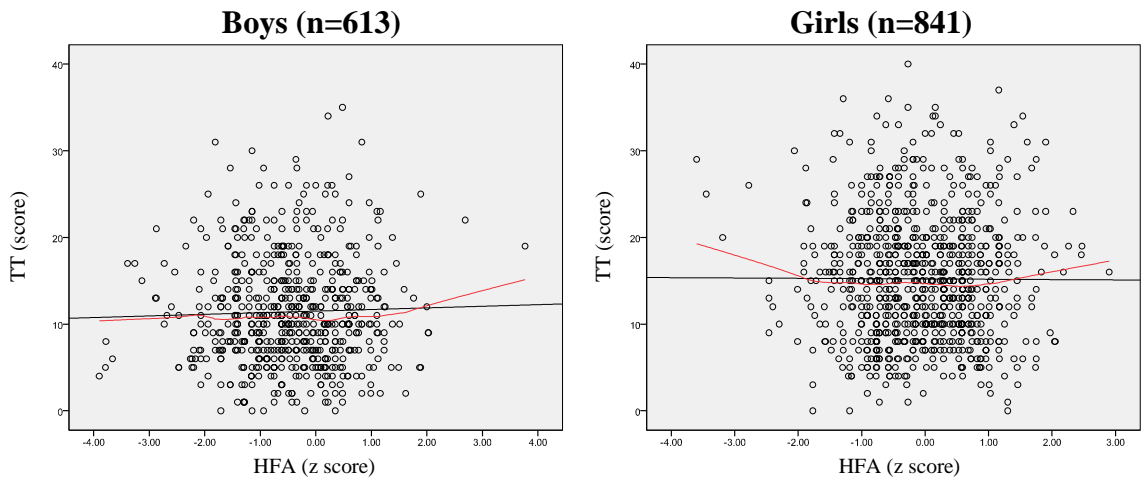


FIGURE A14.8: Scatter plot of crude results for HFA vs. K in the Gulu sample at M1

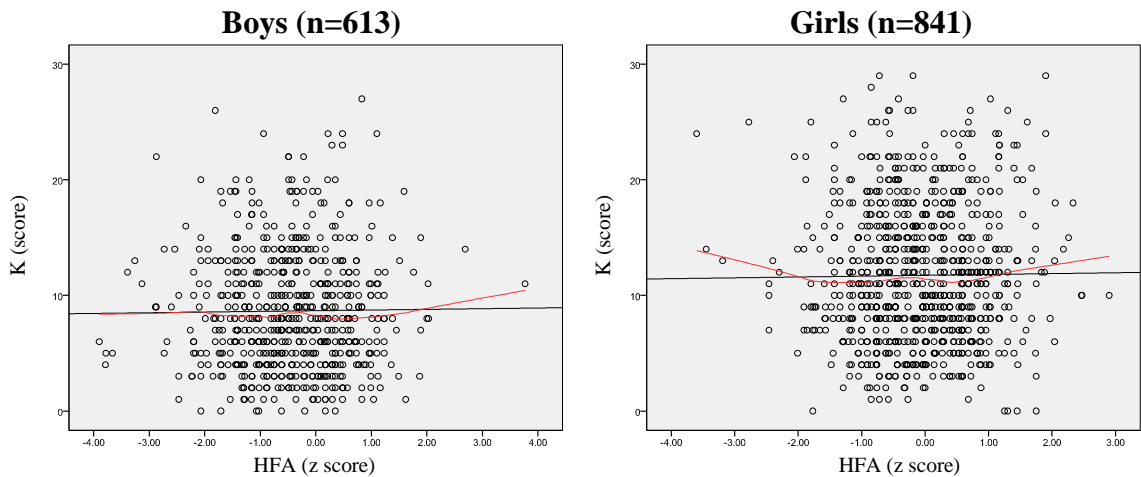


FIGURE A14.9: Scatter plot of crude results for HFA vs. P in the Gulu sample at M1

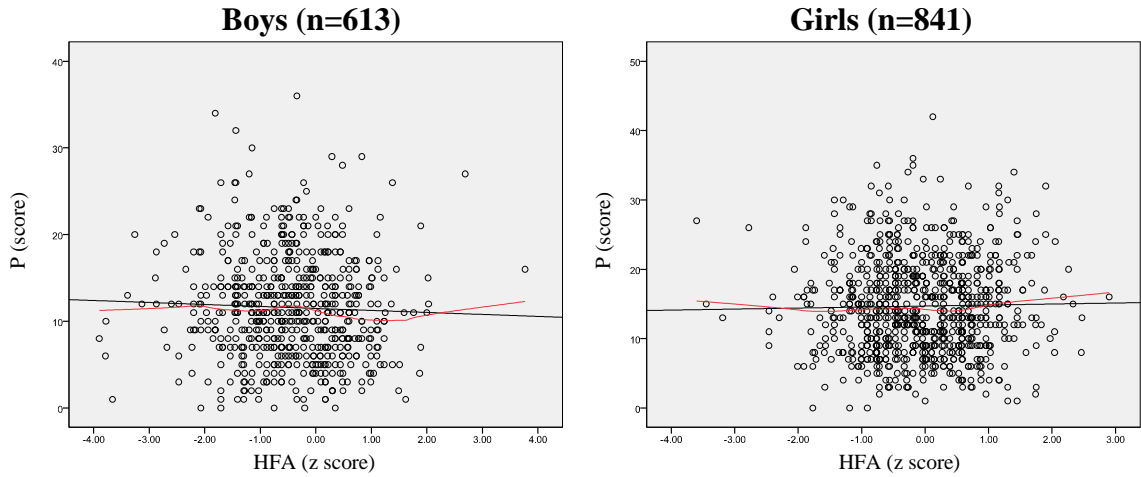


FIGURE A14.10: Scatter plot of crude results for HFA vs. ML in the Gulu sample at M1

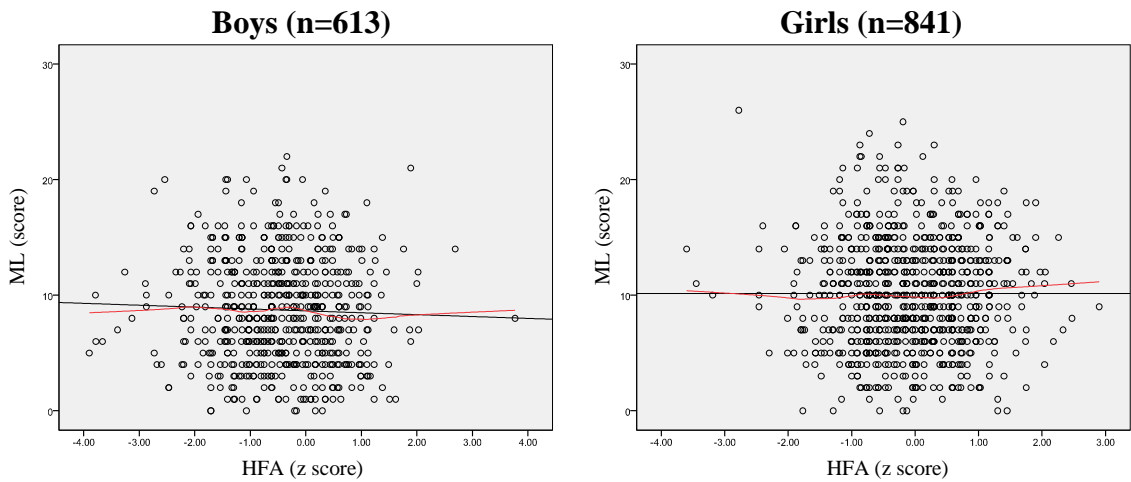


FIGURE A14.11: Scatter plot of crude results for HFA vs. DT in the Gulu sample at M1

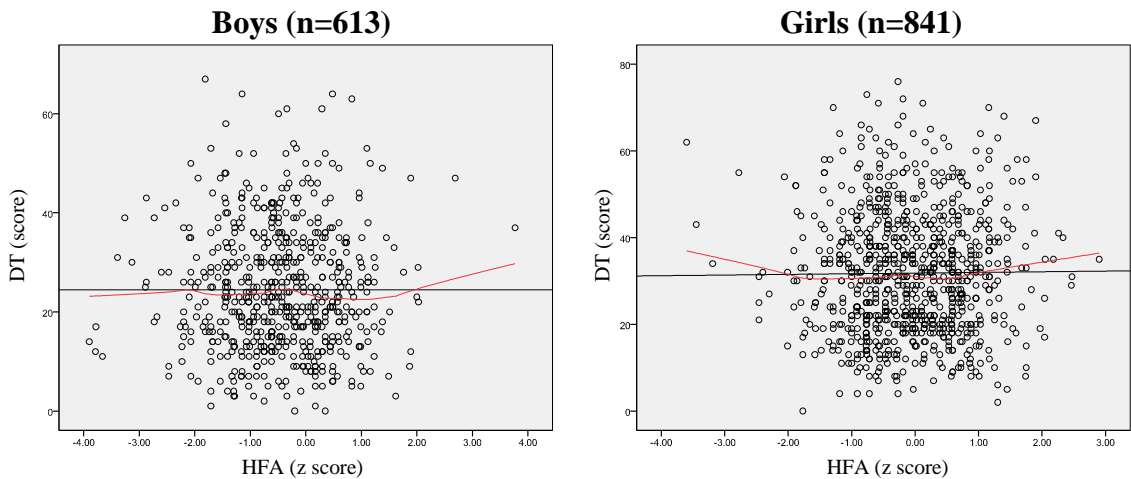


FIGURE A14.12: Scatter plot of crude results for HFA vs. AT in the Gulu sample at M1

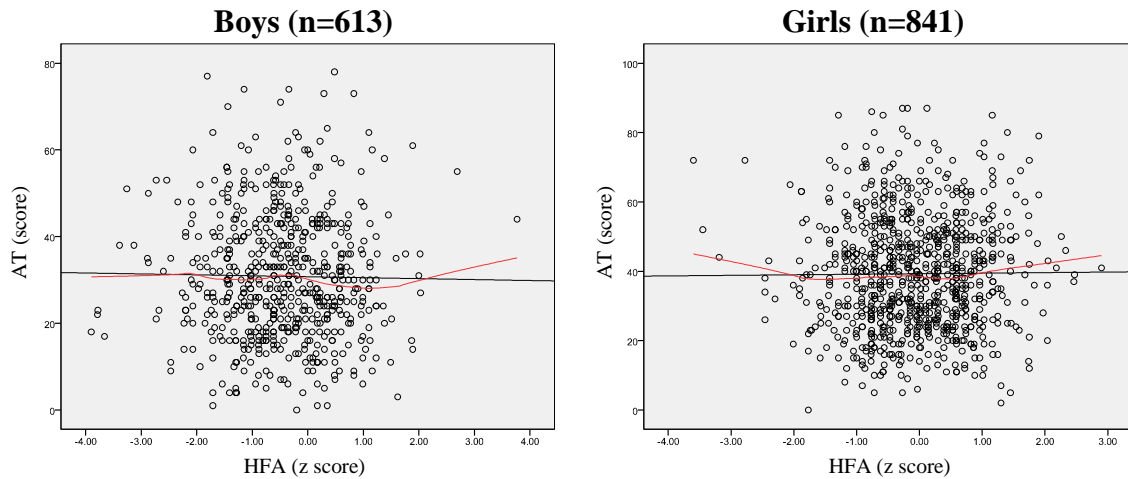


FIGURE A14.13: Scatter plot of crude results for SBJ vs. TT in the Gulu sample at M1

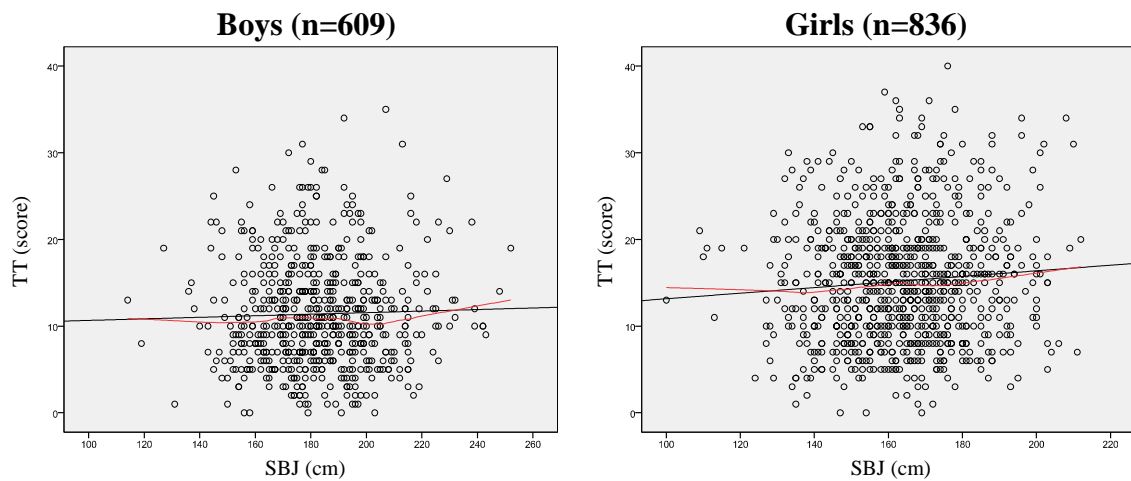


FIGURE A14.14: Scatter plot of crude results for SBJ vs. K in the Gulu sample at M1

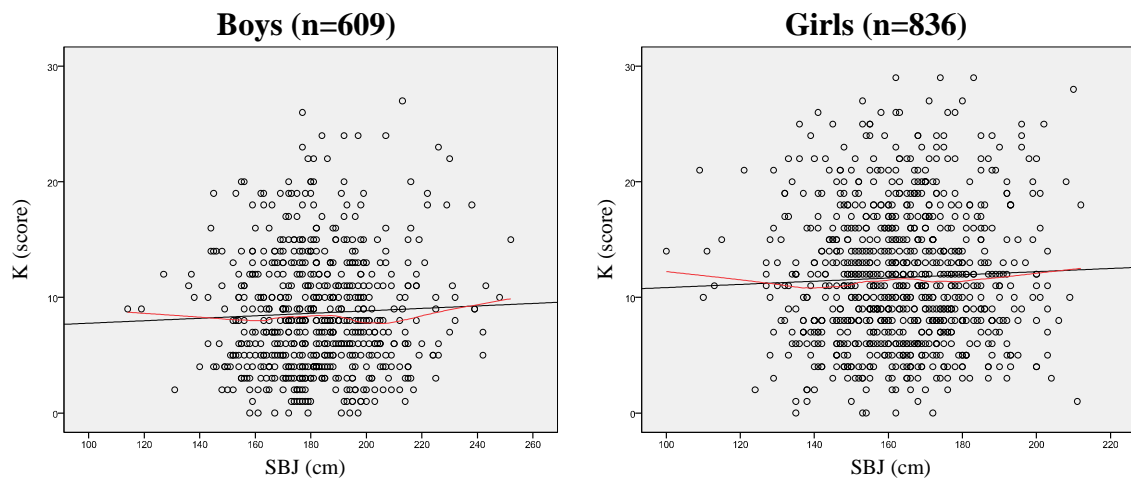


FIGURE A14.15: Scatter plot of crude results for SBJ vs. P in the Gulu sample at M1

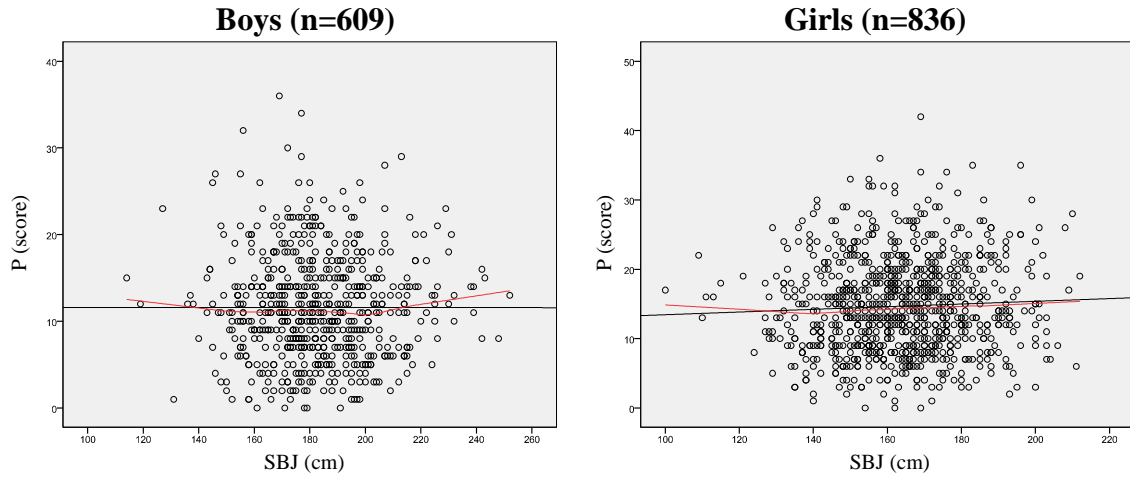


FIGURE A14.16: Scatter plot of crude results for SBJ vs. ML in the Gulu sample at M1

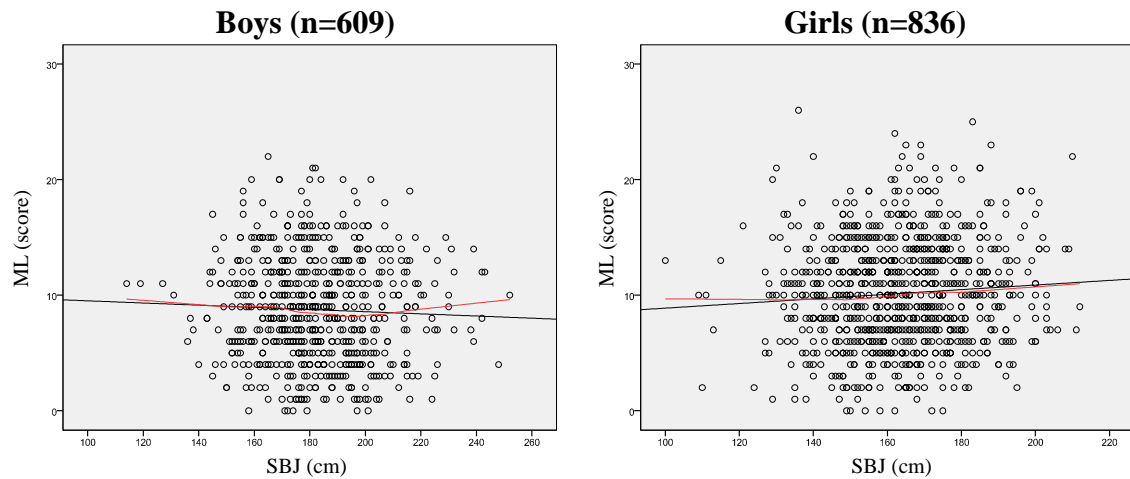


FIGURE A14.17: Scatter plot of crude results for SBJ vs. DT in the Gulu sample at M1

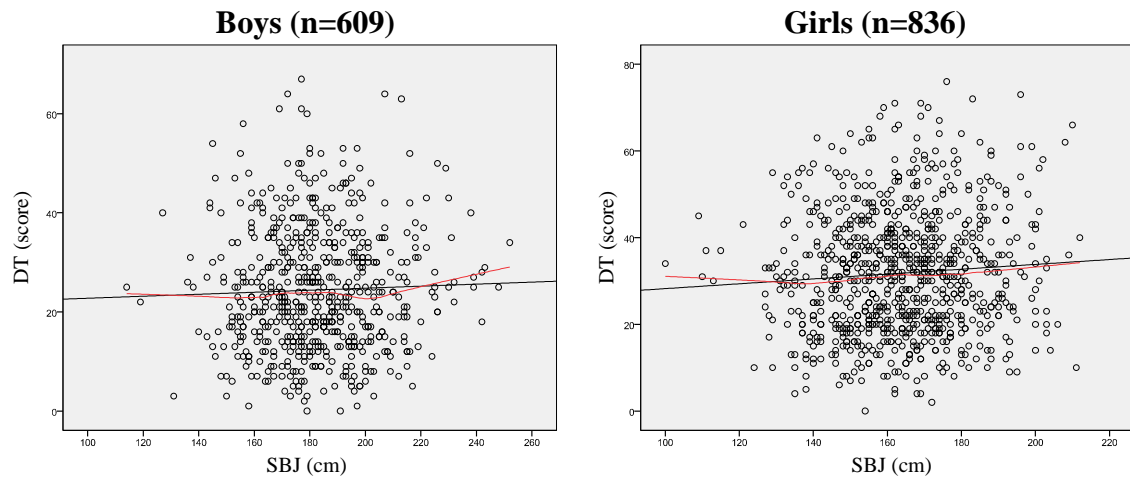


FIGURE A14.18: Scatter plot of crude results for SBJ vs. AT in the Gulu sample at M1

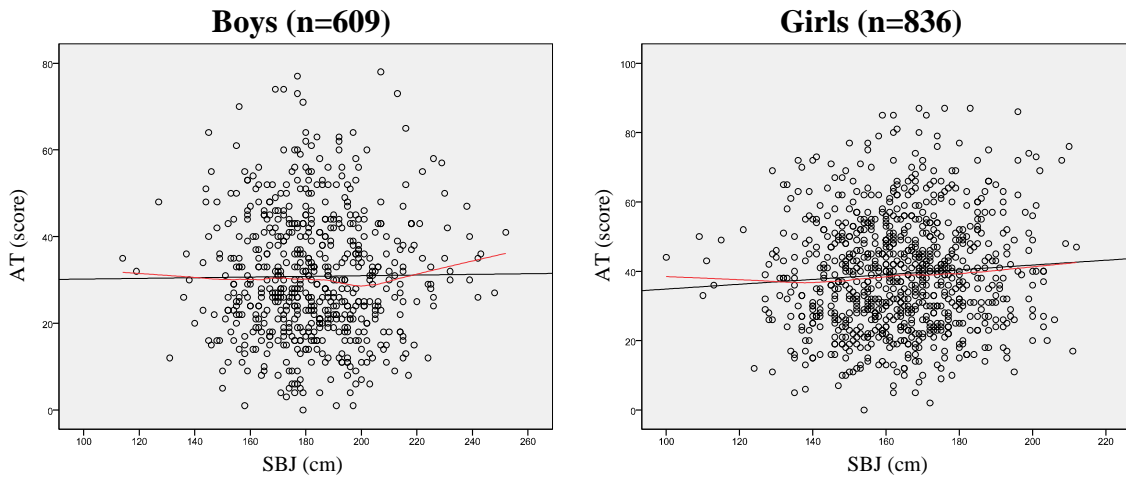


FIGURE A14.19: Scatter plot of crude results for MSFT vs. TT in the Gulu sample at M1

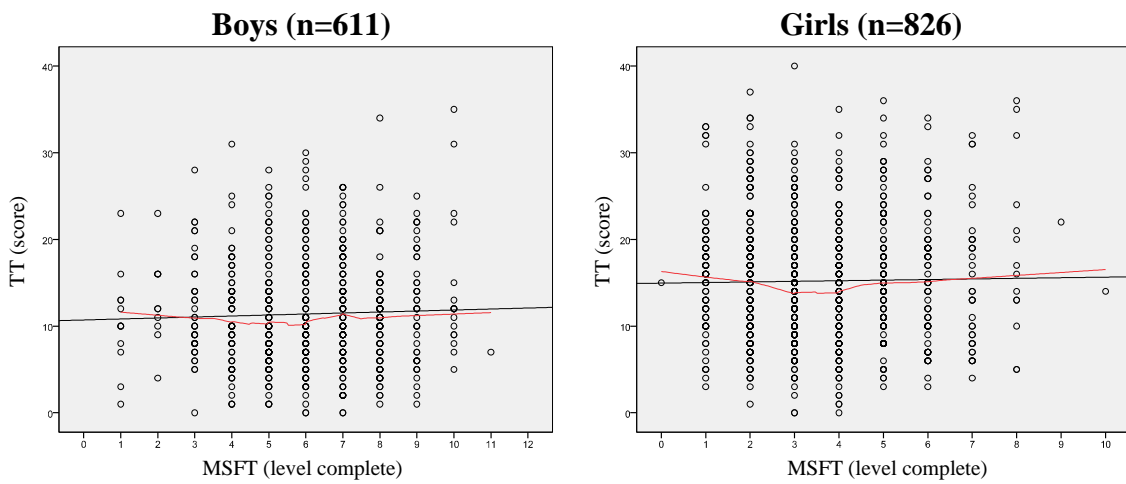


FIGURE A14.20: Scatter plot of crude results for MSFT vs. K in the Gulu sample at M1

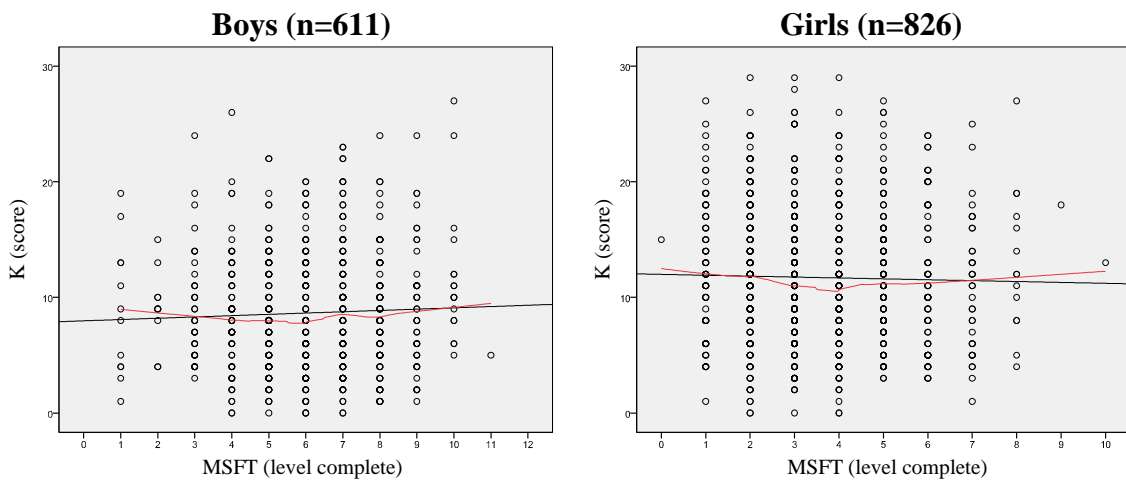


FIGURE A14.21: Scatter plot of crude results for MSFT vs. P in the Gulu sample at M1

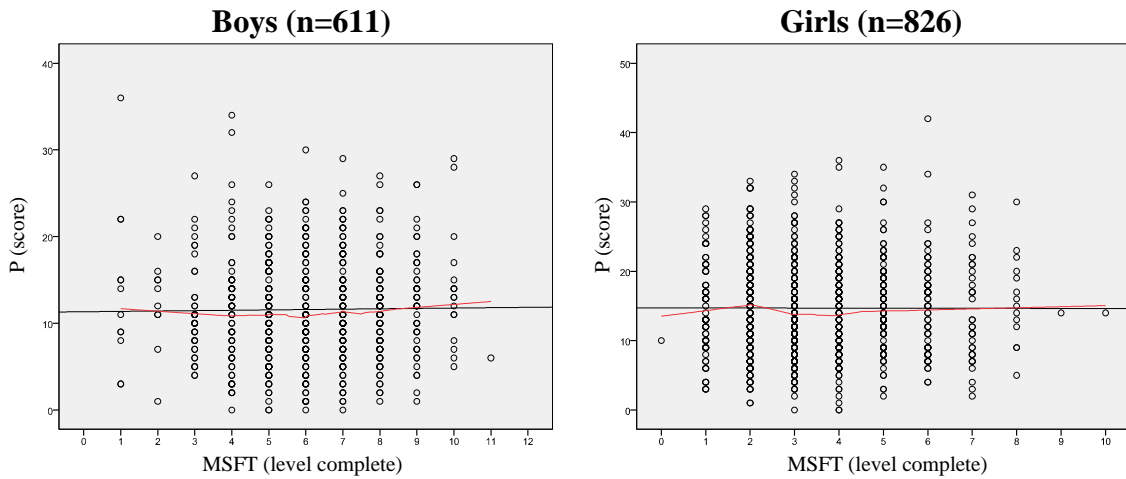


FIGURE A14.22: Scatter plot of crude results for MSFT vs. ML in the Gulu sample at M1

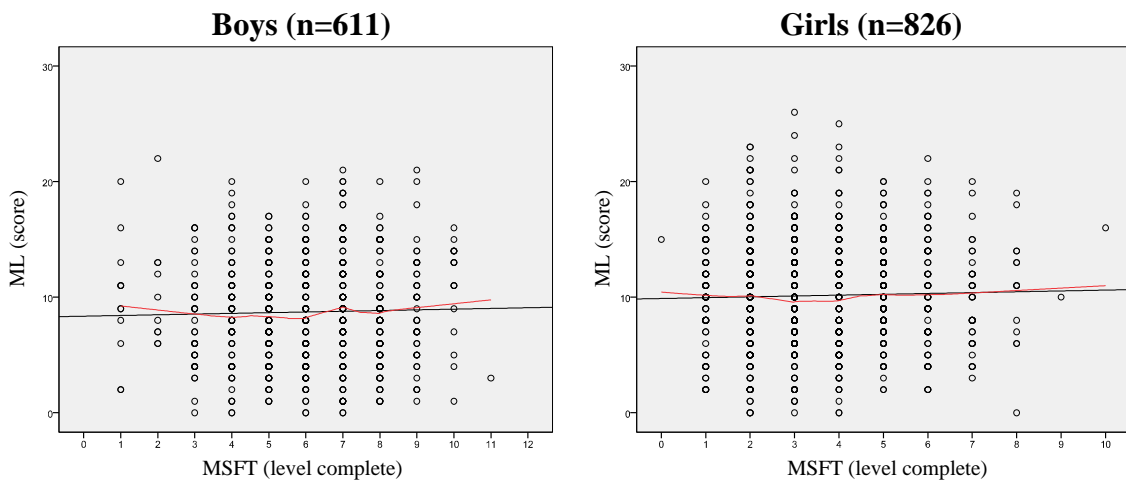


FIGURE A14.23: Scatter plot of crude results for MSFT vs. DT in the Gulu sample at M1

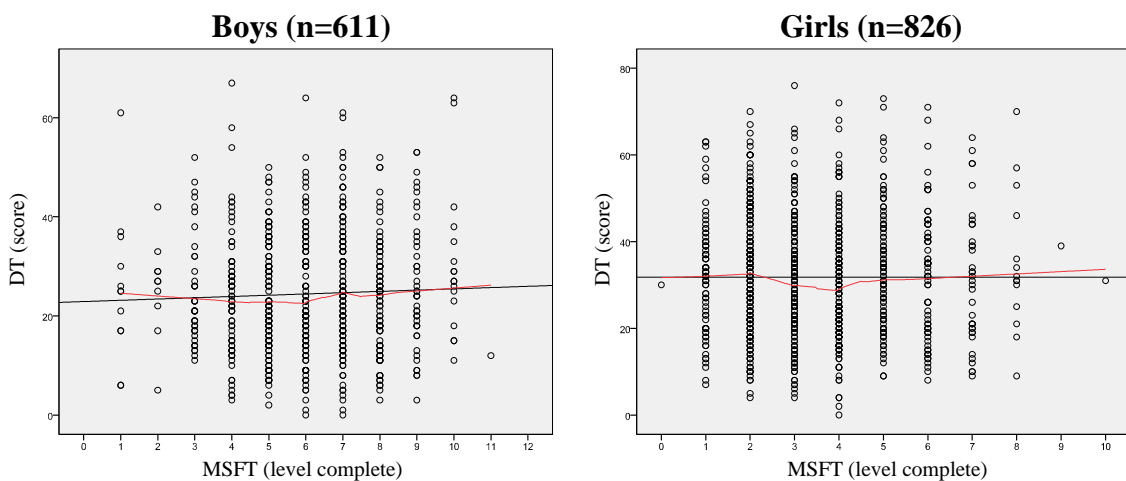
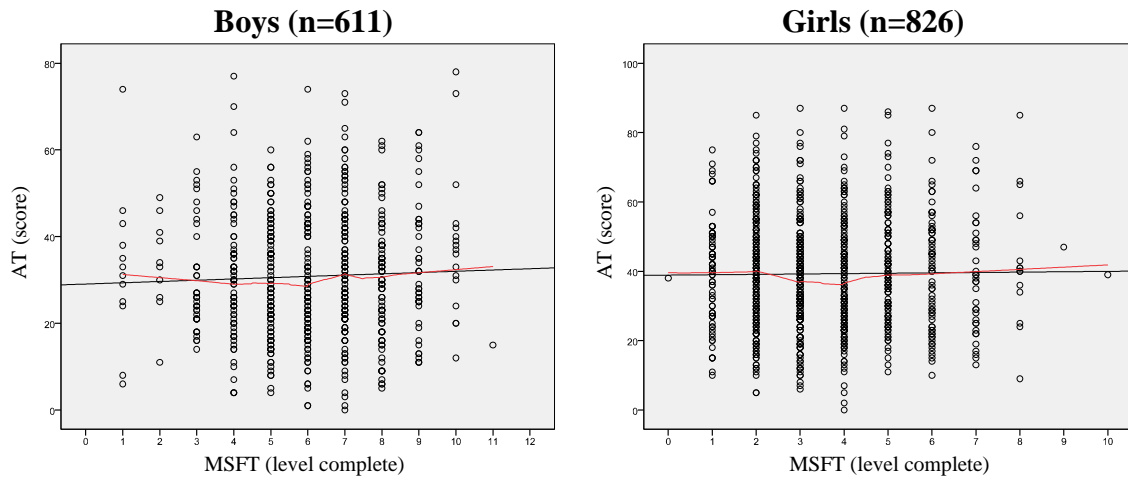


FIGURE A14.24: Scatter plot of crude results for MSFT vs. AT in the Gulu sample at M1



The crude results of the PF outcomes indicating body composition (BFA, HFA) were plotted against the indicators of muscle strength (SBJ) and cardiorespiratory fitness (MSFT) in the Gulu sample at M1 (FIGURES A14.25 – A14.28).

FIGURE A14.25: Scatter plot of crude results for BFA vs. SBJ in the Gulu sample at M1

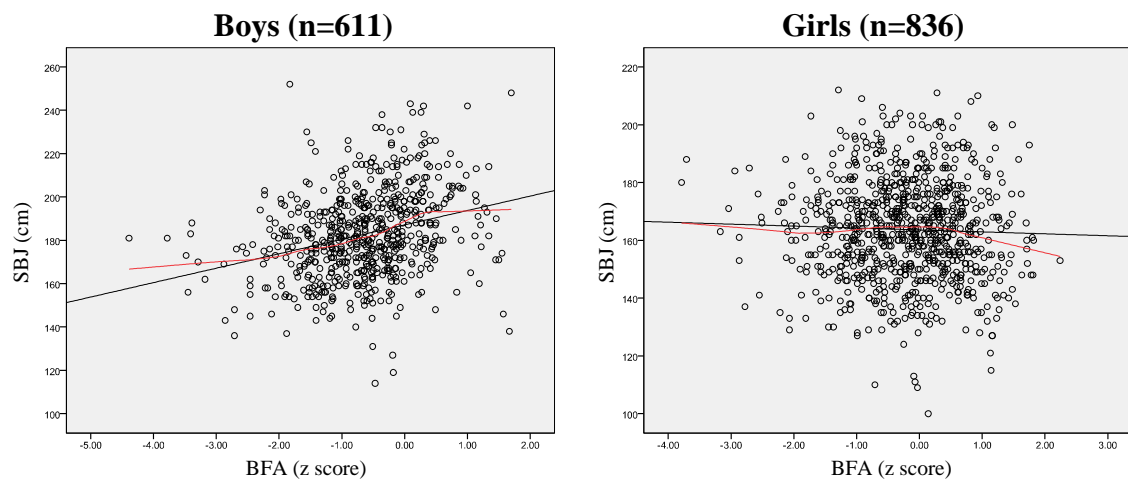


FIGURE A14.26: Scatter plot of crude results for BFA vs. MSFT in the Gulu sample at M1

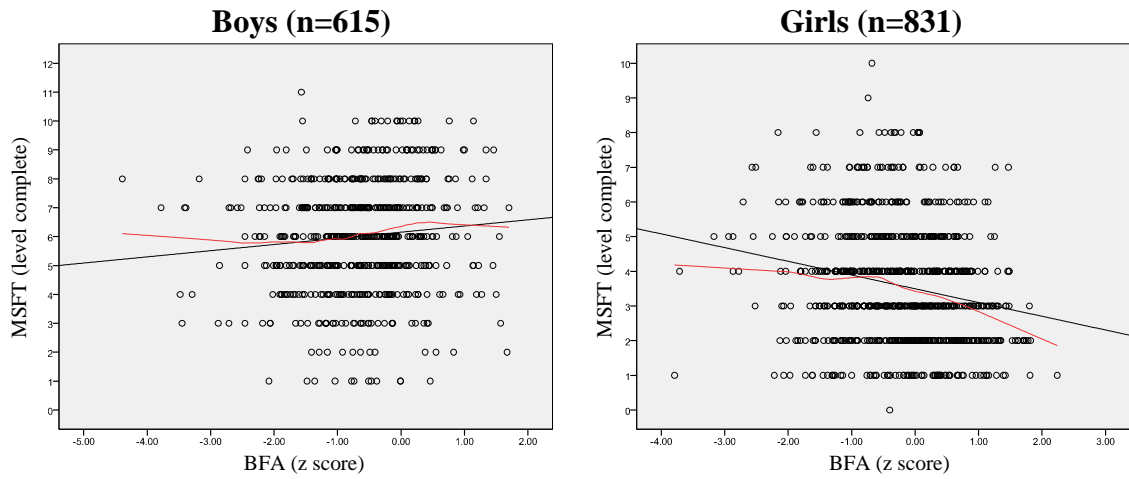


FIGURE A14.27: Scatter plot of crude results for HFA vs. SBJ in the Gulu sample at M1

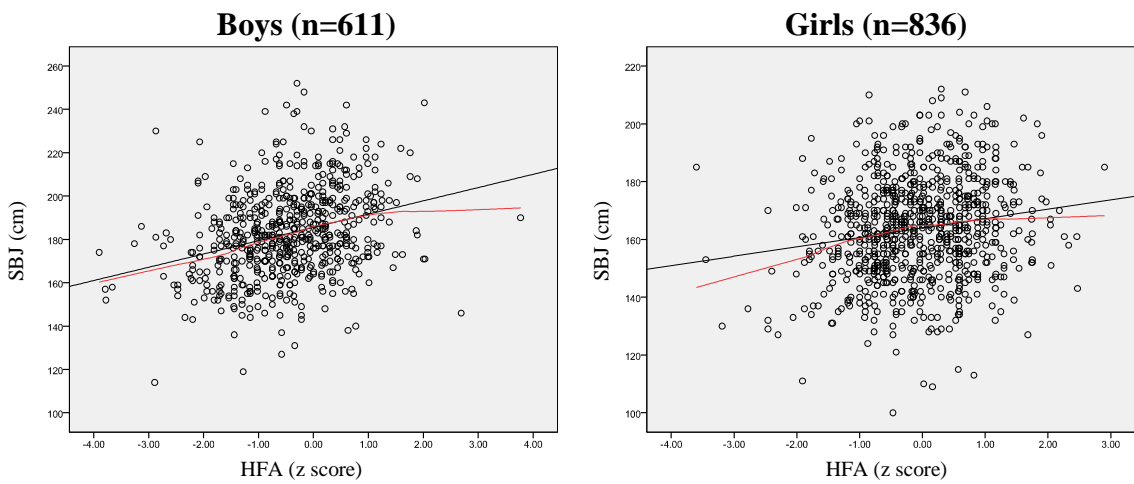
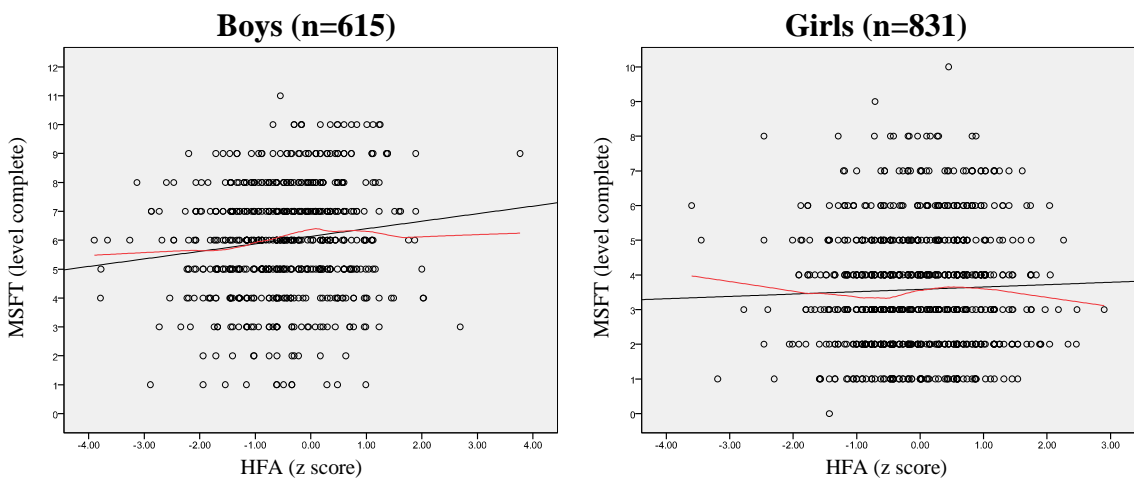


FIGURE A14.28: Scatter plot of crude results for HFA vs. MSFT in the Gulu sample at M1



APPENDIX A15: Relationship between Outcome Variables at M1 (Assumptions)

The statistical methods used to assess the relationship between the outcome variables at baseline assumed an approximately normally distributed sample. This was initially assessed visually using a histogram of the eligible sample for the crude results of each of the dependent variables (FIGURES A15.1 – A15.10).

FIGURE A15.1: Histograms of the distribution of the crude BFA results in the Gulu sample at M1 - stratified by gender

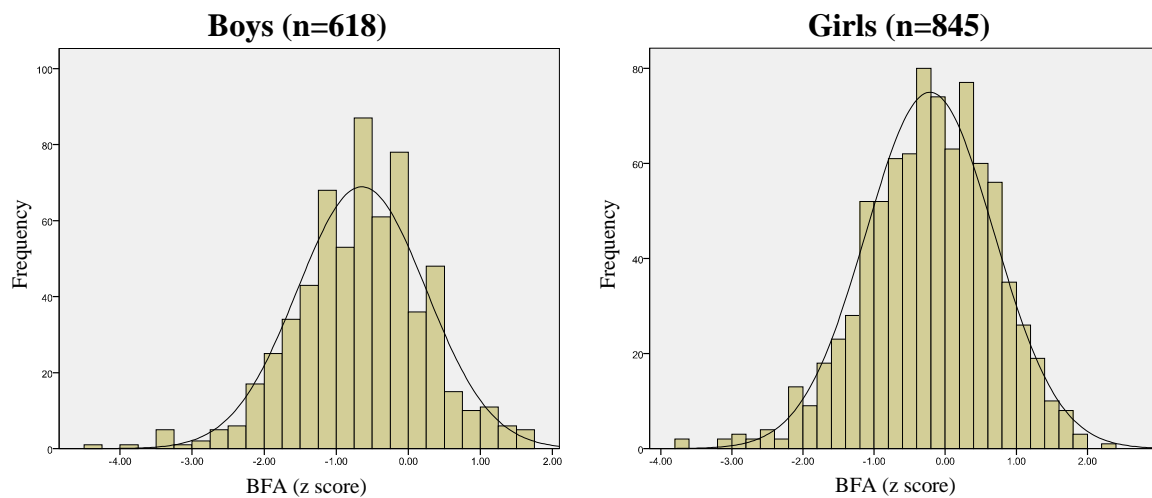


FIGURE A15.2: Histograms of the distribution of the crude HFA results in the Gulu sample at M1 - stratified by gender

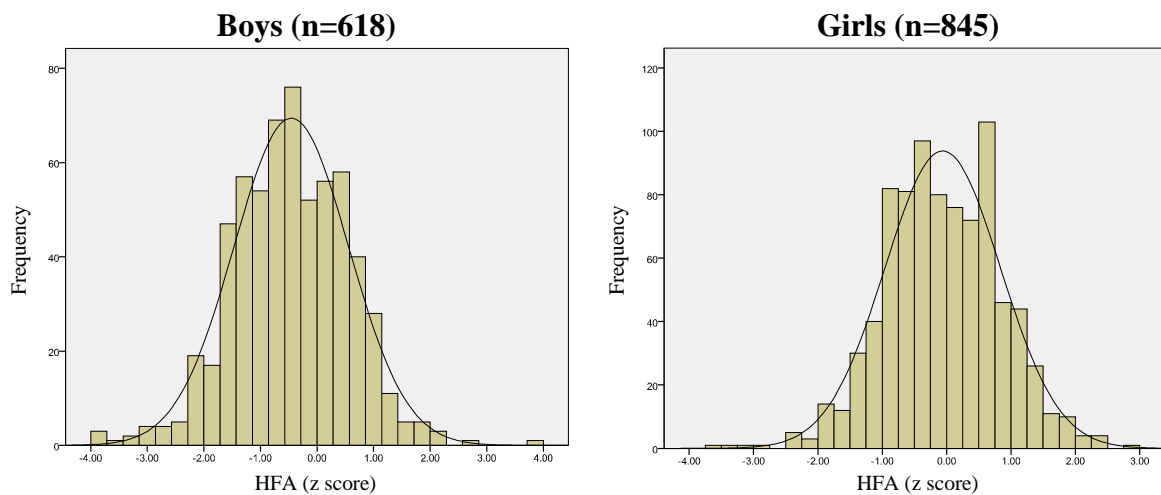


FIGURE A15.3: Histograms of the distribution of the crude SBJ results in the Gulu sample at M1 - stratified by gender

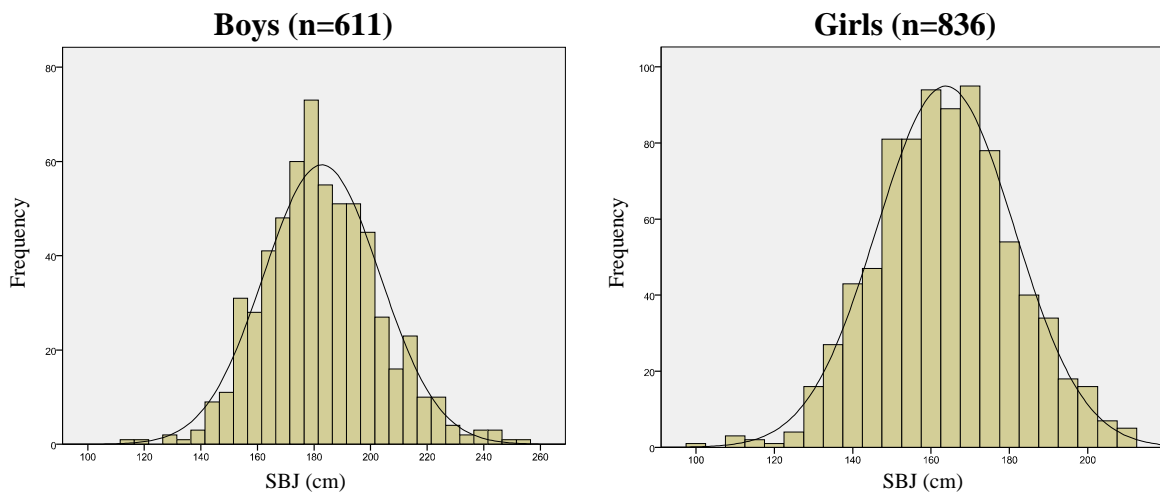


FIGURE A15.4: Histograms of the distribution of the crude MSFT results in the Gulu sample at M1 - stratified by gender

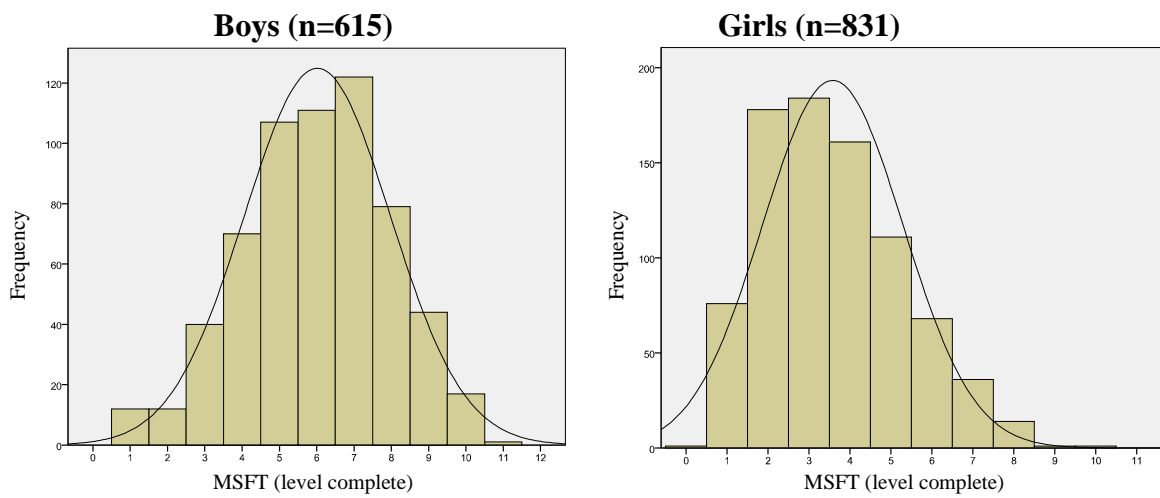


FIGURE A15.5: Histograms of the distribution of the crude TT results in the Gulu sample at M1 - stratified by gender

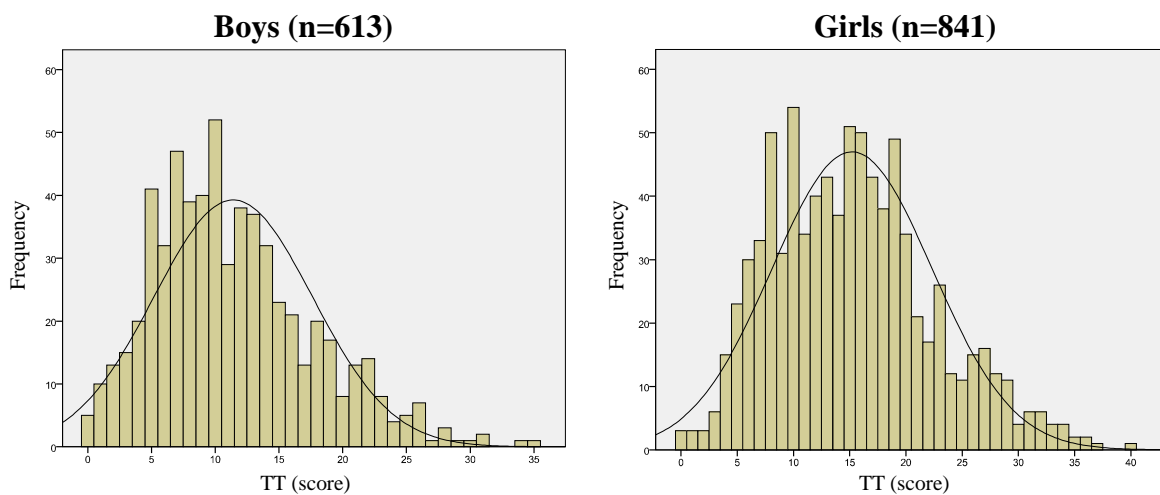


FIGURE A15.6: Histograms of the distribution of the crude K results in the Gulu sample at M1 - stratified by gender

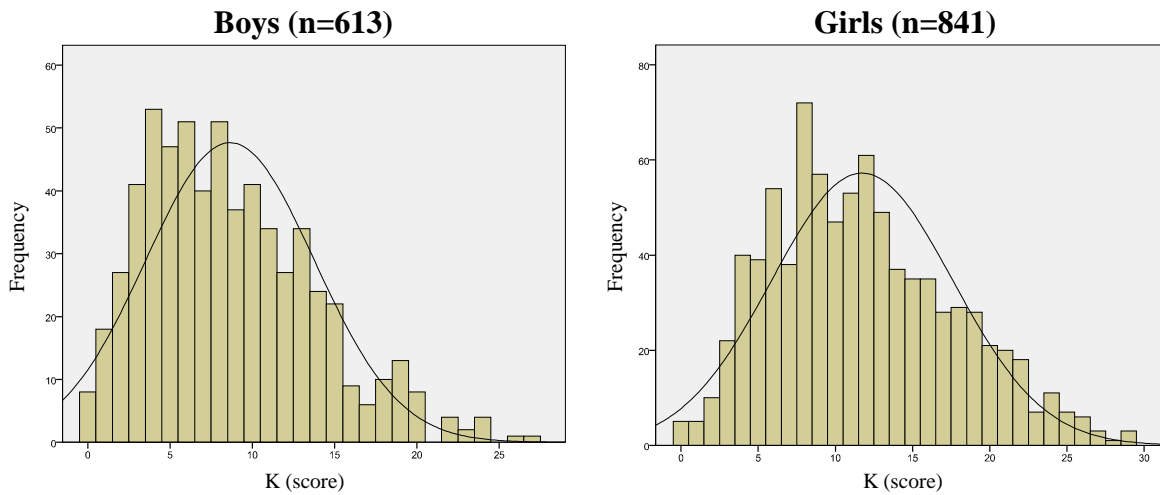


FIGURE A15.7: Histograms of the distribution of the crude P results in the Gulu sample at M1 - stratified by gender

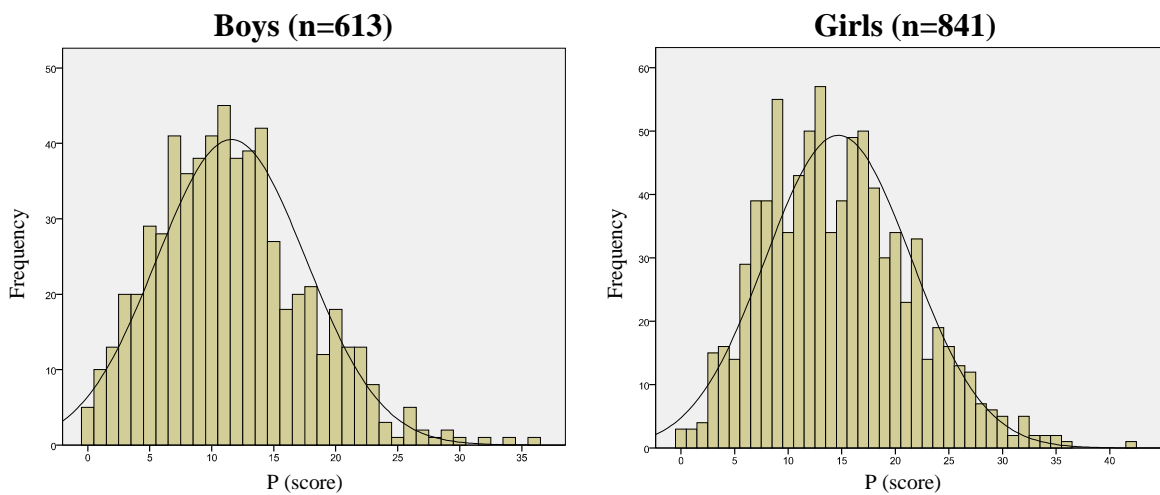


FIGURE A15.8: Histograms of the distribution of the crude ML results in the Gulu sample at M1 - stratified by gender

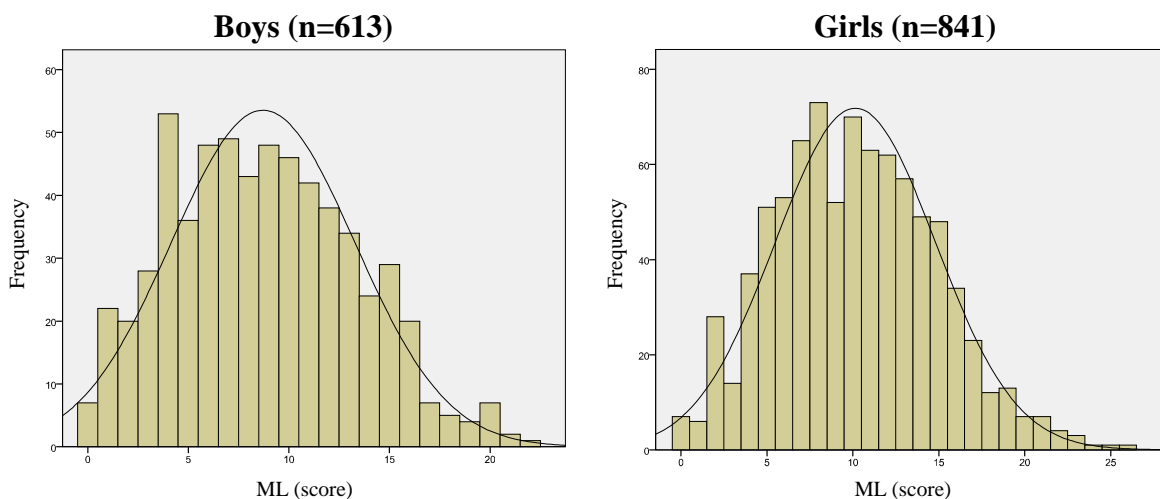


FIGURE A15.9: Histograms of the distribution of the crude DT results in the Gulu sample at M1 - stratified by gender

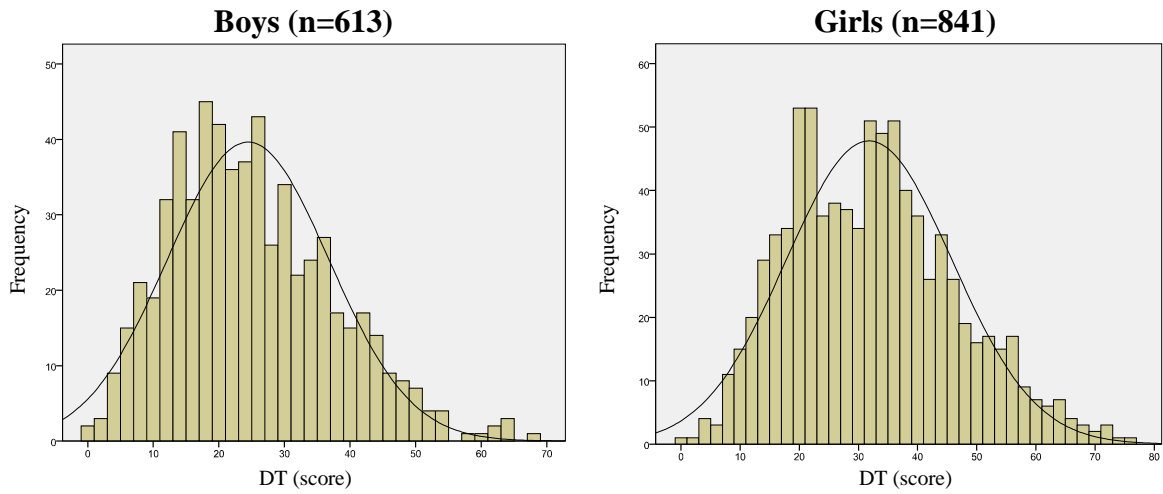
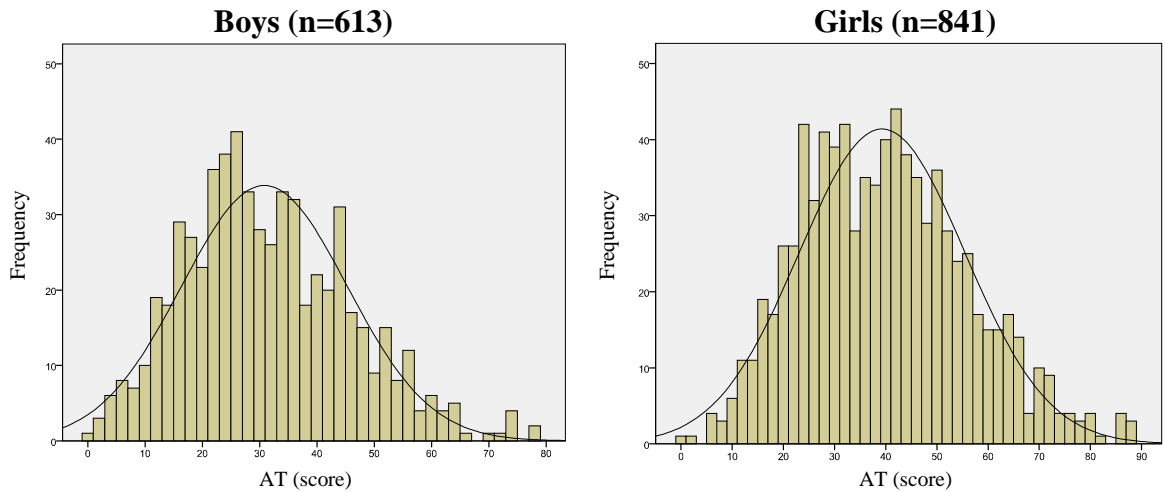


FIGURE A15.10: Histograms of the distribution of the crude AT results in the Gulu sample at M1 - stratified by gender



Although the distribution of all of the outcome variables appeared approximately normal, there were several deviations that warranted further assessment. Statistical tests for normality for the crude results were also completed (TABLE A15.1)

TABLE A15.1: Statistical tests for normality of the crude results for all of the outcomes in the Gulu sample at M1 - stratified by gender

Gender	Outcome variable	n	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
Boys (N=618)	BFA (z score)	618	-0.342	0.098	0.834	0.196	0.989	<0.001
	HFA (z score)	618	-0.119	0.098	0.661	0.196	0.994	0.012
	SBJ (cm)	611	0.270	0.099	0.407	0.197	0.993	0.006
	MSFT (level complete)	615	-0.210	0.099	-0.252	0.197	0.972	<0.001
	TT (score)	613	0.727	0.099	0.399	0.197	0.963	<0.001
	K (score)	613	0.717	0.099	0.227	0.197	0.958	<0.001
	P (score)	613	0.588	0.099	0.431	0.197	0.976	<0.001
	ML (score)	613	0.288	0.099	-0.485	0.197	0.981	<0.001
	DT (score)	613	0.574	0.099	0.097	0.197	0.976	<0.001
	AT (score)	613	0.473	0.099	0.010	0.197	0.983	<0.001
Girls (N=846)	BFA (z score)	845	-0.388	0.084	0.397	0.168	0.991	<0.001
	HFA (z score)	845	-0.093	0.084	0.335	0.168	0.995	0.010
	SBJ (cm)	836	-0.006	0.085	0.047	0.169	0.997	0.215
	MSFT (level complete)	831	0.548	0.085	-0.163	0.169	0.942	<0.001
	TT (score)	841	0.516	0.084	-0.051	0.168	0.977	<0.001
	K (score)	841	0.476	0.084	-0.339	0.168	0.974	<0.001
	P (score)	841	0.469	0.084	0.028	0.168	0.982	<0.001
	ML (score)	841	0.274	0.084	-0.189	0.168	0.988	<0.001
	DT (score)	841	0.426	0.084	-0.234	0.168	0.983	<0.001
	AT (score)	841	0.365	0.084	-0.215	0.168	0.988	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

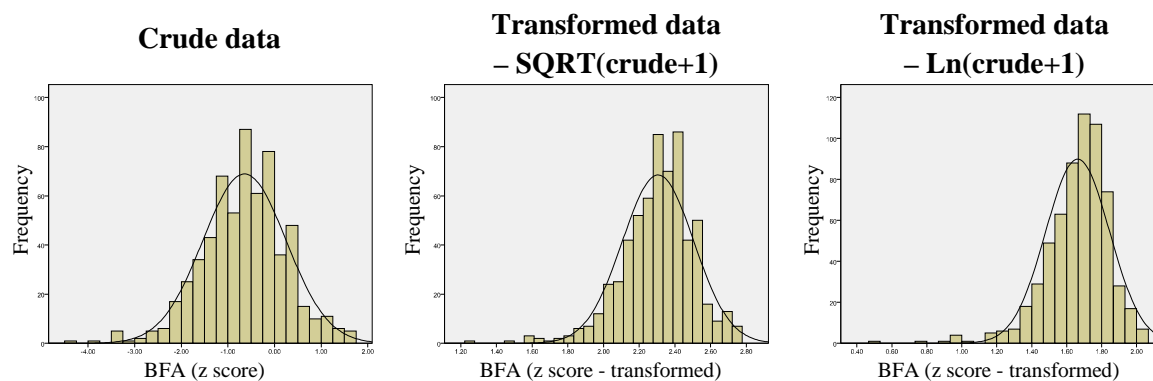
NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

The statistical tests for normality suggested that there were some significant deviations from an approximately normally distributed sample for the crude data. Standard transformations were performed for each of the outcome variables to assess if this improved the normality of the distribution. The square root (SQRT) and logarithmic (Ln) transformations were most effective, but the improvements were marginal at best (FIGURES A15.11 – A15.20 & TABLES A15.2 – A15.11).

FIGURE A15.11: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for BFA in the Gulu sample at M1 - stratified by gender

Boys (n=618)



Girls (n=845)

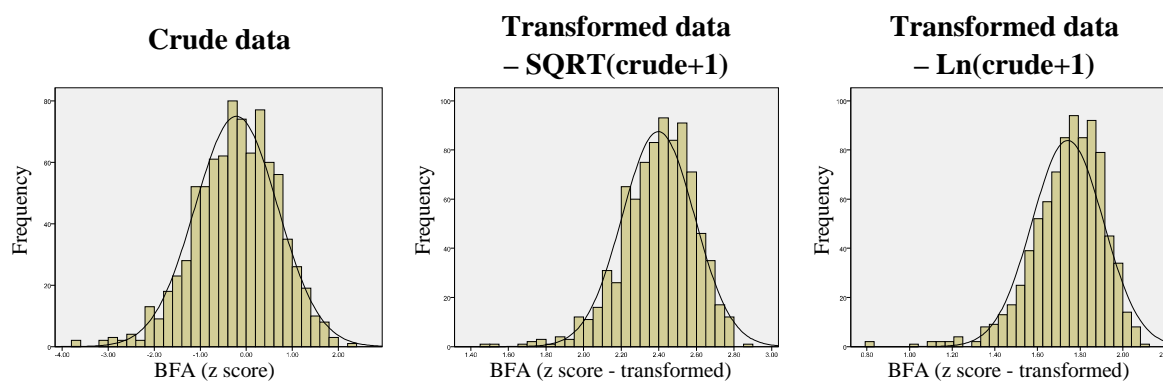


TABLE A15.2: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for BFA in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	618	<i>-0.342</i>	<i>-0.770</i>	<i>-1.334</i>	<i>0.834</i>	<i>2.004</i>	<i>4.639</i>	<i>0.989</i>	<i>0.969</i>	<i>0.930</i>
Girls	845	<i>-0.388</i>	<i>-0.709</i>	<i>-1.112</i>	<i>0.397</i>	<i>1.251</i>	<i>2.904</i>	<i>0.991</i>	<i>0.974</i>	<i>0.946</i>

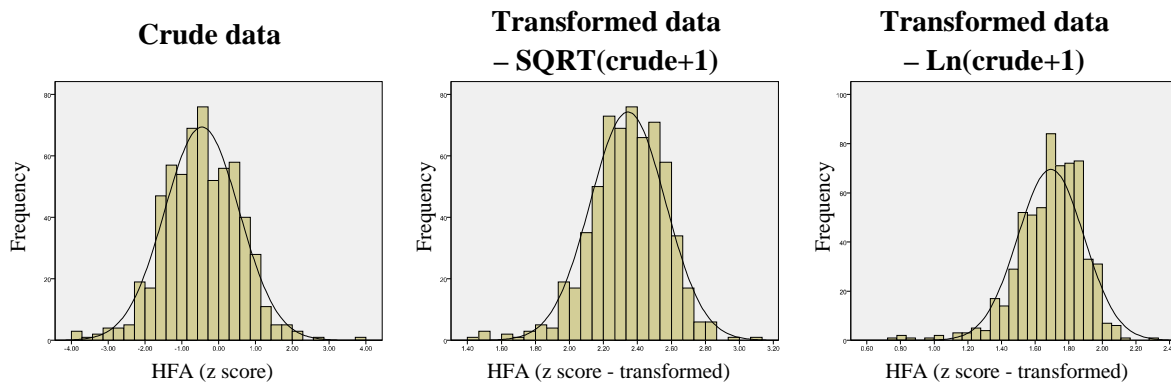
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.12: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for HFA in the Gulu sample at M1 - stratified by gender

Boys (n=618)



Girls (n=845)

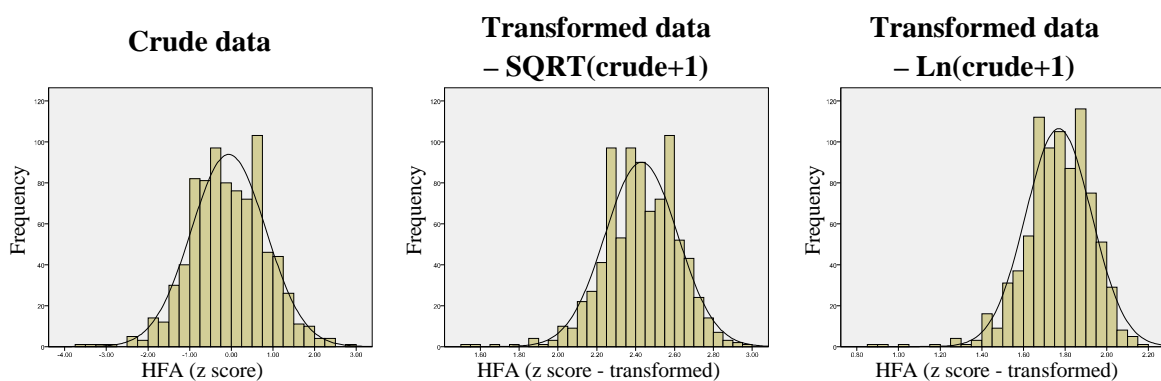


TABLE A15.3: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for HFA in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	618	-0.119	<i>-0.526</i>	-0.110	<i>0.661</i>	<i>1.199</i>	<i>2.669</i>	<i>0.994</i>	<i>0.981</i>	<i>0.952</i>
Girls	845	-0.093	<i>-0.403</i>	<i>-0.791</i>	0.335	<i>0.958</i>	<i>2.355</i>	<i>0.995</i>	<i>0.987</i>	<i>0.968</i>

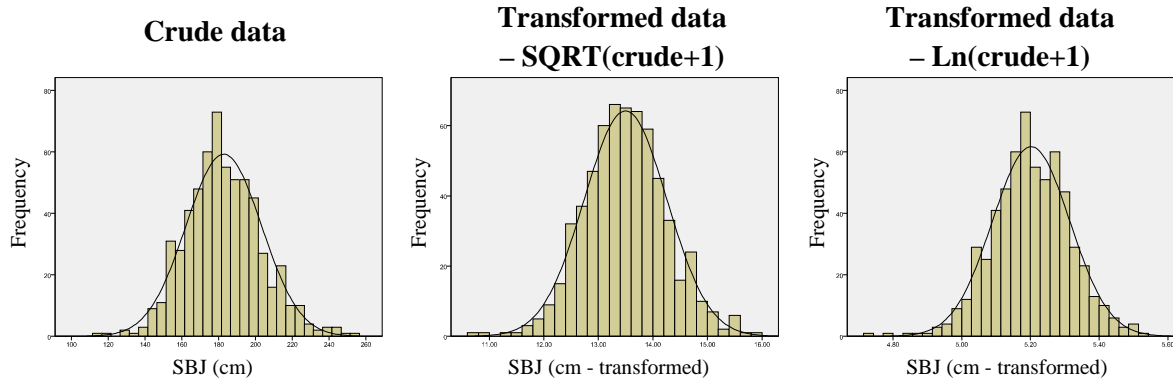
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.13: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for the SBJ in the Gulu sample at M1 - stratified by gender

Boys (n=611)



Girls (n=836)

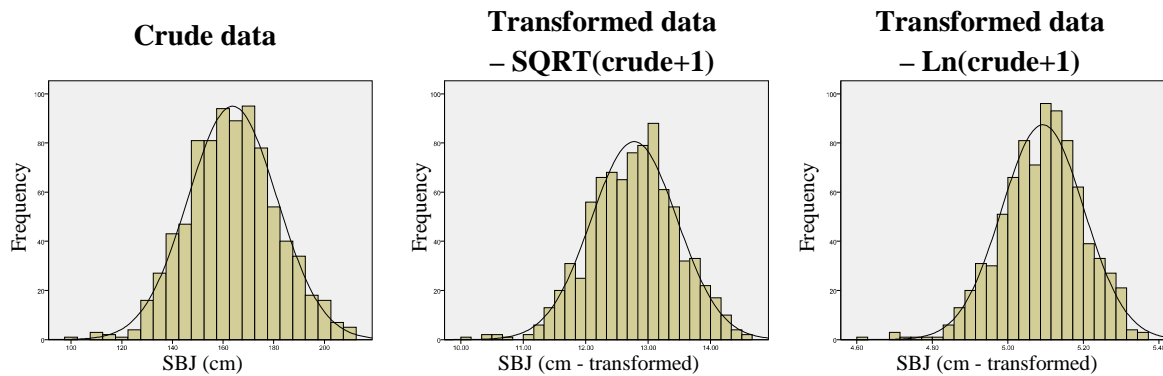


TABLE A15.4: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for SBJ in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	611	0.270	0.070	-0.139	0.407	0.409	0.578	0.993	0.996	0.995
Girls	836	-0.006	-0.181	-0.369	0.047	0.224	0.549	0.997	0.996	0.990

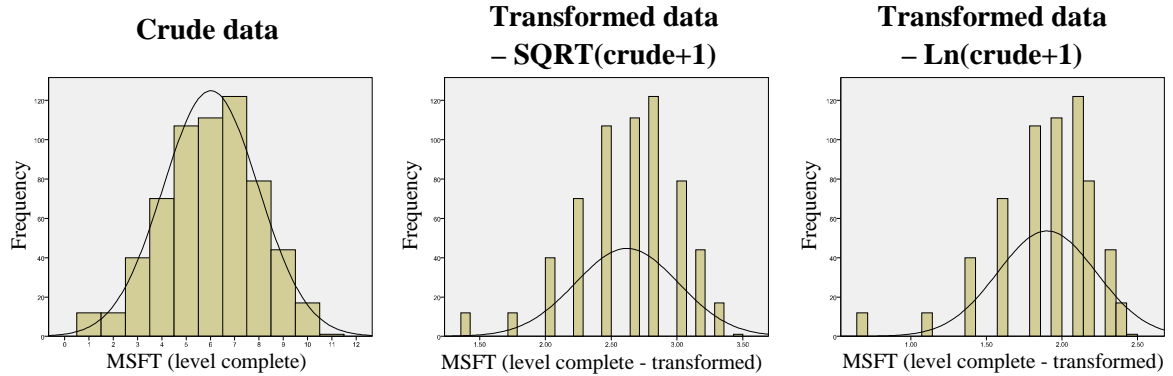
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.14: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for the MSFT in the Gulu sample at M1 - stratified by gender

Boys (n=615)



Girls (n=831)

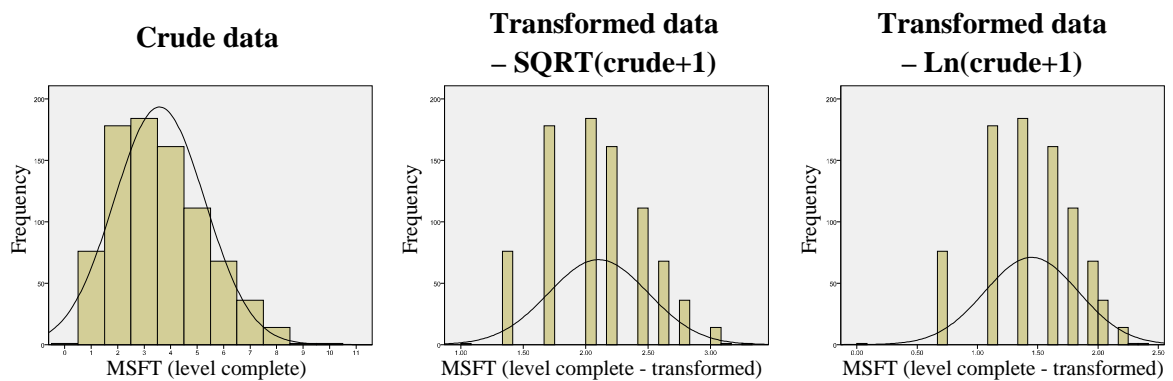


TABLE A15.5: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for MSFT in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	615	-0.210	-0.666	-1.250	-0.252	0.500	2.278	0.972	0.951	0.900
Girls	831	0.548	0.143	-0.279	-0.163	-0.543	-0.419	0.942	0.957	0.950

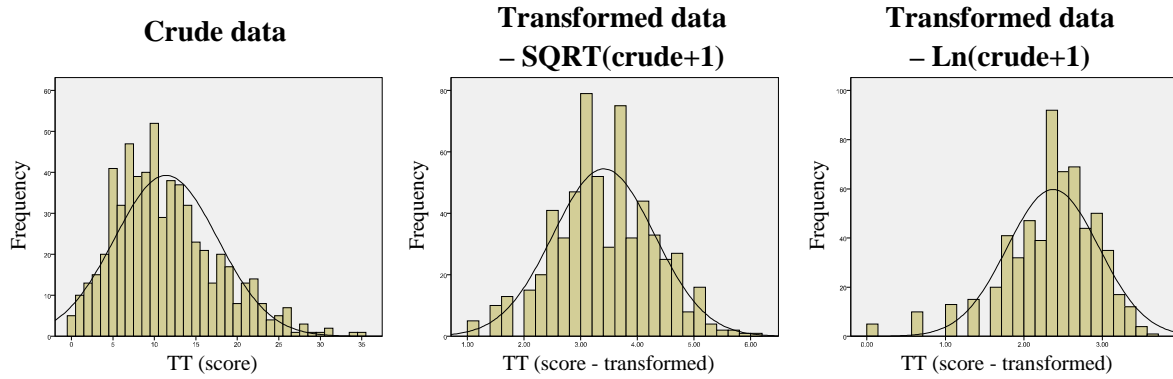
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.15: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for TT in the Gulu sample at M1 - stratified by gender

Boys (n=613)



Girls (n=841)

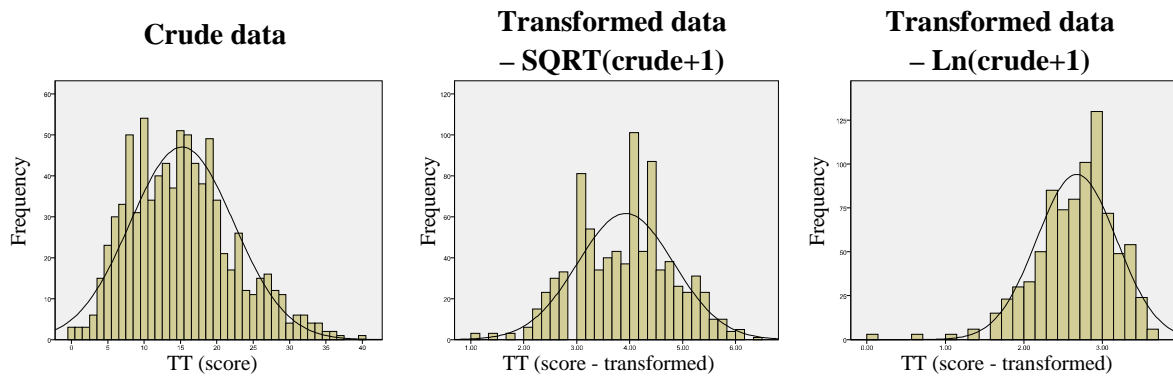


TABLE A15.6: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for TT in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	613	<i>0.727</i>	0.013	<i>-0.986</i>	<i>0.399</i>	-0.099	<i>1.841</i>	<i>0.963</i>	<i>0.995</i>	<i>0.948</i>
Girls	841	<i>0.516</i>	-0.089	<i>-1.043</i>	0.051	-0.161	<i>2.643</i>	<i>0.977</i>	<i>0.995</i>	<i>0.947</i>

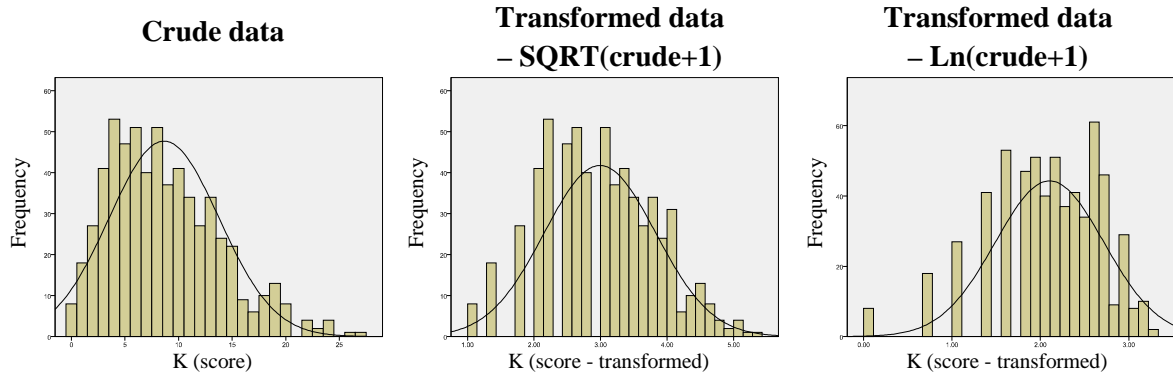
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.16: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for K in the Gulu sample at M1 - stratified by gender

Boys (n=613)



Girls (n=841)

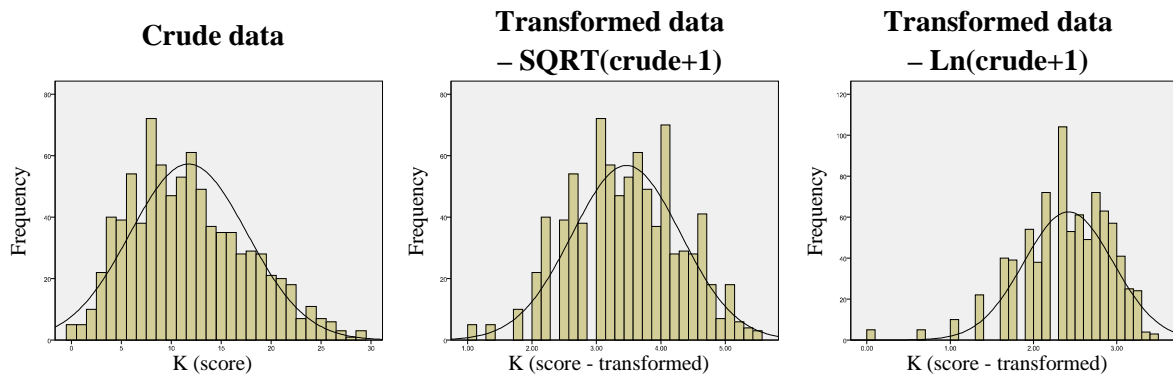


TABLE A15.7: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for K in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	613	0.717	0.062	-0.772	0.227	-0.382	0.770	0.958	0.991	0.959
Girls	841	0.476	-0.089	-0.965	-0.339	-0.353	1.891	0.974	0.992	0.949

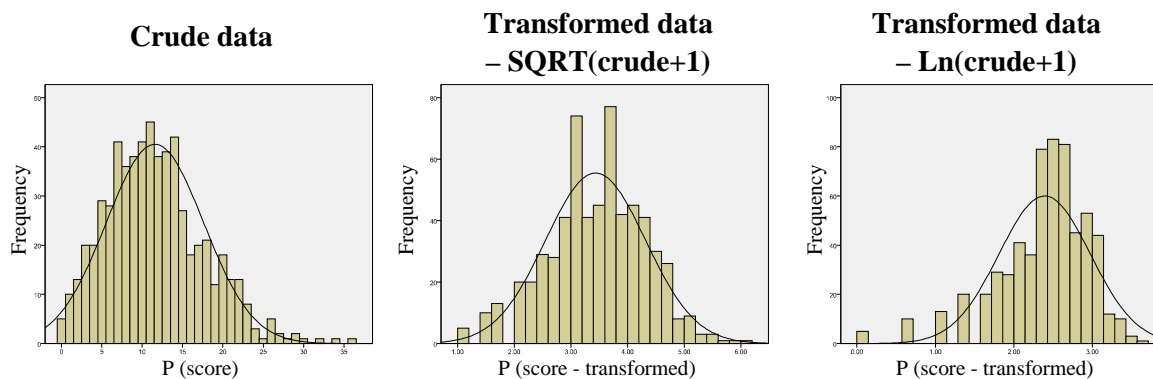
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.17: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for P in the Gulu sample at M1 - stratified by gender

Boys (n=613)



Girls (n=841)

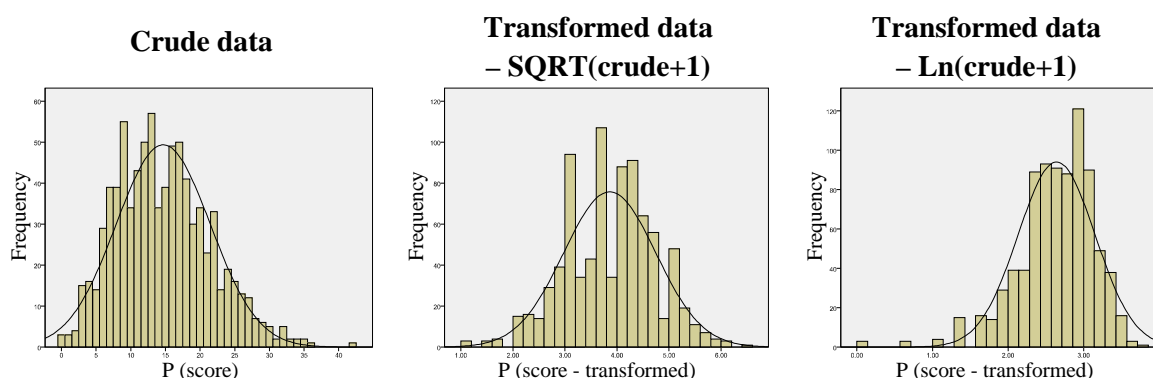


TABLE A15.8: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for P in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	613	<i>0.588</i>	-0.158	<i>-1.138</i>	<i>0.431</i>	-0.020	<i>2.055</i>	<i>0.976</i>	<i>0.994</i>	<i>0.933</i>
Girls	841	<i>0.469</i>	-0.160	<i>-1.113</i>	0.028	-0.082	<i>2.679</i>	<i>0.982</i>	<i>0.995</i>	<i>0.942</i>

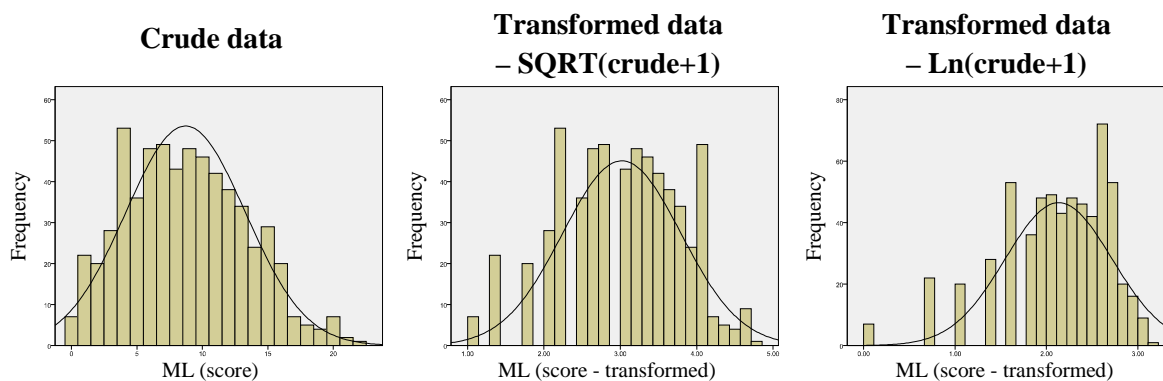
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.18: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for ML in the Gulu sample at M1 - stratified by gender

Boys (n=613)



Girls (n=841)

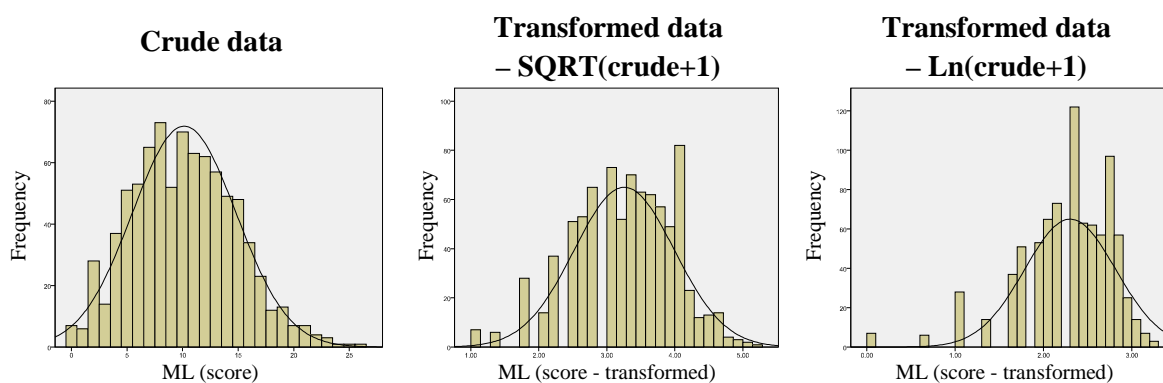


TABLE A15.9: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for ML in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	613	<i>0.288</i>	<i>-0.292</i>	<i>-1.074</i>	<i>-0.485</i>	-0.390	<i>1.310</i>	<i>0.981</i>	<i>0.984</i>	<i>0.926</i>
Girls	841	<i>0.274</i>	<i>-0.359</i>	<i>-1.293</i>	-0.189	0.032	<i>2.807</i>	<i>0.988</i>	<i>0.986</i>	<i>0.917</i>

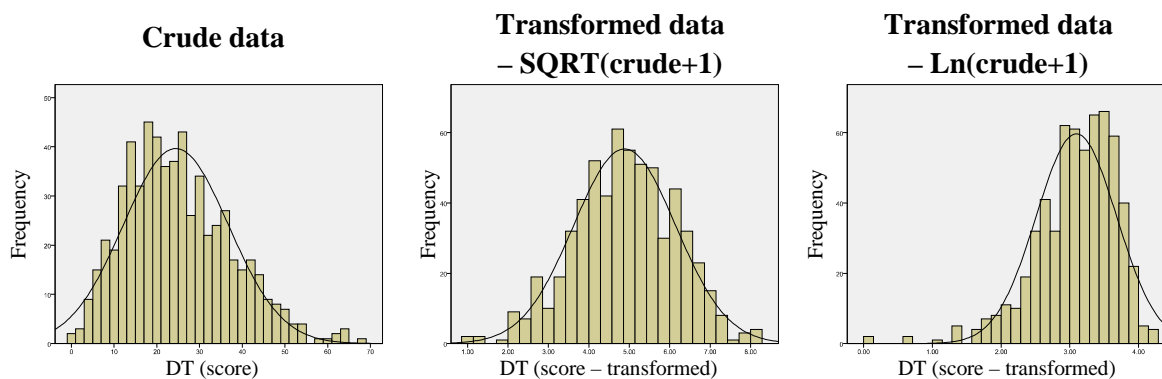
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.19: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for DT in the Gulu sample at M1 - stratified by gender

Boys (n=613)



Girls (n=841)

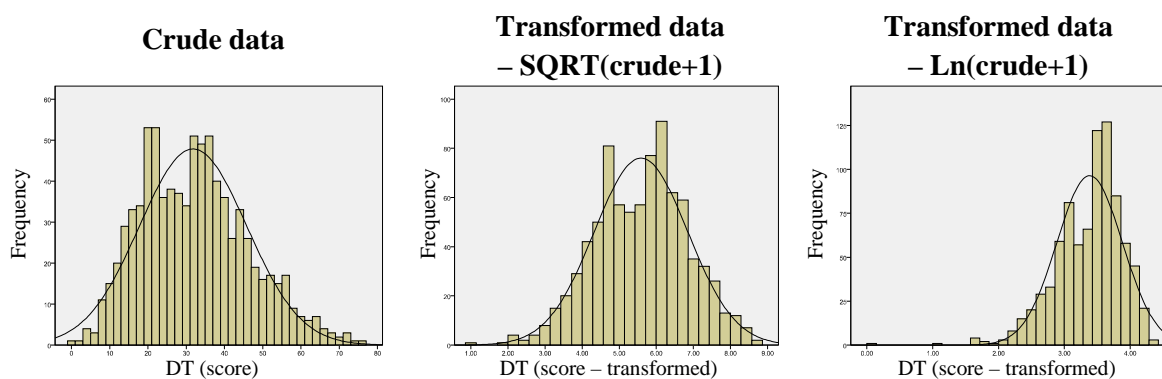


TABLE A15.10: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for DT in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	613	0.574	-0.112	-1.203	0.097	-0.138	3.023	0.976	0.997	0.936
Girls	841	0.426	-0.140	-1.080	-0.234	-0.229	3.085	0.983	0.996	0.948

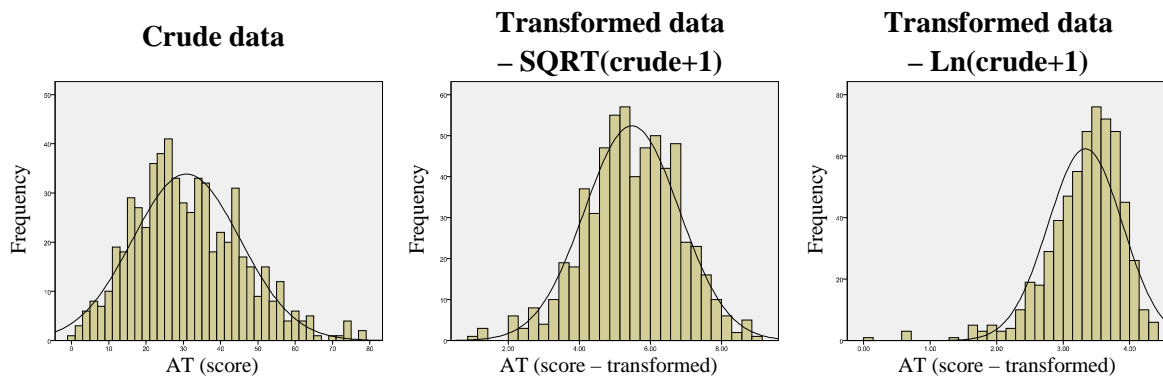
NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

FIGURE A15.20: Histograms comparing the distribution of the crude data to SQRT and Ln transformations for AT in the Gulu sample at M1 - stratified by gender

Boys (n=613)



Girls (n=841)

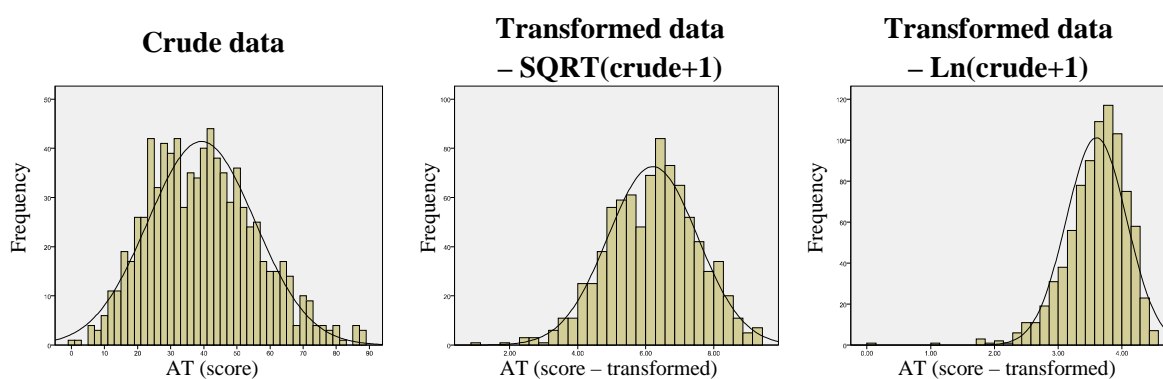


TABLE A15.11: Statistical tests for normality comparing the crude data to SQRT and Ln transformations for AT in the Gulu sample at M1 - stratified by gender

Gender	n	Skew test statistic			Kurtosis test statistic			Shapiro Wilk test stat		
		Crude	SQRT	Ln	Crude	SQRT	Ln	Crude	SQRT	Ln
Boys	613	0.473	-0.229	-1.426	0.010	0.074	4.217	0.983	0.995	0.918
Girls	841	0.365	-0.218	-1.322	-0.215	-0.020	5.013	0.988	0.995	0.934

NOTE: A perfectly normally distributed population has a skew and kurtosis = 0 ^[10].

NOTE: A perfectly normally distributed population has a Shapiro Wilk test statistic = 1 ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

In light of these results, the crude data for the outcome variables in the eligible population were treated as approximately normally distributed due to the following factors:

- Large sample size;
- Approximate normality on visual assessment;
- High sensitivity of the statistical tests utilised to assess normality;
- Marginal improvements from using standard transformations.

Therefore, statistical testing to assess the relationship between the outcome variables at baseline was completed using parametric procedures.

APPENDIX A16: Gum Marom Kids League Model (Intervention)

A16.1) Programme partners and stakeholders

The GMKL was delivered as a partnership between the Youth Coalition for Peace (YCP) and The Kids League (TKL). Since 2008, OA Projects (OAP) had funded and built the capacity of YCP to deliver peace-building activities through football interventions in Gulu. They had implemented youth football camps to promote peace-building and rehabilitated local sports infrastructure. TKL is a subsidiary of the Kampala Kids League (KKL) and was initially established to assist with the reintegration of child soldiers in northern Uganda. However, this focus diversified to incorporate a broad spectrum of objectives including positive health outcomes. Their programme in Gulu at the time of this research was funded by the Australian Sports Commission (ASC).

The GMKL also had the full support of local government officials and primary schools.

Other stakeholders within the community were identified at an open meeting that took place on 13 August, 2010. All interested parties were invited.

A16.2) Programme personnel

The programme staff included seven paid employees:

- Administration and Finance Coordinator
- Communications Coordinator
- Football Logistics Coordinator
- Peace-Building Coordinator
- Football Logistics Assistant x 2
- Women's Officer

40 volunteers from the local community were nominated to undergo training to become the GMKL coaches. They participated in a 6-day workshop that was coordinated and delivered by the GMKL programme staff. The training programme focused on developing football coaching and peace-building facilitation skills. 32 of these volunteers were selected based on workshop attendance and performance to become the GMKL coaches. The volunteer coaches were compensated for their transport costs to and from the GMKL activities.

All football games were refereed by six volunteers who had previously received official training and qualifications. They were compensated for their transport costs to and from the GMKL activities.

The RCI provided managerial oversight for the staff and volunteers delivering the intervention. Programme M&E for the health-related outcomes was also coordinated by the RCI. It was implemented by an independently funded research team that comprised five additional members. The peace-building components of M&E were assessed by selected coaches who were compensated for their time.

A16.3) Target population

Young adolescents aged 11-14 years in Gulu municipality.

A16.4) Dates of intervention

The season of the football league coincided with Term 3 of the 2010 school calendar (i.e. Monday 6 September – Friday 3rd December, 2010). It was intended that subsequent seasons of the GMKL would continue to follow the school calendar (i.e. three seasons per year).

A16.5) League structure

The league comprised eight clubs that were named after local bird species:

- Akuri FC
- Amam FC
- Aweno FC
- Ocwak FC
- Okok FC
- Owelo FC
- Twongwno FC
- Udu FC

Each club comprised four teams:

- U12 Boys
 - U12 Girls
 - U14 Boys
 - U14 Girls
- } (11-12 years of age)
- } (13-14 years of age)

Body weight and height were measured for all registrants. These measurements were used to validate their ages and/or provide an estimate for the adolescents who did not know how old they were in years. The age estimation value (Z) was calculated using a method described by the Gulu District Sports Office:

$$Z = 1.5 \times \text{height (cm)} + \text{weight (kg)}$$

The Z calculated for each registrant was then compared to parameters that were recommended by the District Sports Office and were modified for each age group (TABLE A16.1).

Registrants outside of these parameters were excluded to ensure the safety of all the GMKL participants.

TABLE A16.1: Estimating age for the GMKL using height and weight measurements in Gulu (Z)

Gender	Age Group	Z Range
BOYS	U12	220 - 270
	U14	270 - 320
GIRLS	U12	230 - 280
	U14	280 - 330

The number of players in each team was dictated by anticipated interest and demand prior to programme registration. The squads for the boys comprised 15 players and played 11-a-side on a full-size football field. The teams of girls had squads of 10 players and played 7-a-side football on a half-size field. Therefore, a total of 400 young adolescents (240 boys, 160 girls) in 32 teams participated in the football league.

The season took place over an 11-week period with activities routinely occurring each Saturday. The only exception was the rescheduling of the programme from Saturday 9 October to Sunday 10 October due to local celebrations for Independence Day (TABLE A16.2). The season commenced with programme registration (week 1) and team allocation (week 2). All clubs then played each other on one occasion in a round-robin competition (weeks 3 – 9). This was followed by a full-day tournament (week 10) and parents/guardians day with final presentations (week 11). Therefore, a total of nine weeks of football intervention was delivered (weeks 3 – 11).

TABLE A16.2: Weekly programme for the GMKL in Gulu

Week 1: Programme registration Saturday 11 September			
Week 2: Team allocation Saturday 18 September			
Week 3: Round-robin competition (1) Saturday 25 September			
Game 1	Amam FC	vs.	Akuri FC
Game 2	Aweno FC	vs.	Ocwak FC
Game 3	Okok FC	vs.	Owelo FC
Game 4	Udu FC	vs.	Twongwno FC
Week 4: Round-robin competition (2) Saturday 02 October			
Game 1	Amam FC	vs.	Aweno FC
Game 2	Akuri FC	vs.	Ocwak FC
Game 3	Okok FC	vs.	Twongwno FC
Game 4	Owelo FC	vs.	Udu FC
Week 5: Round-robin competition (3) Sunday 10 October			
Game 1	Amam FC	vs.	Ocwak FC
Game 2	Akuri FC	vs.	Aweno FC
Game 3	Okok FC	vs.	Udu FC
Game 4	Owelo FC	vs.	Twongwno FC
Week 6: Round-robin competition (4) Saturday 16 October			
Game 1	Amam FC	vs.	Okok FC
Game 2	Akuri FC	vs.	Owelo FC
Game 3	Aweno FC	vs.	Twongwno FC
Game 4	Ocwak FC	vs.	Udu FC
Week 7: Round-robin competition (5) Saturday 23 October			
Game 1	Amam FC	vs.	Owelo FC
Game 2	Akuri FC	vs.	Twongwno FC
Game 3	Aweno FC	vs.	Udu FC
Game 4	Ocwak FC	vs.	Okok FC
Week 8: Round-robin competition (6) Saturday 30 October			
Game 1	Amam FC	vs.	Twongwno FC
Game 2	Akuri FC	vs.	Udu FC
Game 3	Aweno FC	vs.	Okok FC
Game 4	Ocwak FC	vs.	Owelo FC
Week 9: Round-robin competition (7) Saturday 06 November			
Game 1	Amam FC	vs.	Udu FC
Game 2	Akuri FC	vs.	Okok FC
Game 3	Aweno FC	vs.	Owelo FC
Game 4	Ocwak FC	vs.	Twongwno FC
Week 10: The GMKL tournament Saturday 13 November			
Full day programme to include football tournament and peace-building activities for all clubs and their members.			
Week 11: The parents/guardians day Saturday 20 November			
Football and peace-building activities to include all club members and their parents/guardians.			
FINAL PRESENTATIONS			

During the season all teams also participated in at least one independent training session each week that occurred after school hours (i.e. 4:30 – 6:00 pm).

It is important to note that although the league was thoroughly scheduled, it was rare that the timetable was followed. Most activities started late due to poor planning and time management by the programme personnel, volunteers and players. Changes in schedule were common in response to demands as they arose. Flexibility was a key factor in ensuring that all of the programme activities were successfully completed.

A16.5a) Location of league activities

The programme registration, team allocation, round-robin competition, tournament and parents/guardians day took place at Pece Stadium (U14) and Gulu Public Primary School (U12).

For the training sessions during the week, two clubs were located in each of the four divisions within Gulu municipality:

- Bardege – Amam FC and Akuri FC
- Laroo – Aweno FC and Ocwak FC
- Layibi – Okok FC and Owelo FC
- Pece – Twongwno FC and Udu FC

Each team was allocated a training venue and day (TABLE A16.3).

TABLE A16.3: Club training schedule for the GMKL in Gulu

Club	Training Ground	Training Day	Team
Amam FC	Bar Mony Barracks	Monday	U12 Boys
			U12 Girls
		Wednesday	U14 Boys
			U14 Girls
Akuri FC	Bar Mony Barracks	Tuesday	U12 Boys
			U12 Girls
		Thursday	U14 Boys
			U14 Girls
Aweno FC	Gulu Town Primary School	Monday	U12 Boys
			U12 Girls
		Wednesday	U14 Boys
			U14 Girls
Ocwak FC	Gulu Town Primary School	Tuesday	U12 Boys
			U12 Girls
		Thursday	U14 Boys
			U14 Girls
Okok FC	Layibi Central Primary School	Monday	U12 Boys
			U12 Girls
		Wednesday	U14 Boys
			U14 Girls
Owelo FC	Layibi Central Primary School	Tuesday	U12 Boys
			U12 Girls
		Thursday	U14 Boys
			U14 Girls
Twongwono FC	Pece P7 Primary School	Monday	U12 Boys
			U12 Girls
		Wednesday	U14 Boys
			U14 Girls
Udu FC	Pece P7 Primary School	Tuesday	U12 Boys
			U12 Girls
		Thursday	U14 Boys
			U14 Girls

The coaches from each club were able to negotiate alternative training times by sharing the facilities that were organised by the programme staff. They could also organise additional training sessions at other venues at their own discretion and in cooperation with their players.

A16.5b) Logistics for programme registration

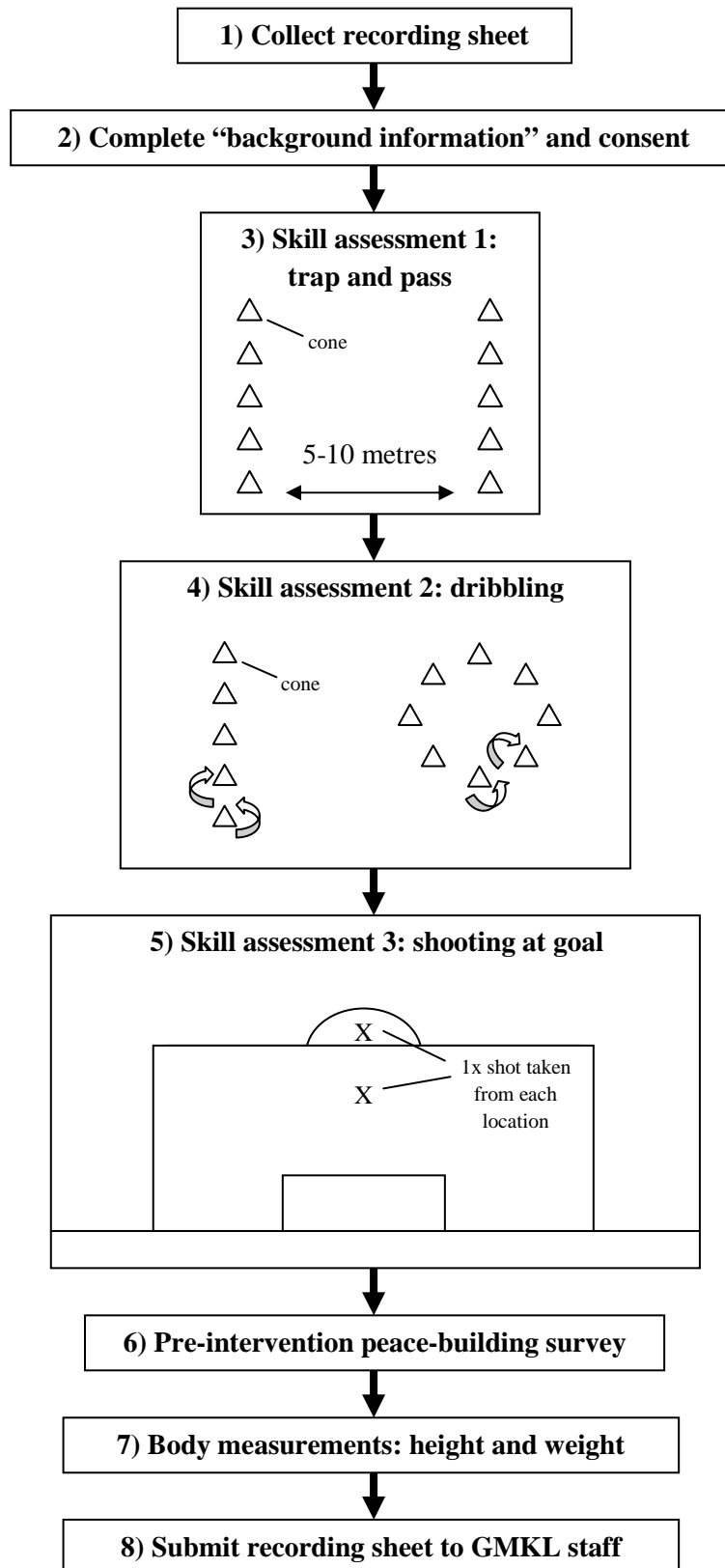
Programme registration was promoted using bulletin board public notices, radio announcements, newspaper advertisements and word of mouth. Registration was open to eligible adolescents with a parent/guardian in attendance who consented to their child participating in the GMKL. The parent/guardian was also required to provide information for programme risk management (allergies, medical conditions, emergency contact).

Registration procedures included the collection of other background data that assisted with player identification and team allocation (name, gender, age, school, class, residential division). Assessment of the height (cm), weight (kg) and basic football skills (score: 5 = excellent; 4 = very good; 3 = good; 2 = fair; 1 = poor) of the registrants was also used for these purposes.

All registrants were interviewed individually to complete a questionnaire about various peace-building objectives. This was part of the pre-intervention peace-building assessment for M&E purposes.

The registration process comprised a number of assessment stations where programme staff and volunteers assisted the participants to complete each section of the prepared recording sheet (FIGURE A16.1).

FIGURE A16.1: Flow-diagram of the GMKL registration process in Gulu



When submitting their registration form the participants were advised to return to the same location the following week for the announcement of team allocation.

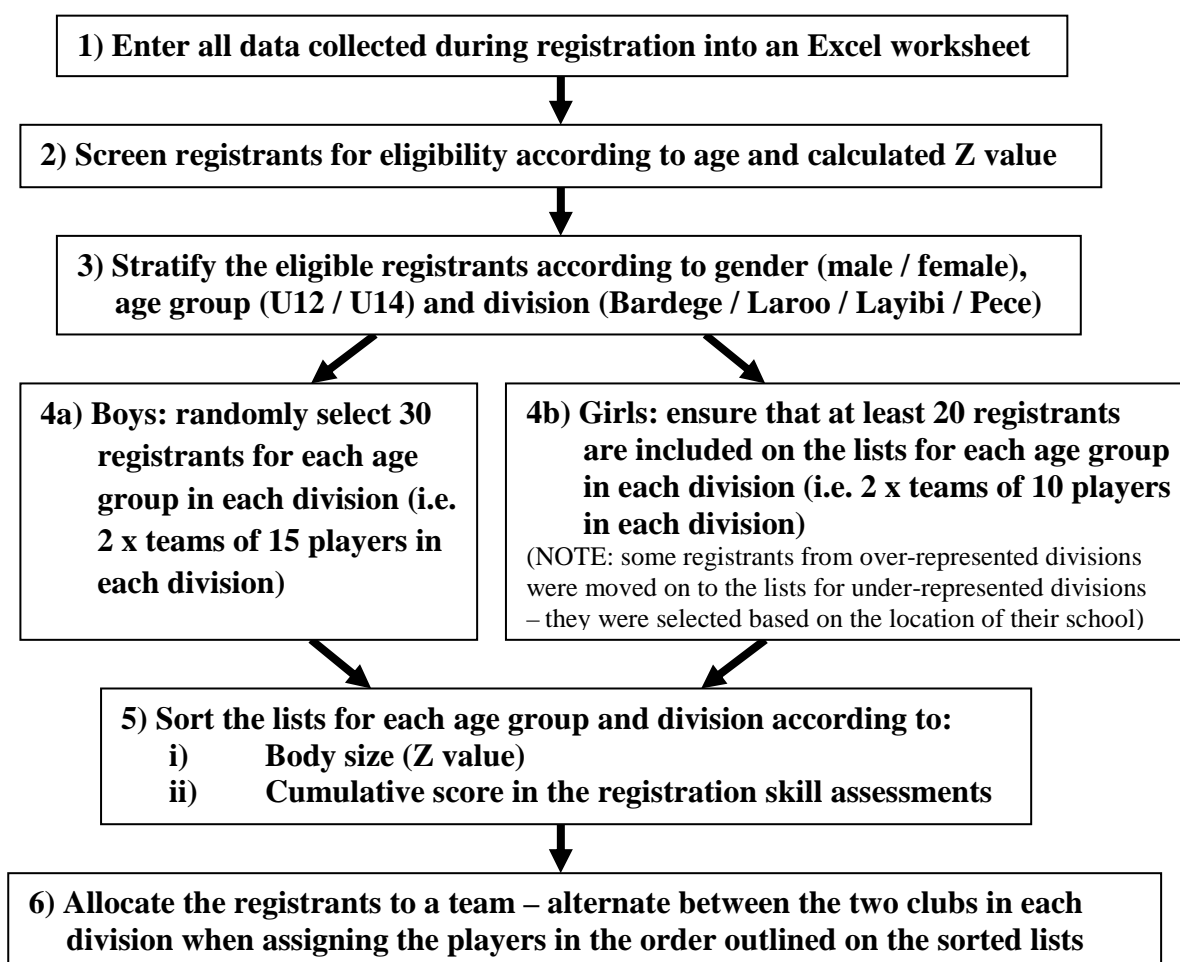
A16.5c) Methods for team allocation

Team allocation was designed to:

- Ensure the eligibility of all adolescents included in the GMKL;
- Randomly select eligible adolescents to participate in the GMKL or be wait-listed for the next season if the programme was over-subscribed;
- Assign participants and coaches to teams that would enable them to conveniently attend training;
- Balance different teams evenly according to body size and football skill level.

These objectives were achieved by applying a systematic team allocation procedure (FIGURE A16.2)

FIGURE A16.2: Flow-diagram of the team allocation procedure for the GMKL in Gulu



The boys who were not randomly selected for inclusion were wait-listed and given priority entrance into the next GMKL season (i.e. commencing January 2011). The league for the girls was not over-subscribed and there was no waiting list generated.

Selected coaches were assigned to a team that trained in the division in which they lived or worked. Male and female coaches were mixed to ensure that all clubs had at least one coach of each gender.

A16.5d) Logistics for round-robin competition

The GMKL activities on game days during the round robin competition were scheduled to commence at 8:00am and finish by 2:00pm. Football games comprised 2 x 20 minute halves with a five-minute half-time break. All players in each squad had to play at least one half of football each week and the players who started on the bench were rotated by the volunteer coaches. Bottled water and a fruit snack were provided for all participants at the completion of the game.

All games were preceded by the performance of a peace-building task by the two opposing teams. This activity changed each week and took approximately 20 minutes.

Therefore, the time for peace-building performances and football games totalled approximately one hour each week (TABLE A16.4).

TABLE A16.4: The GMKL game day schedule for round-robin competition in Gulu

TIME	GAME
8:00am - 9:00am	Boys – peace-building and game 1 (full field)
9:00am - 10:00am	Boys - peace-building and game 2 (full field)
10:00am - 11:00am	Girls - peace-building and games 1 and 2 (half field)
11:00am - 12:00pm	Girls - peace-building and games 3 and 4 (half field)
12:00pm - 1:00pm	Boys - peace-building and game 3 (full field)
1:00pm - 2:00pm	Boys - peace-building and game 4 (full field)

The volunteer coaches also spent time preparing and debriefing their team members before and after the round-robin activities each week. The content and duration of this varied among coaches and teams.

A16.5e) Logistics for tournament

The tournament was scheduled to commence at 8:00am and finish by 3:00pm. Football games comprised 2 x 8 minute halves with no half-time break (i.e. swap ends only). The draw was designed so that all teams would play three games and a final ranking could be established for the teams in each age group. The round one fixtures were determined randomly and drawn games were decided by a penalty shootout (3 shots). A peace-building drama performance was delivered by selected coaches and staff before commencing round 2 of the tournament (TABLE A16.5).

TABLE A16.5: The GMKL tournament schedule in Gulu

Game	Time	Field	Round 1 - Boys		
1	8:00am	Full	Akuri FC	vs.	Udu FC
2	8:20am	Full	Twongwno FC	vs.	Aweno FC
3	8:40am	Full	Amam FC	vs.	Okok FC
4	9:00am	Full	Ocwak FC	vs.	Owelo FC

Game	Time	Field	Round 1 - Girls		
5	9:20am	A	Akuri FC	vs.	Udu FC
6	9:20am	B	Twongwno FC	vs.	Aweno FC
7	9:40am	A	Amam FC	vs.	Okok FC
8	9:40am	B	Ocwak FC	vs.	Owelo FC

10:00am	Peace-building activity (drama performance)
----------------	--

Game	Time	Field	Round 2 - Boys		
9	11:00am	Full	Loser of game 1	vs.	Loser of game 2
10	11:20am	Full	Loser of game 3	vs.	Loser of game 4
11	11:40am	Full	Winner of game 1	vs.	Winner of game 2
12	12:00pm	Full	Winner of game 3	vs.	Winner of game 4

Game	Time	Field	Round 2 - Girls		
13	12:20pm	A	Loser of game 5	vs.	Loser of game 6
14	12:20pm	B	Loser of game 7	vs.	Loser of game 8
15	12:40pm	A	Winner of game 5	vs.	Winner of game 6
16	12:40pm	B	Winner of game 7	vs.	Winner of game 8

Game	Time	Field	Playoff for 7th and 8th - Girls		
17	1:00pm	A	Loser of game 13	vs.	Loser of game 14

Game	Time	Field	Playoff for 5th and 6th - Girls		
18	1:00pm	B	Winner of game 13	vs.	Winner of game 14

Game	Time	Field	Playoff for 3rd and 4th - Girls		
19	1:20pm	A	Loser of game 15	vs.	Loser of game 16

Game	Time	Field	Playoff for 1st and 2nd - Girls		
20	1:20pm	B	Winner of game 15	vs.	Winner of game 16

Game	Time	Field	Playoff for 7th and 8th - Boys		
21	1:40pm	Full	Loser of game 9	vs.	Loser of game 10

Game	Time	Field	Playoff for 5th and 6th - Boys		
22	2:00pm	Full	Winner of game 9	vs.	Winner of game 10

Game	Time	Field	Playoff for 3rd and 4th - Boys		
23	2:20pm	Full	Loser of game 11	vs.	Loser of game 12

Game	Time	Field	Playoff for 1st and 2nd - Boys		
24	2:40pm	Full	Winner of game 11	vs.	Winner of game 12

The players and volunteer coaches were encouraged to support and participate in all of the football games for their club in the tournament. Bottled water and fruit snacks were provided for all participants throughout the day.

A16.5f) Logistics for parents / guardians day

Parents/guardians day was scheduled to commence at 8:00am and finish by 3:00pm. Football games comprised 2 x 10 minute halves with a 5-minute break at half time. The children and parents/guardians from each team played one game against each other. Volunteer coaches and the GMKL staff filled the gaps in the parents/guardians teams if there were an insufficient number in attendance. The games were preceded and followed by peace-building activities coordinated by each of the volunteer coaches. One hour was also put aside at the end of the day for final presentations to all the participants and to recognise the GMKL club champions (TABLE A16.6).

TABLE A16.6: The parents/guardians day schedule in Gulu

Game	Time	Field	Boys vs. parents/guardians
1	8:00am	Full	Akuri FC
2	8:30am	Full	Amam FC
3	9:00am	Full	Aweno FC
4	9:30am	Full	Ocwak FC

Game	Time	Field	Girls vs. parents/guardians
5	10:00am	A	Akuri FC
6	10:00am	B	Amam FC
7	10:30am	A	Aweno FC
8	10:30am	B	Ocwak FC
9	11:00am	A	Okok FC
10	11:00am	B	Owelo FC
11	11:30am	A	Twongwno FC
12	11:30am	B	Udu FC

Game	Time	Field	Boys vs. parents/guardians
13	12:00	Full	Okok FC
14	12:30pm	Full	Owelo FC
15	1:00pm	Full	Twongwno FC
16	1:30pm	Full	Udu FC

2:00pm - 3:00pm	Final presentations
------------------------	----------------------------

The participants and their parents/guardians were encouraged to enjoy all of the football and peace-building activities and stay for the final presentations. Bottled water and fruit snacks were supplied for all participants.

A16.5g) League points allocation

The focus of the GMKL was to win the Club Trophy. This required all four teams (U12 boys, U12 girls, U14 boys, U14 girls) to work together to maximise the club points.

It is important to note that the focus of the GMKL was on the community rather than the football results. Only 30% of the points in the competition could be attained from winning football games. The remaining points were earned for maintaining good on-field behaviour, contributing to peace-building activities and completing community service (TABLE A16.7).

TABLE A16.7: The GMKL points allocation system in Gulu

Activity	Points awarded
Football league and football tournament (MAXIMUM 30 POINTS)	Teams were awarded points according to results: - League (7 weeks): Win = 3 points, Draw = 1 point, Loss = 0 points. - Tournament: 1 st = 9 points, 2 nd = 8 points, 3 rd = 7 points, 4 th = 6 points, 5 th = 5 points, 6 th = 4 points, 7 th = 3 points, 8 th = 2 points.
On field behaviour (MAXIMUM 25 POINTS)	Teams commenced the season with maximum (25) points and deductions were made for poor on-field behaviour: - Deduct 3 points for a red card. - Deduct 1 point for a yellow card.
Contribution to peace-building activities (MAXIMUM 25 POINTS)	Teams were awarded points based on attendance and engagement in the peace-building activities that took place before each game: - Maximum of 3 points before each league game. - Maximum of 4 points for parental/guardian participation on the parents/guardians day.
Community service (MAXIMUM 20 POINTS)	Teams received 4 points for each community service activity (approx. 1 hour) completed as an inclusive group (up to maximum of 20 points). Maintenance and slashing of club training grounds could contribute to community service points.

A weekly newsletter was produced and disseminated to notify the community about the GMKL activities, promote programme highlights and distribute an updated club points table.

A16.6) Programme equipment

Each coach was provided with the following equipment to assist with the GMKL activities for the duration of the season:

- 3x footballs
- 1x ball pump and needle
- 1x bib (goalkeeper)
- 15 (boys) / 10 (girls) x club uniform shirts
- 5x coloured markers (small plastic bowl)

Portable goals and corner posts for games were also provided. All of the equipment was purchased and/or made in Gulu.

A16.7) Programme facilities

Both the partner organisations rented office space in Gulu. This was used to complete administrative duties and provide a meeting space for programme staff and volunteers. Programme equipment was also stored in each of the organisation offices.

OAP and YCP were also engaged in rehabilitating two football grounds within Gulu municipality for future GMKL seasons. This process involved clearing the allocated land, levelling the surface, planting the grass seed, fencing the area, demarcating the playing arena and constructing goal posts. These tasks were completed by local people using equipment available in Gulu. Both of the fields were handed over to the local community for use and ongoing maintenance after the rehabilitation was complete.

A16.8) Programme budget

The projected budget to cover all expenses for one season of the GMKL was 25,769,000

Ugandan shillings (~£7,800). The actual costs were below budget and totalled 24,451,000

Ugandan shillings (~£7,400). These costs were exclusive of ground rehabilitation expenses in preparation for the following GMKL season.

APPENDIX A17: Reach of the Intervention (Assumptions)

The statistical methods used to assess whether the PF and MH of the adolescents who registered for the football league were different to their non-registered peers assumed that both groups were approximately normally distributed. This was initially assessed visually using a histogram of the eligible sample stratified according to registration for the crude results of each of the dependent variables (FIGURES A17.1 – A17.10).

FIGURE A17.1: Histograms of the distribution of the crude BFA results in the Gulu sample at M1 - stratified by gender and registration for the GMKL

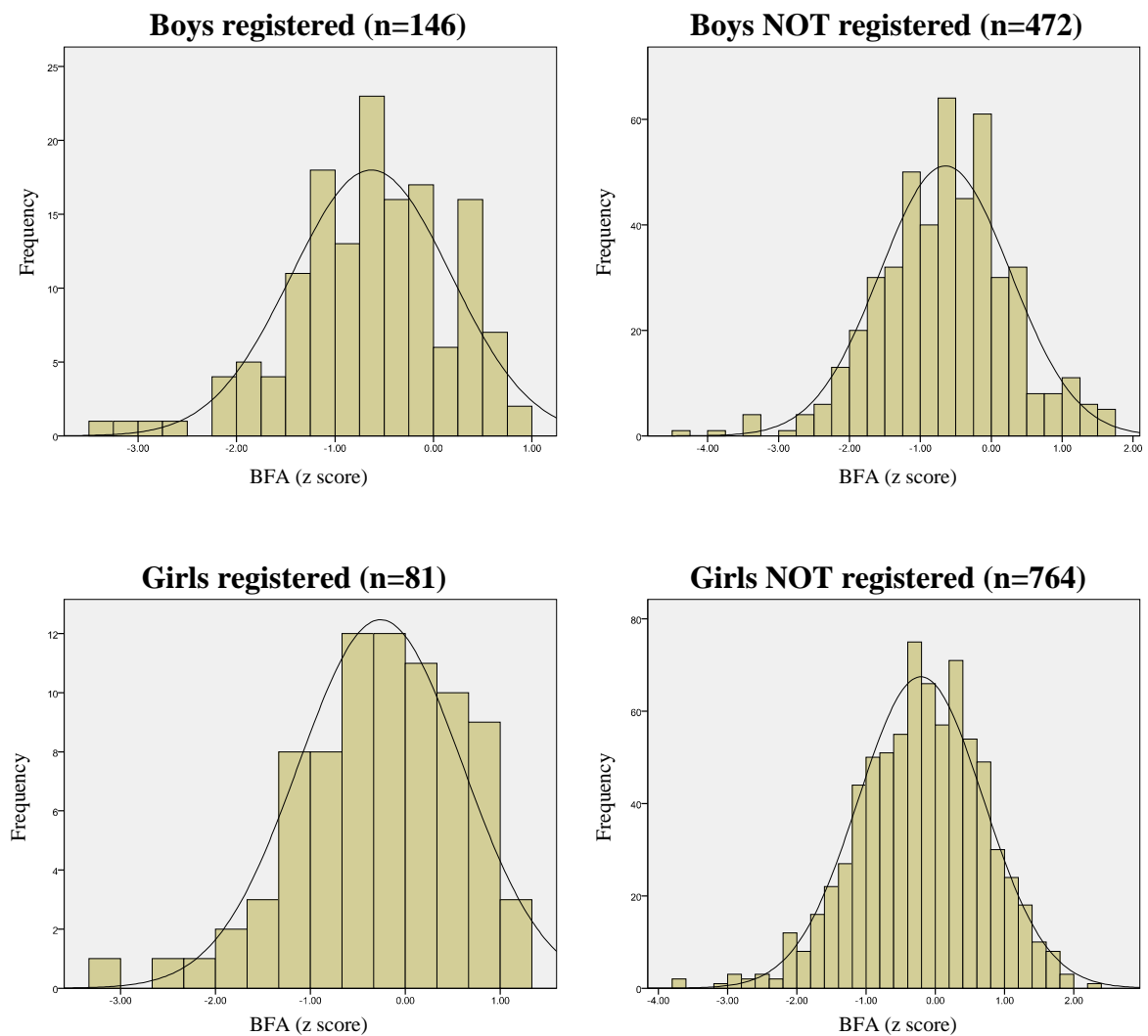


FIGURE A17.2: Histograms of the distribution of the crude HFA results in the Gulu sample at M1 - stratified by gender and registration for the GMKL

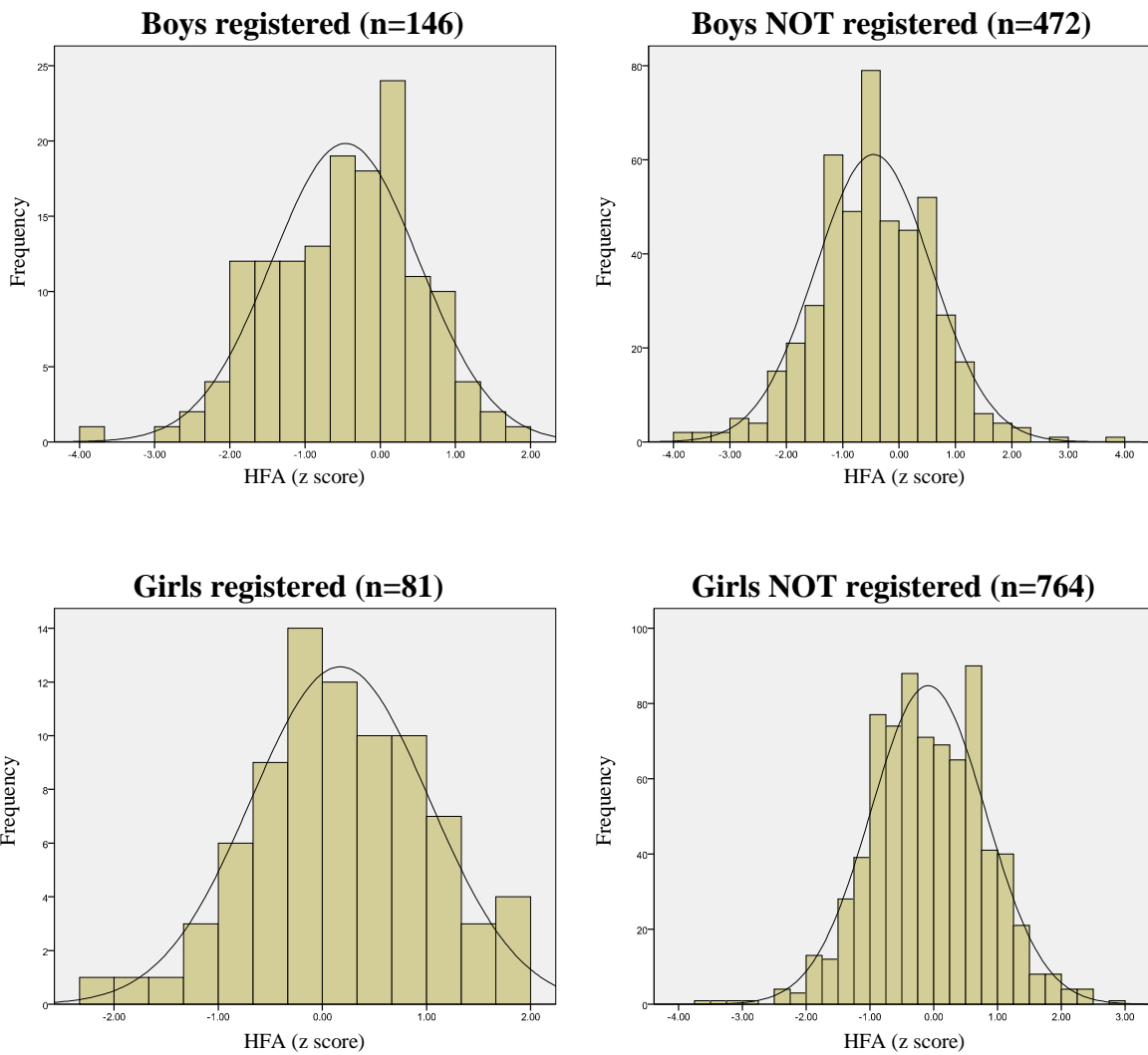
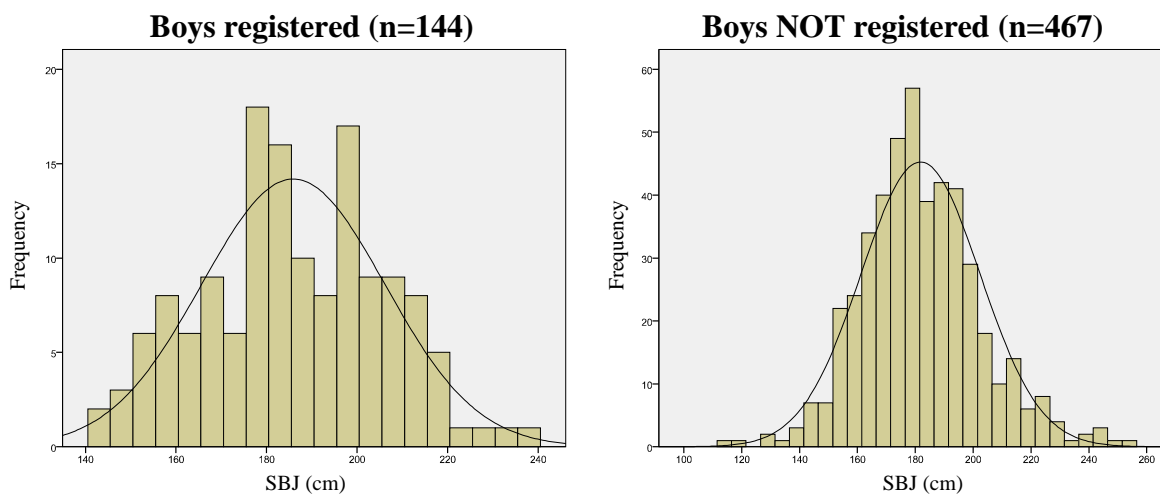


FIGURE A17.3: Histograms of the distribution of the crude SBJ results in the Gulu sample at M1 - stratified by gender and registration for the GMKL



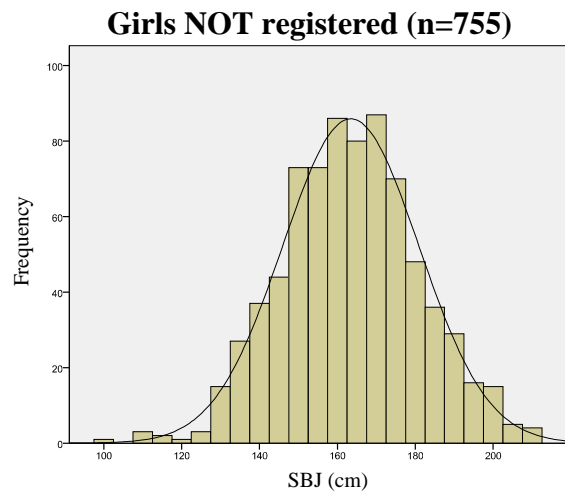
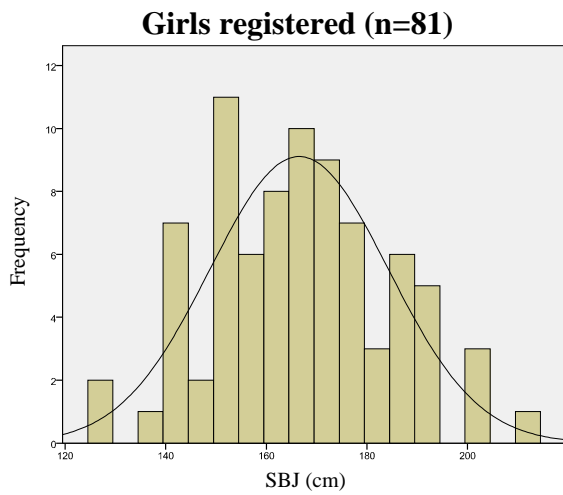


FIGURE A17.4: Histograms of the distribution of the crude MSFT results in the Gulu sample at M1 - stratified by gender and registration for the GMKL

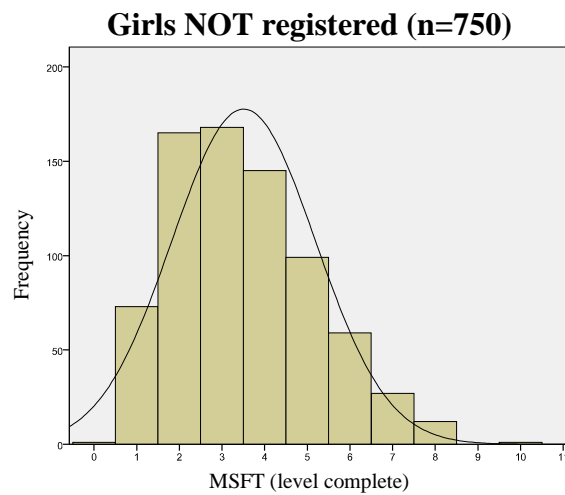
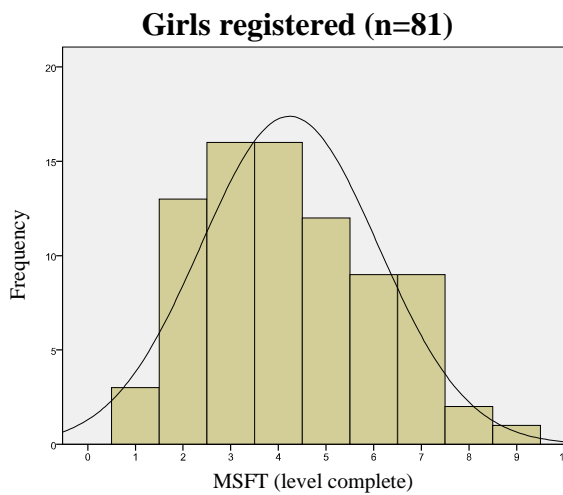
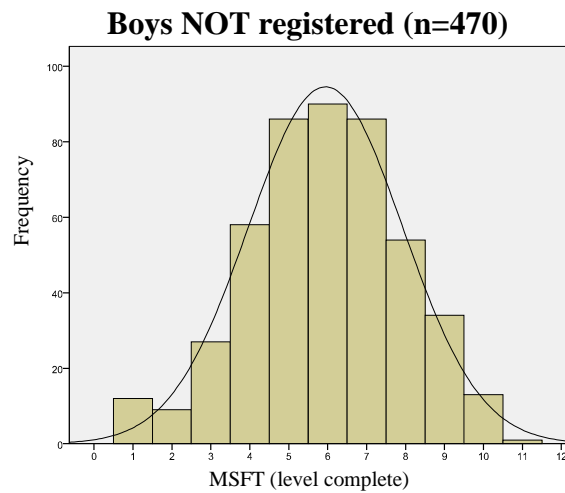
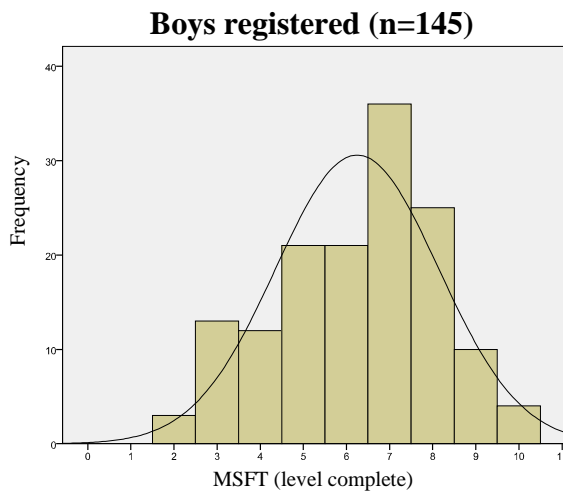


FIGURE A17.5: Histograms of the distribution of the crude TT results in the Gulu sample at M1 - stratified by gender and registration for the GMKL

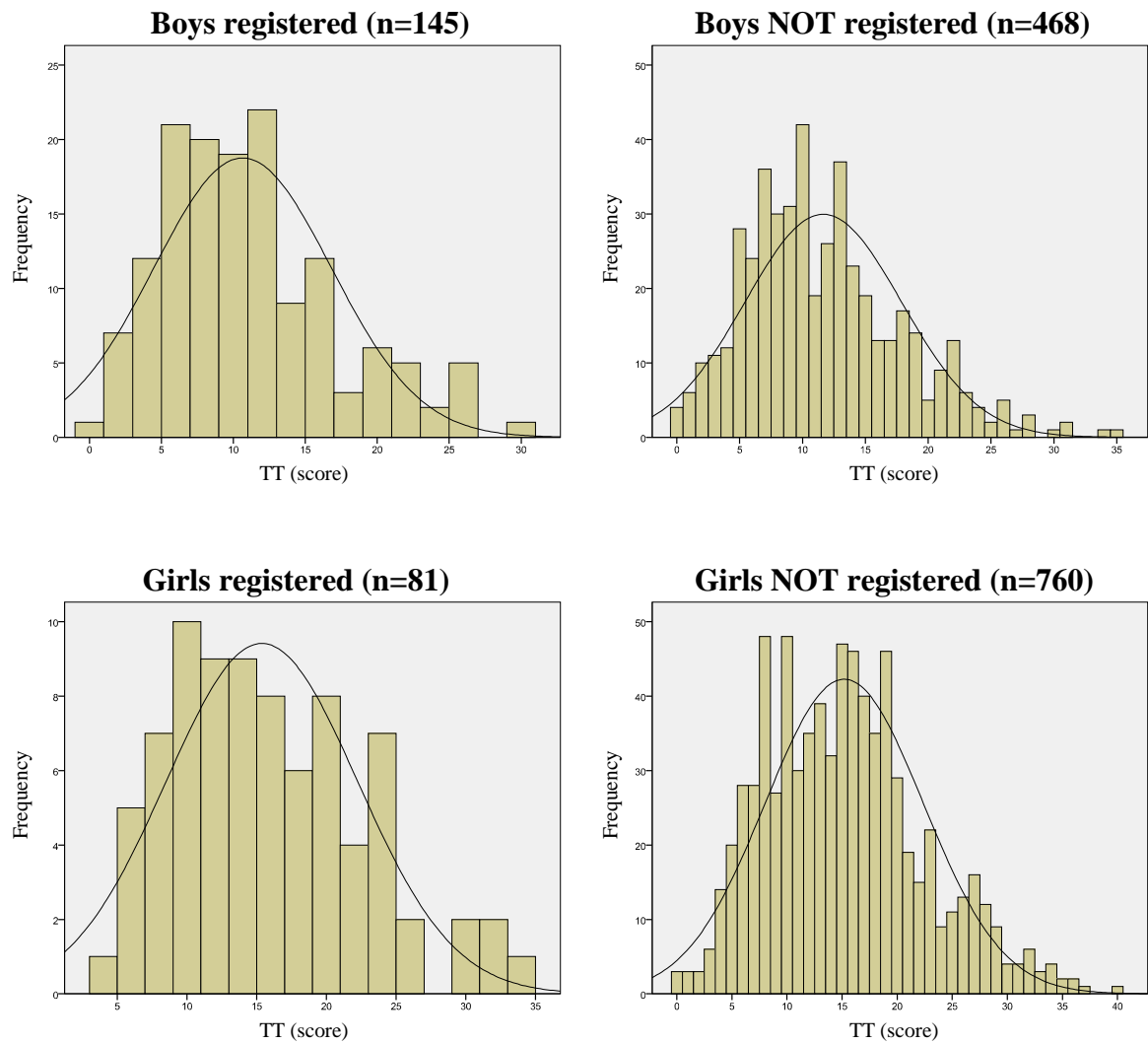
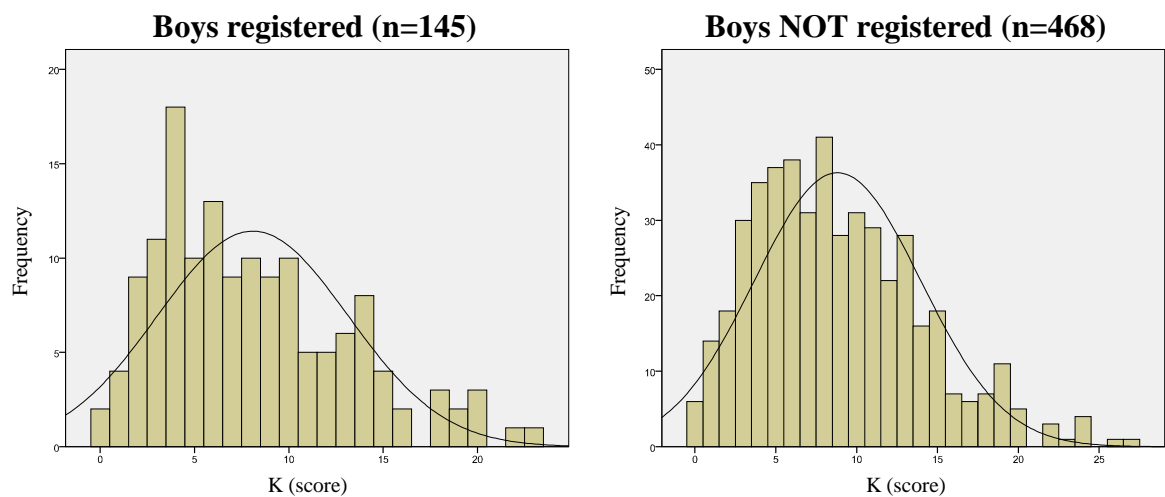


FIGURE A17.6: Histograms of the distribution of the crude K results in the Gulu sample at M1 - stratified by gender and registration for the GMKL



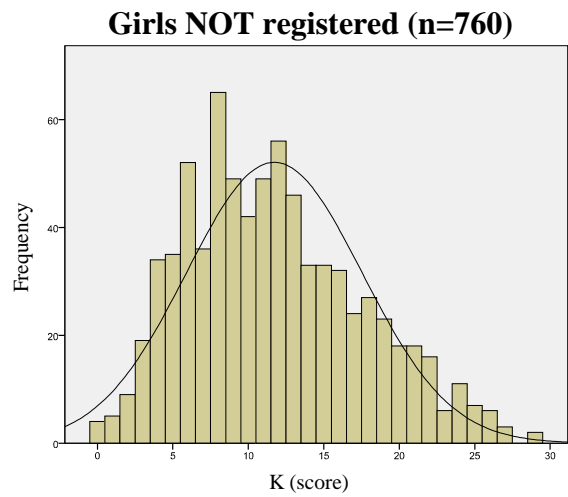
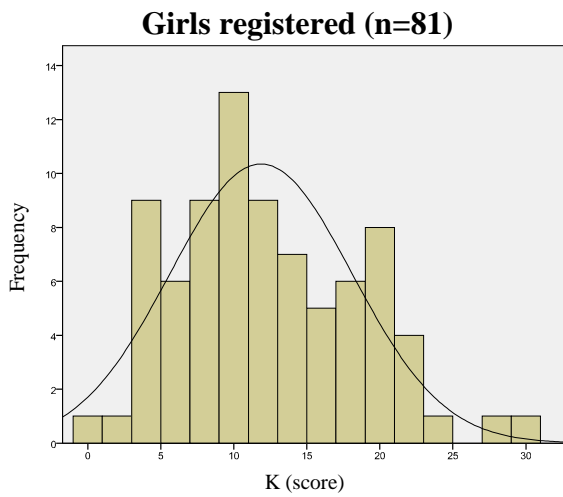


FIGURE A17.7: Histograms of the distribution of the crude P results in the Gulu sample at M1 - stratified by gender and registration for the GMKL

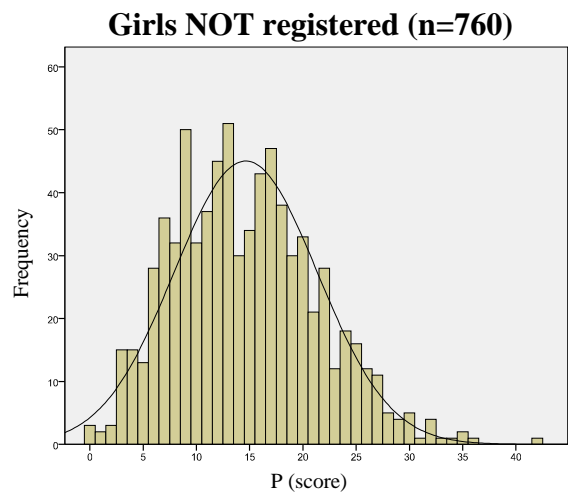
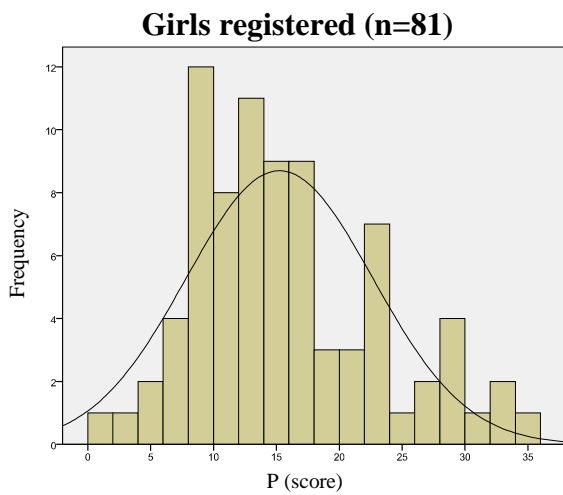
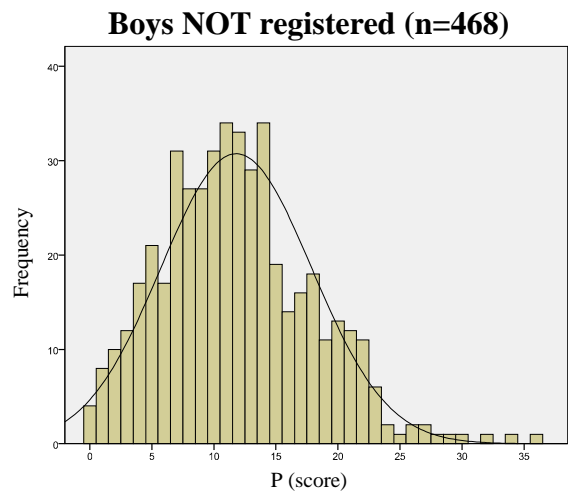
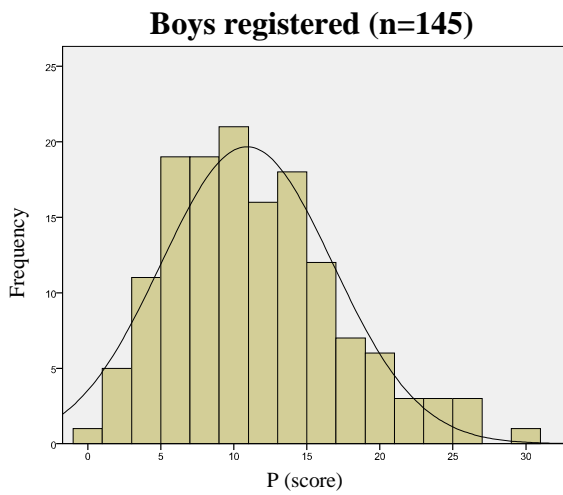


FIGURE A17.8: Histograms of the distribution of the crude ML results in the Gulu sample at M1 - stratified by gender and registration for the GMKL

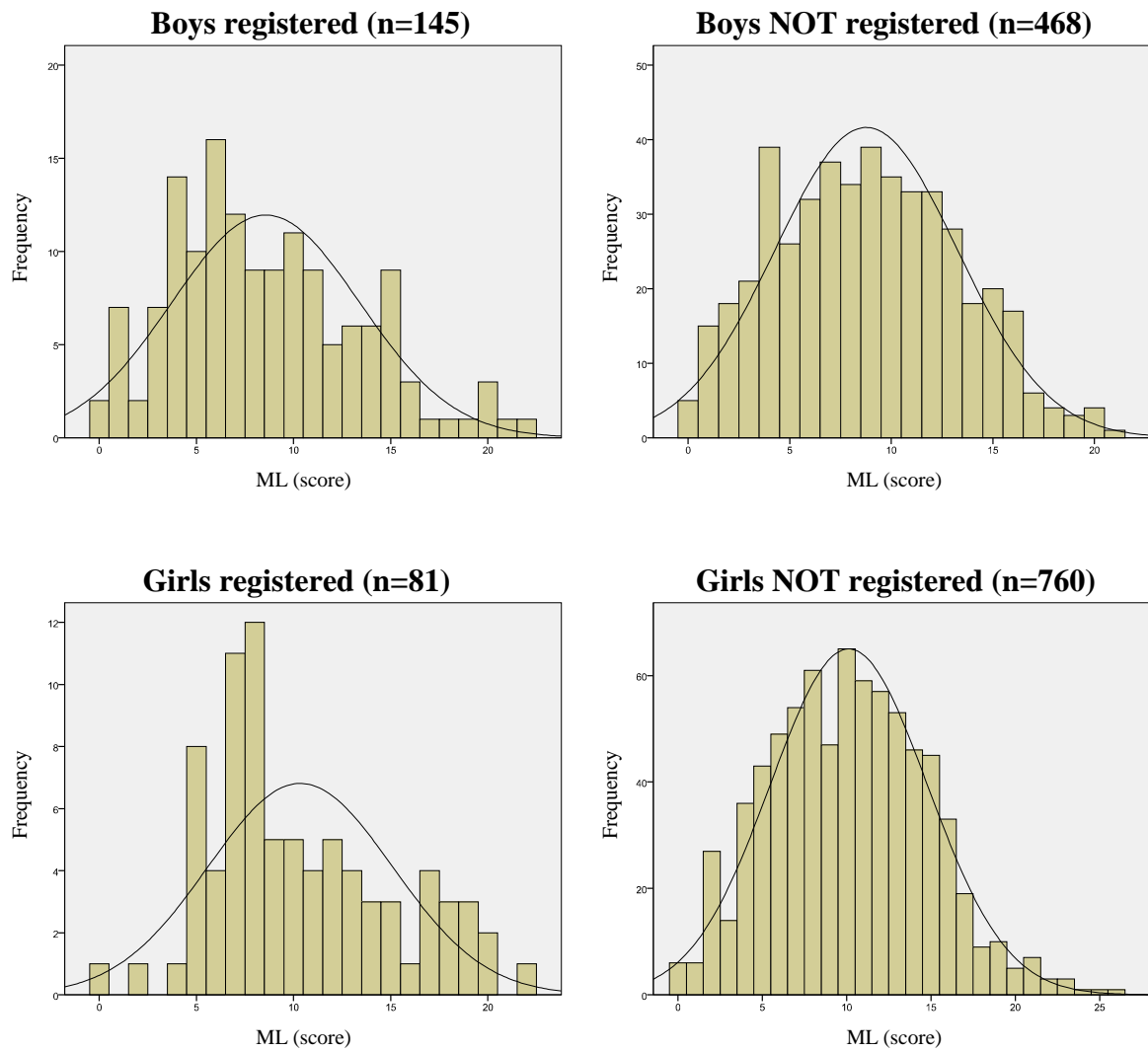
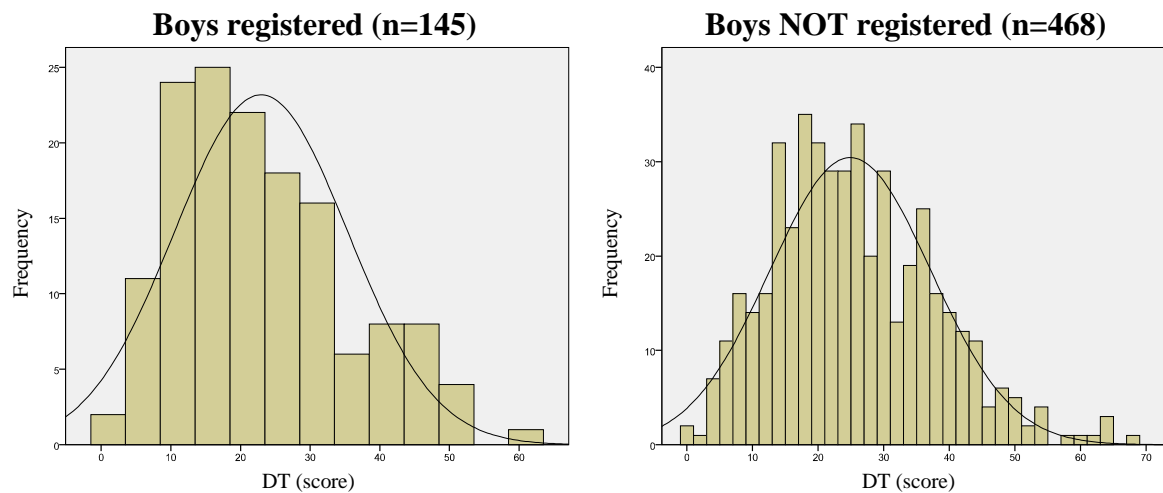


FIGURE A17.9: Histograms of the distribution of the crude DT results in the Gulu sample at M1 - stratified by gender and registration for the GMKL



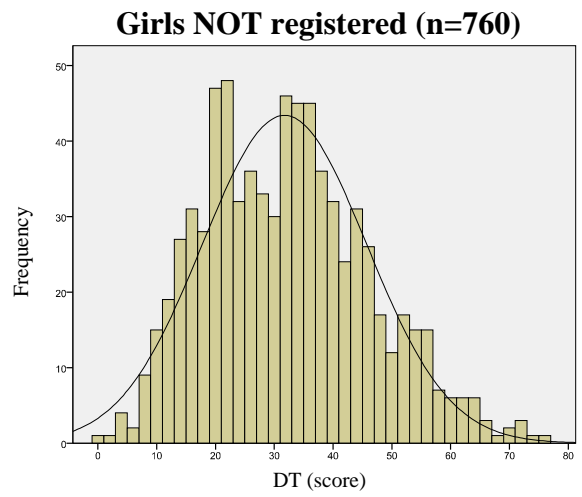
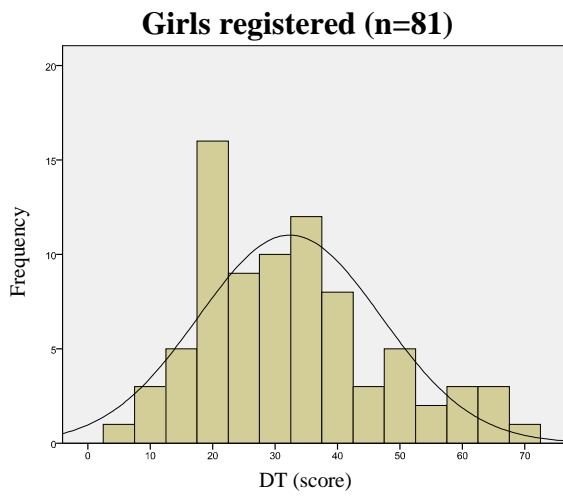
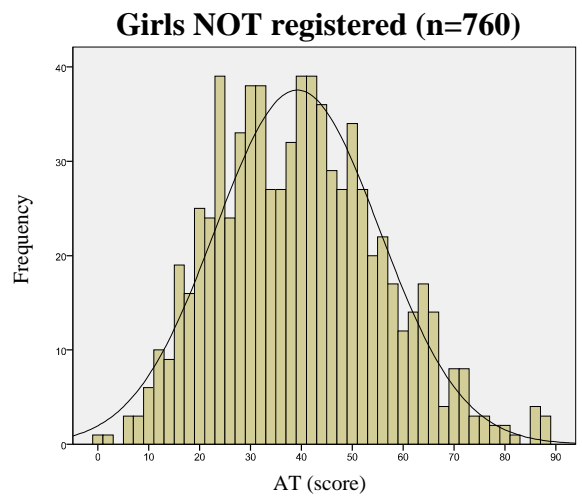
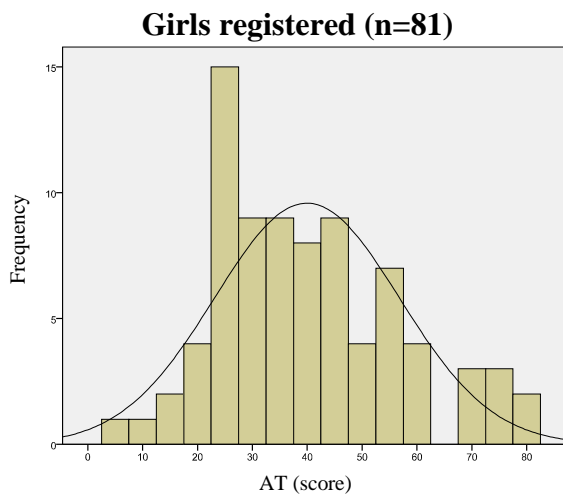
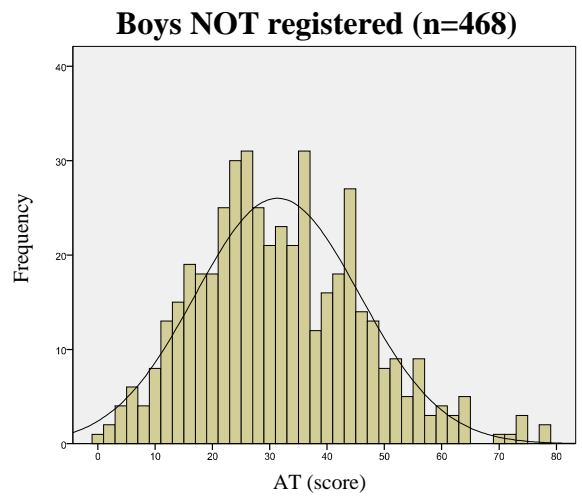
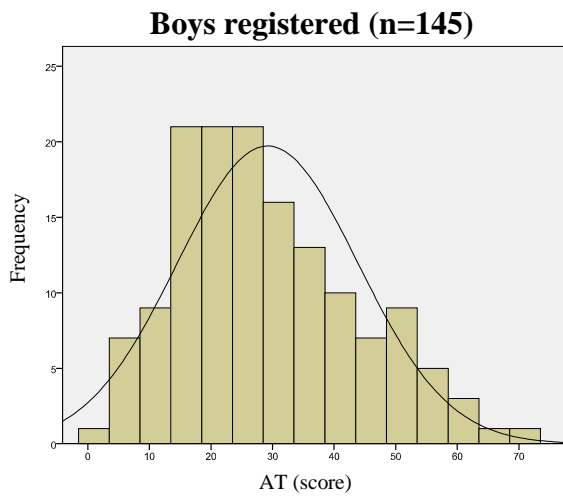


FIGURE A17.10: Histograms of the distribution of the crude AT results in the Gulu sample at M1 - stratified by gender and registration for the GMKL



Inspection of the histograms warranted further assessment of normality and statistical tests of the crude results were also completed (TABLES A17.1 & A17.2).

TABLE A17.1: Statistical tests for normality of the crude results for all of the outcomes for the boys in the Gulu sample at M1 – stratified by registration for the GMKL

Outcome variable	Registered for GMKL	n	Skew		Kurtosis		Shapiro Wilk	
			Stat	SES	Stat	SEK	Stat	p
BFA (z score)	Yes	146	-0.659	0.201	0.857	0.399	0.971	0.003
	No	472	-0.273	0.112	0.800	0.224	0.990	0.002
HFA (z score)	Yes	146	-0.330	0.201	0.036	0.399	0.988	0.247
	No	472	-0.064	0.112	0.822	0.224	0.992	0.016
SBJ (cm)	Yes	144	0.023	0.202	-0.509	0.401	0.990	0.378
	No	467	0.350	0.113	0.749	0.225	0.987	<0.001
MSFT (level complete)	Yes	145	-0.295	0.201	-0.576	0.400	0.954	<0.001
	No	470	-0.180	0.113	-0.159	0.225	0.972	<0.001
TT (score)	Yes	145	0.736	0.201	0.089	0.400	0.952	<0.001
	No	468	0.732	0.113	0.505	0.225	0.964	<0.001
K (score)	Yes	145	0.767	0.201	0.032	0.400	0.943	<0.001
	No	468	0.707	0.113	0.300	0.225	0.961	<0.001
P (score)	Yes	145	0.669	0.201	0.220	0.400	0.965	<0.001
	No	468	0.564	0.113	0.509	0.225	0.977	<0.001
ML (score)	Yes	145	0.542	0.201	-0.183	0.400	0.966	0.001
	No	468	0.196	0.113	-0.594	0.225	0.983	<0.001
DT (score)	Yes	145	0.673	0.201	-0.126	0.400	0.956	<0.001
	No	468	0.553	0.113	0.205	0.225	0.978	<0.001
AT (score)	Yes	145	0.559	0.201	-0.178	0.400	0.970	0.003
	No	468	0.456	0.113	0.102	0.225	0.985	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A17.2: Statistical tests for normality of the crude results for all of the outcomes for the girls in the Gulu sample at M1 – stratified by registration for the GMKL

Outcome Variable	Registered for GMKL	n	Skew		Kurtosis		Shapiro Wilk	
			Stat	SES	Stat	SEK	Stat	p
BFA (z score)	Yes	81	-0.687	0.267	0.749	0.529	0.965	0.028
	No	764	-0.364	0.088	0.369	0.177	0.992	<0.001
HFA (z score)	Yes	81	-0.206	0.267	-0.044	0.529	0.990	0.779
	No	764	-0.077	0.088	0.390	0.177	0.995	0.010
SBJ (cm)	Yes	81	0.174	0.267	-0.313	0.529	0.992	0.894
	No	755	-0.026	0.089	0.079	0.178	0.997	0.261
MSFT (level complete)	Yes	81	0.366	0.267	-0.607	0.529	0.951	0.004
	No	750	0.557	0.089	-0.111	0.178	0.941	<0.001
TT (score)	Yes	81	0.523	0.267	-0.319	0.529	0.966	0.028
	No	760	0.516	0.089	-0.024	0.177	0.977	<0.001
K (score)	Yes	81	0.463	0.267	-0.254	0.529	0.970	0.056
	No	760	0.477	0.089	-0.348	0.177	0.974	<0.001
P (score)	Yes	81	0.677	0.267	-0.031	0.529	0.953	0.005
	No	760	0.435	0.089	0.019	0.177	0.984	<0.001
ML (score)	Yes	81	0.541	0.267	-0.412	0.529	0.948	0.002
	No	760	0.246	0.089	-0.164	0.177	0.988	<0.001
DT (score)	Yes	81	0.591	0.267	-0.136	0.529	0.962	0.017
	No	760	0.406	0.089	-0.246	0.177	0.984	<0.001
AT (score)	Yes	81	0.576	0.267	-0.139	0.529	0.960	0.013
	No	760	0.340	0.089	-0.224	0.177	0.989	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12].

The assumption of HOV between registered vs. NOT registered for the crude results of each of the dependent variables was assessed using the Levene and Brown-Forsythe tests (TABLE A17.3). These statistical tests for HOV were robust to violations of the normality assumption ^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value

was adjusted accordingly ($p < 0.01$).

TABLE A17.3: Statistical tests of the HOV between registered and NOT registered adolescents for all of the crude outcomes in the Gulu sample at M1 - stratified by gender

Outcome variable	Gender	Registered		Not registered		Levene test		Brown-Forsythe test		Summary
		n	Variance	n	Variance	F	p	F	p	
BFA (z score)	Boys	146	0.655	472	0.846	2.640	0.105	2.793	0.095	√
	Girls	81	0.746	764	0.817	0.205	0.651	0.222	0.637	√
HFA (z score)	Boys	146	0.958	472	1.054	0.015	0.902	0.031	0.861	√
	Girls	81	0.735	764	0.809	0.214	0.644	0.207	0.649	√
SBJ (cm)	Boys	144	410.03	467	423.56	0.289	0.591	0.309	0.579	√
	Girls	81	314.63	755	307.28	0.057	0.812	0.049	0.824	√
MSFT (level complete)	Boys	145	3.577	470	3.933	0.001	0.974	0.027	0.870	√
	Girls	81	3.457	750	2.838	1.751	0.186	1.259	0.262	√
TT (score)	Boys	145	38.02	468	38.84	0.019	0.891	0.028	0.867	√
	Girls	81	47.11	760	51.43	0.055	0.815	0.094	0.759	√
K (score)	Boys	145	25.64	468	26.47	0.009	0.926	0.013	0.908	√
	Girls	81	38.97	760	33.91	0.894	0.345	0.755	0.385	√
P (score)	Boys	145	34.60	468	36.92	0.144	0.705	0.194	0.660	√
	Girls	81	55.21	760	45.33	0.640	0.424	0.355	0.551	√
ML (score)	Boys	145	23.43	468	20.11	1.026	0.311	0.671	0.413	√
	Girls	81	22.50	760	21.76	0.179	0.673	0.003	0.956	√
DT (score)	Boys	145	155.43	468	150.58	0.140	0.709	0.043	0.836	√
	Girls	81	214.91	760	195.11	0.107	0.744	0.082	0.775	√
AT (score)	Boys	145	215.12	468	206.07	0.162	0.687	0.045	0.832	√
	Girls	81	284.35	760	260.87	0.193	0.661	0.083	0.773	√

NOTE: Results in ***BOLD ITALICS*** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

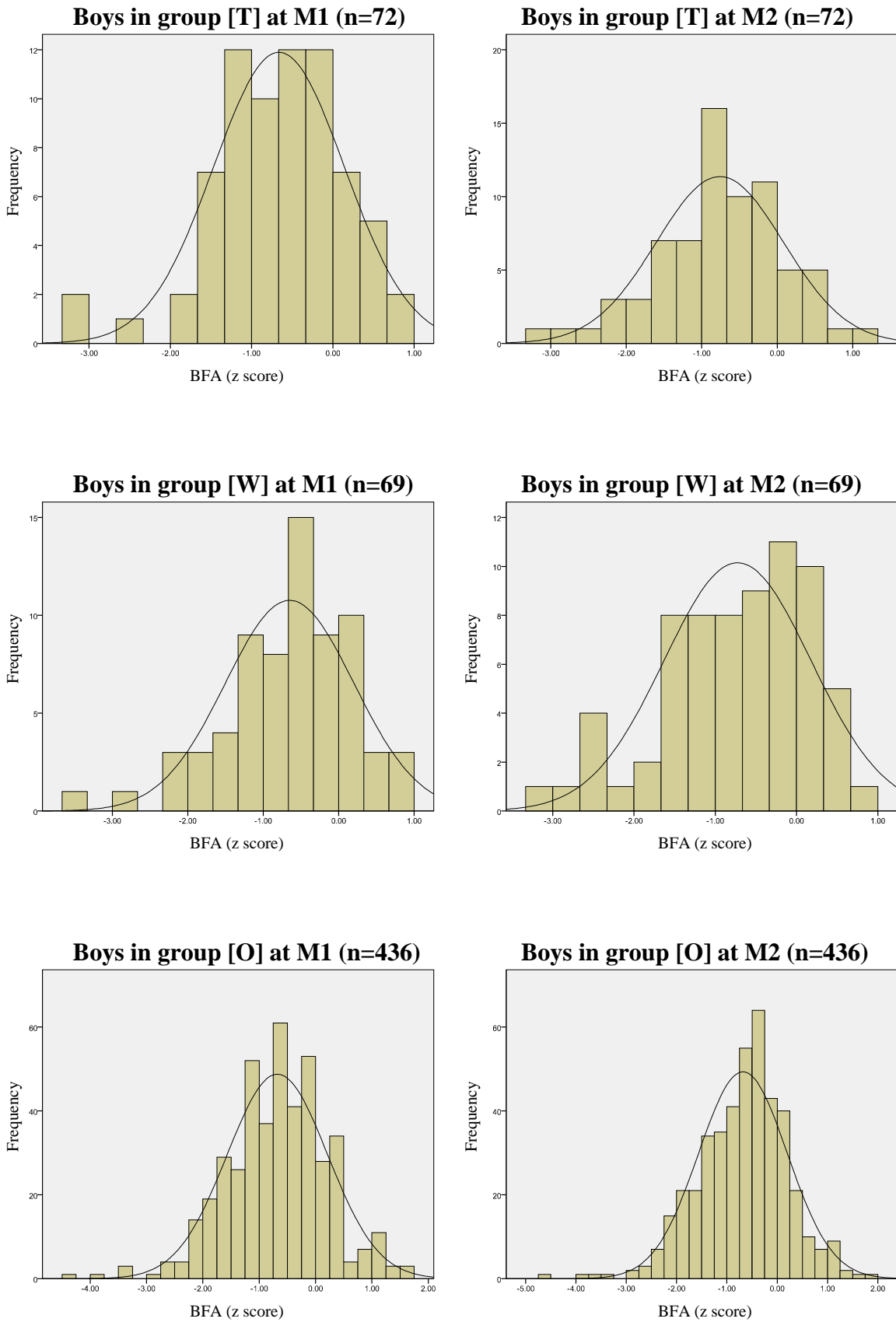
NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

The assumption of HOV held for all of the dependent variables when comparing the outcomes for those registered vs. NOT registered. The statistical analysis could proceed with little concern when the HOV assumption was satisfied ^[12,14].

APPENDIX A18: Impact of the Intervention – Within-Group Differences (Assumptions)

The statistical methods used to assess whether there was any change over time in PF and MH of the adolescents in each of the experiment groups assumed an approximately normal distribution at M1 and M2. This was initially assessed visually using a histogram of the eligible sample at each time point for the crude results of each of the dependent variables (FIGURES A18.1 – A18.9).

FIGURE A18.1: Histograms of the distribution for the crude BFA results in the Gulu sample at M1 and M2 – stratified by gender and intervention group



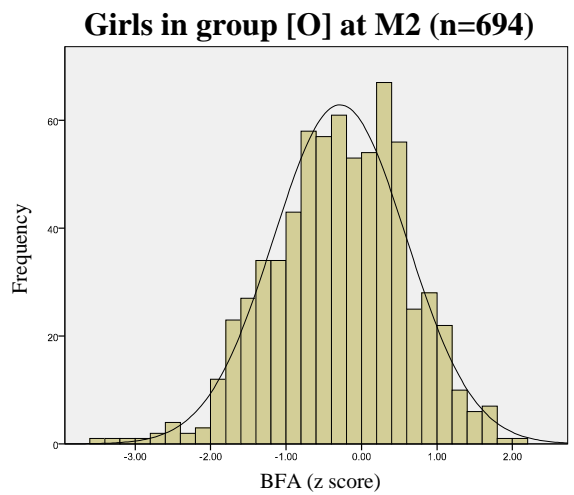
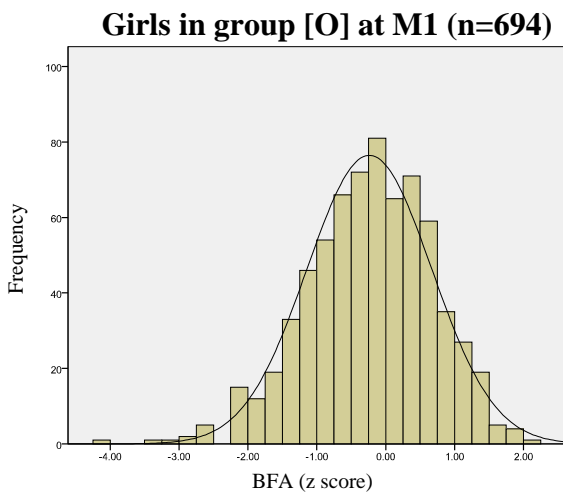
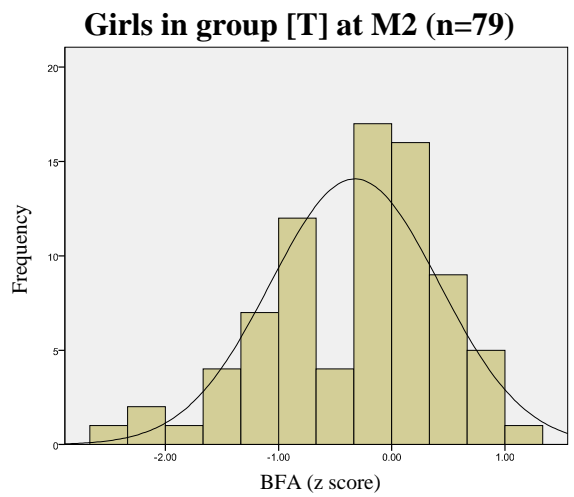
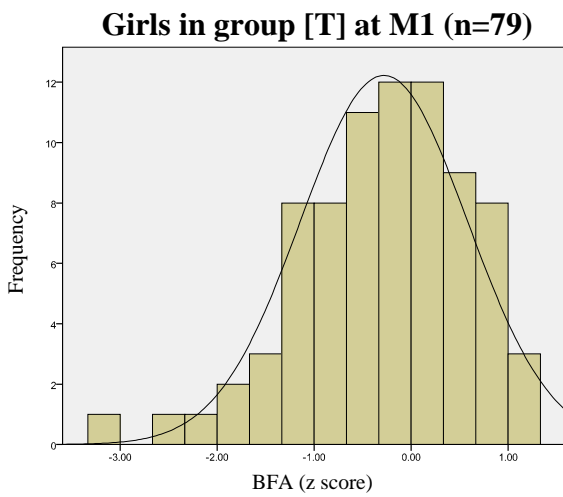
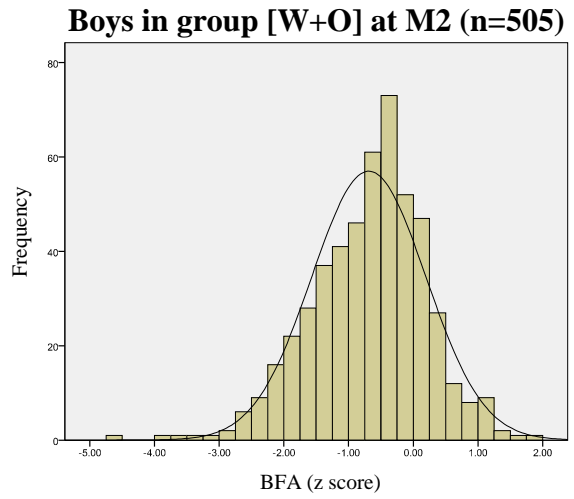
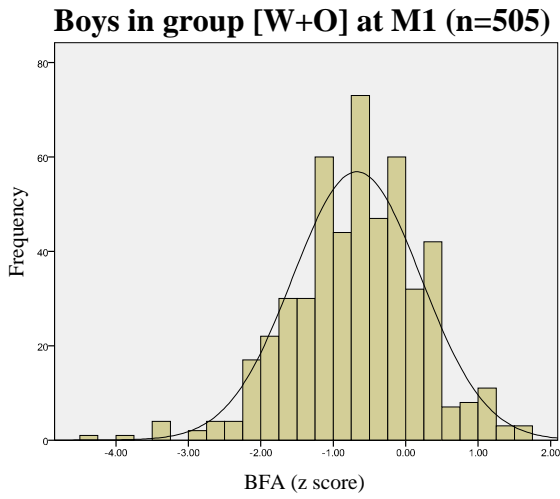
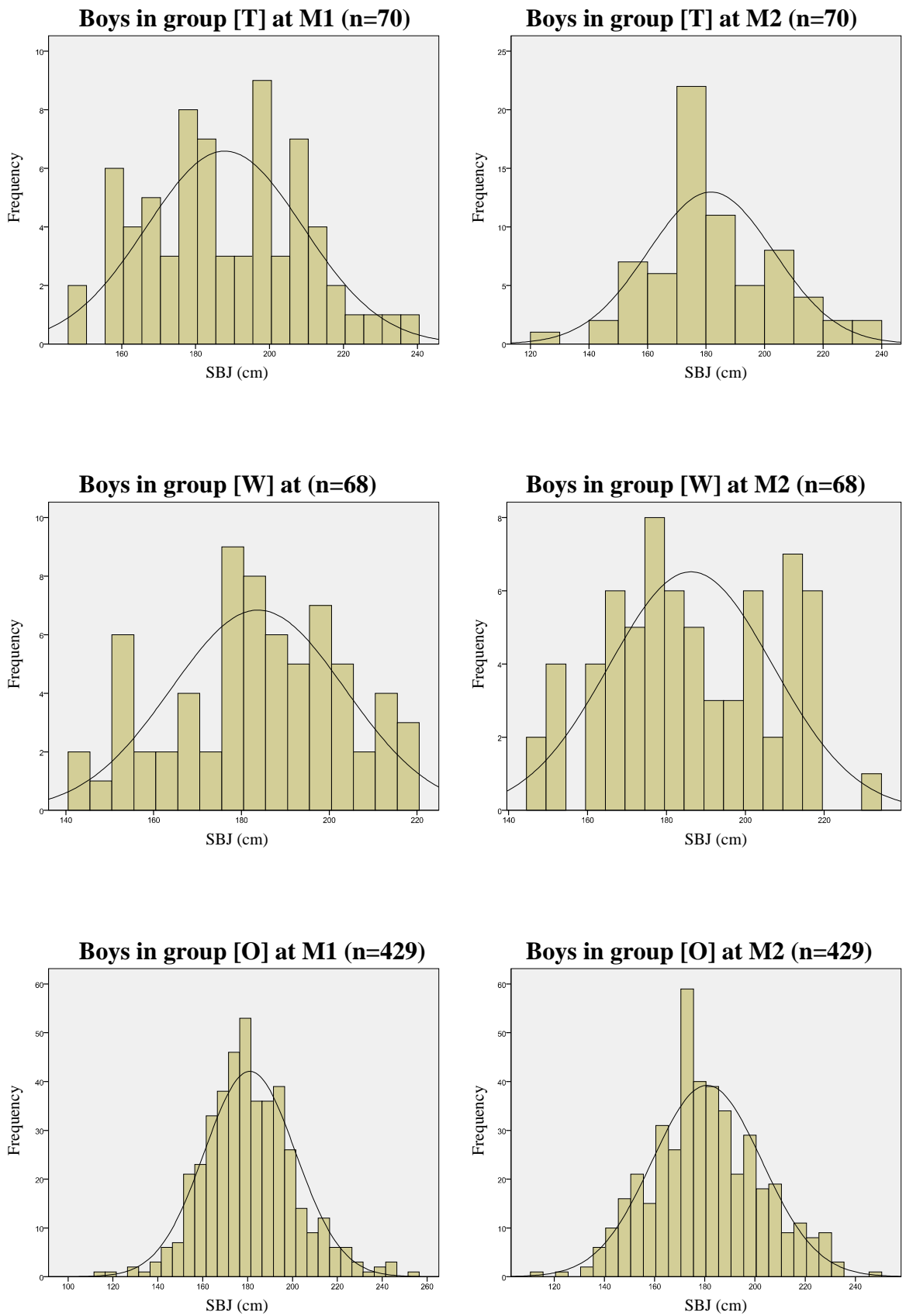


FIGURE A18.2: Histograms of the distribution for the crude SBJ results in the Gulu sample at M1 and M2 – stratified by gender and intervention group



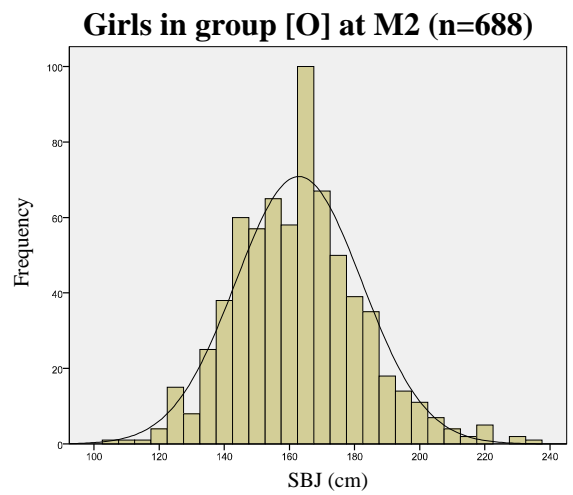
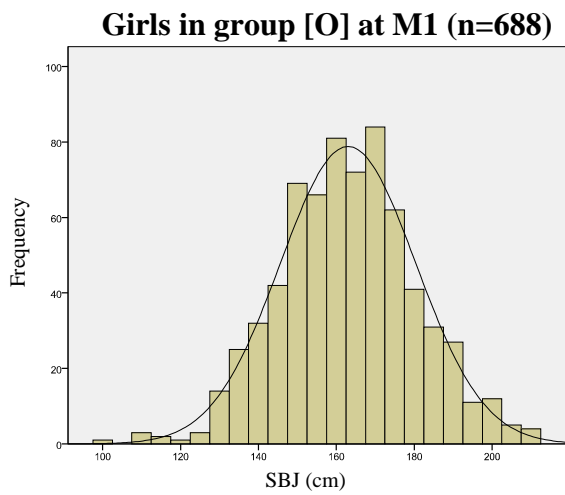
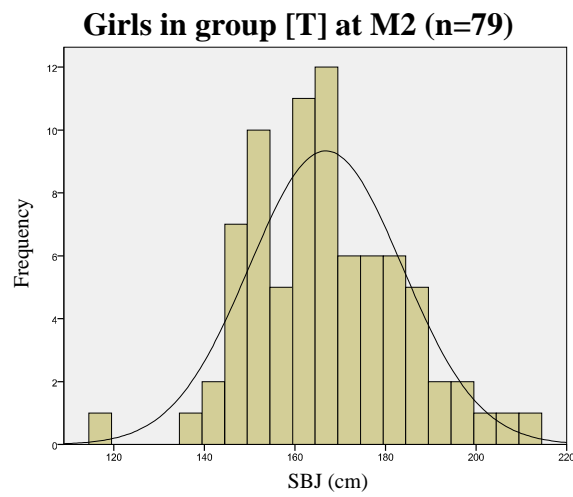
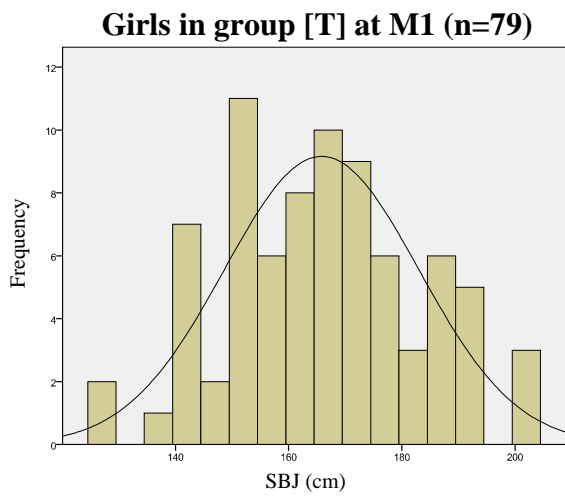
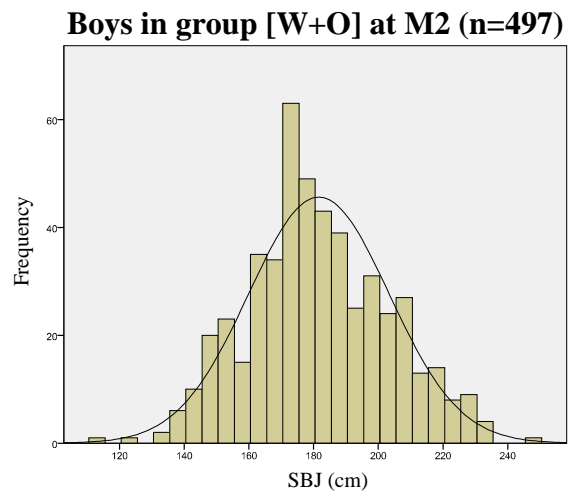
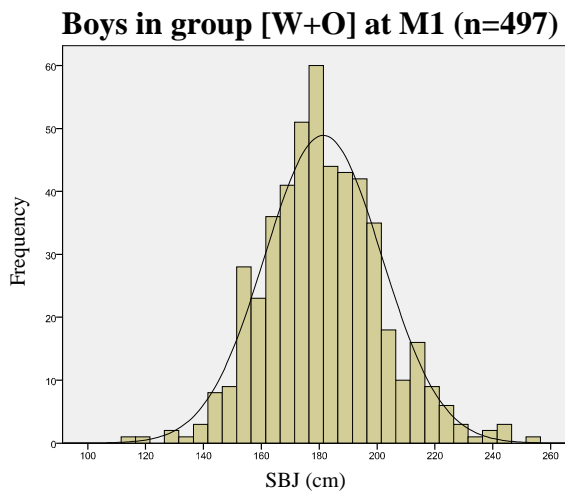
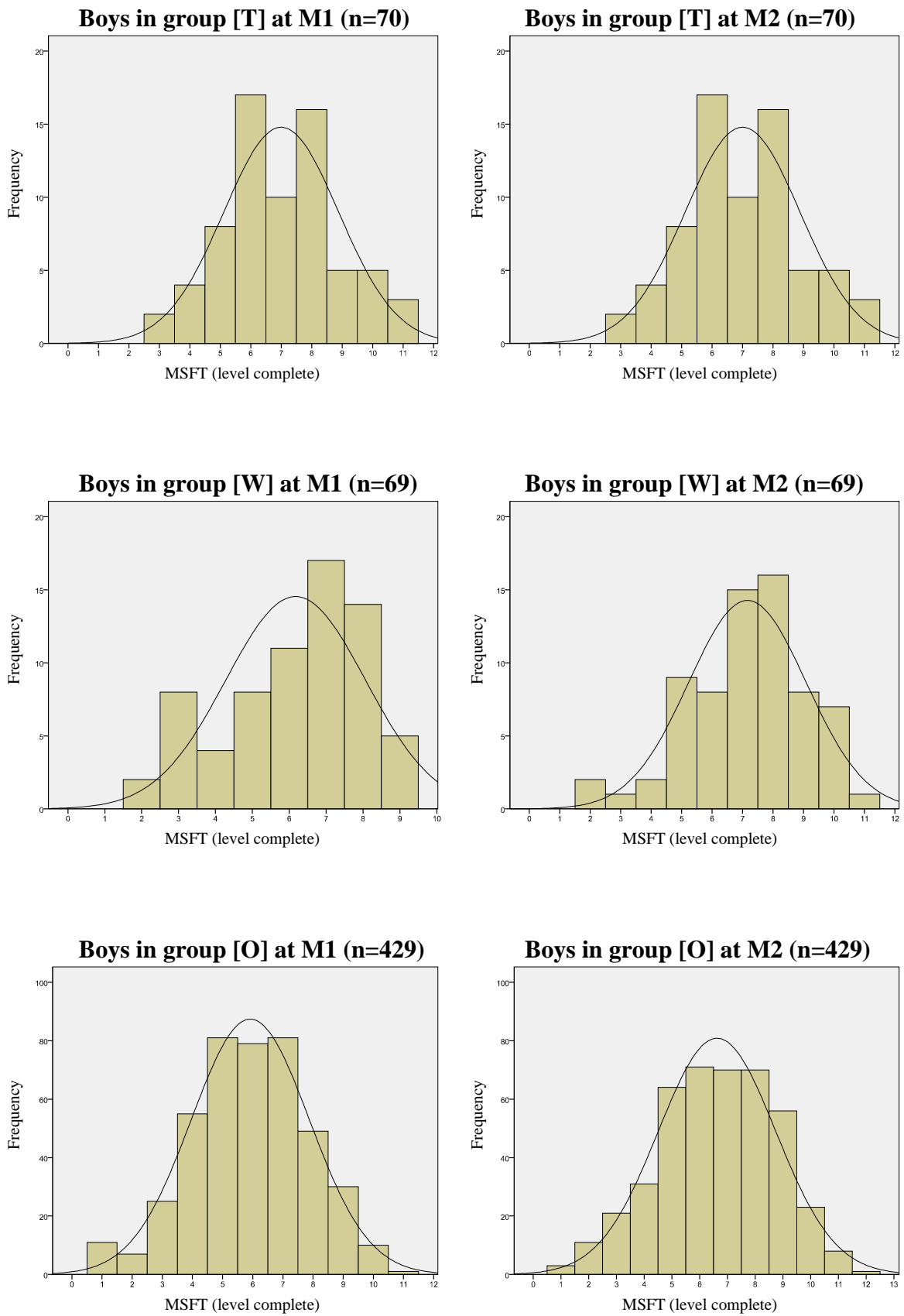
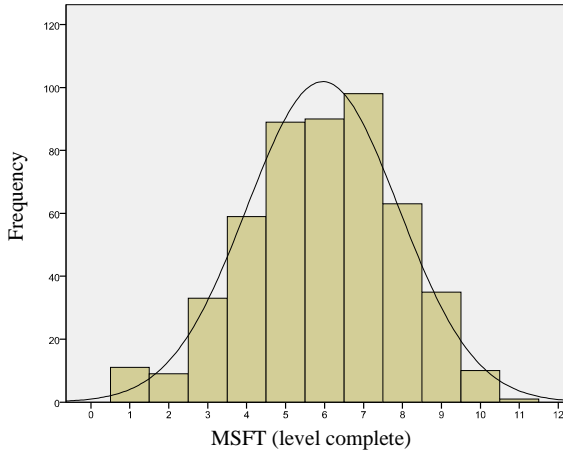


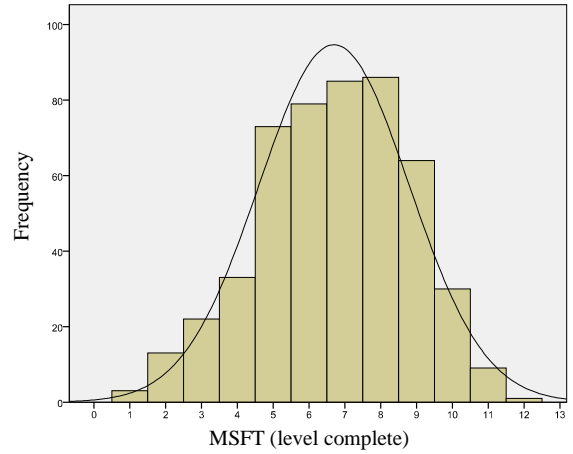
FIGURE A18.3: Histograms of the distribution for the crude MSFT results in the Gulu sample at M1 and M2 – stratified by gender and intervention group



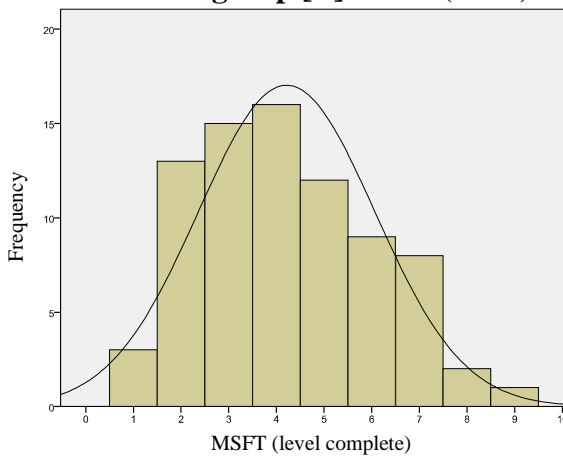
Boys in group [W+O] at M1 (n=498)



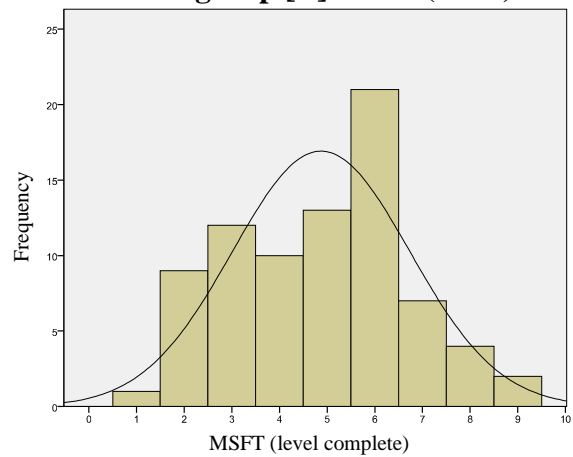
Boys in group [W+O] at M2 (n=498)



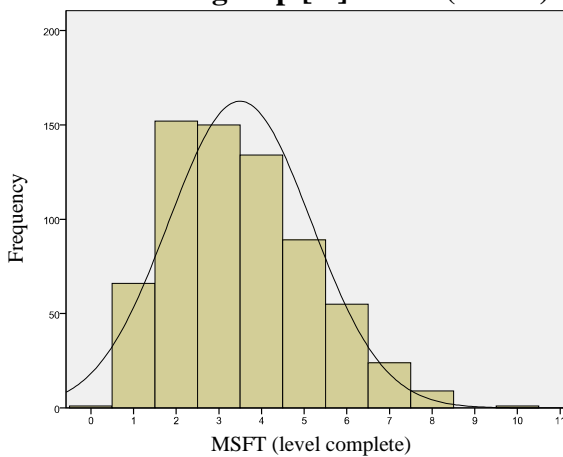
Girls in group [T] at M1 (n=79)



Girls in group [T] at M2 (n=79)



Girls in group [O] at M1 (n=681)



Girls in group [O] at M2 (n=681)

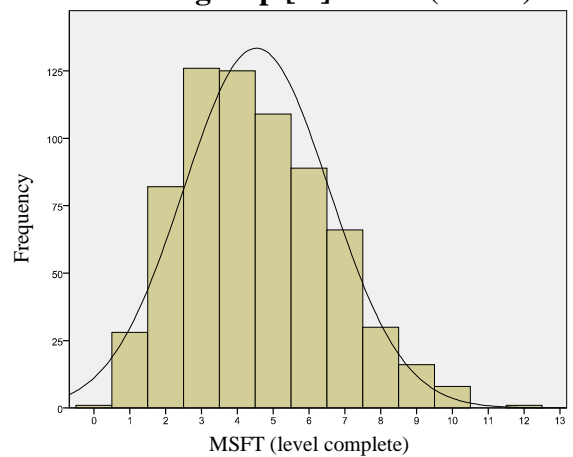
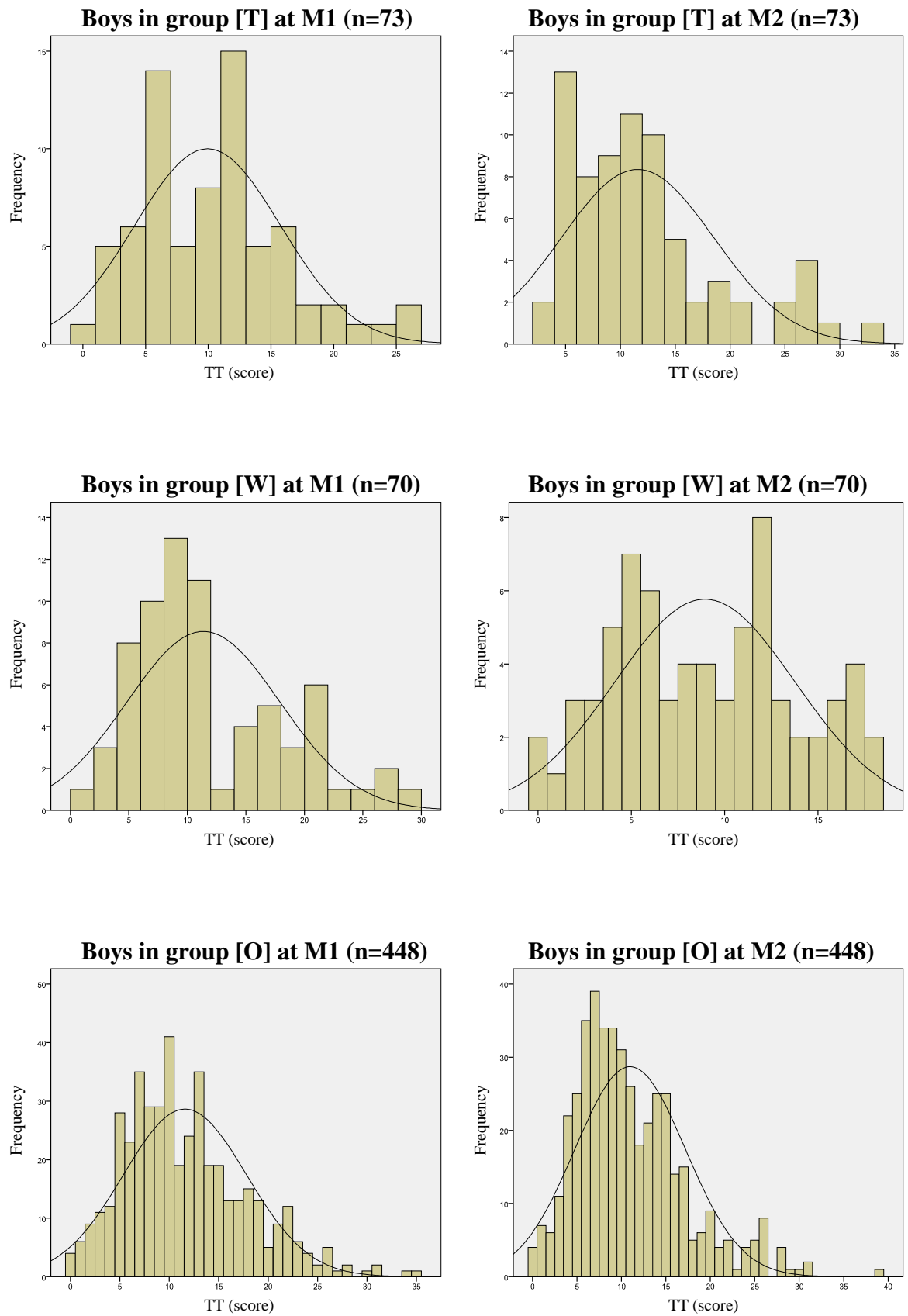


FIGURE A18.4: Histograms of the distribution for the crude TT results at M1 and M2 in the Gulu sample – stratified by gender and intervention group



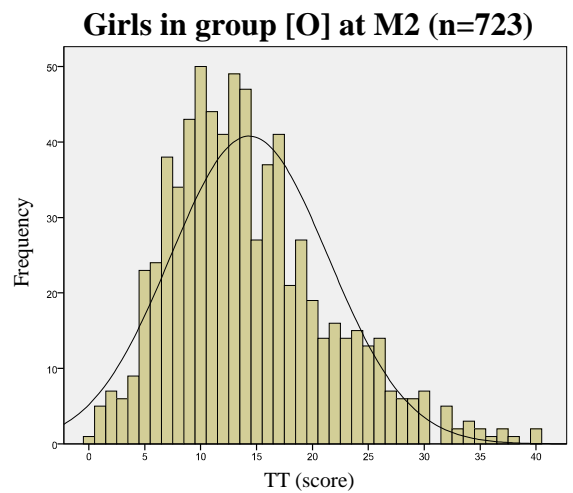
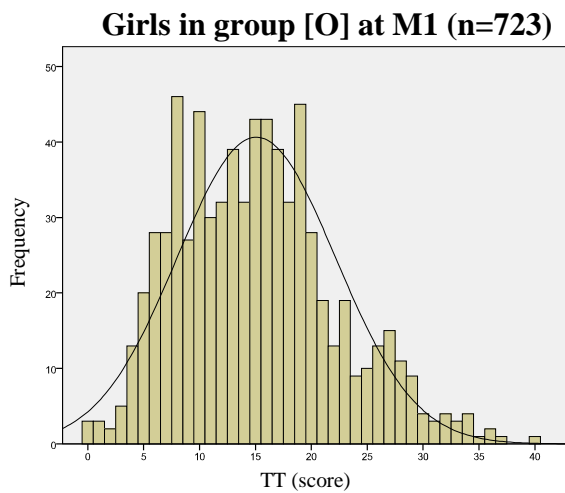
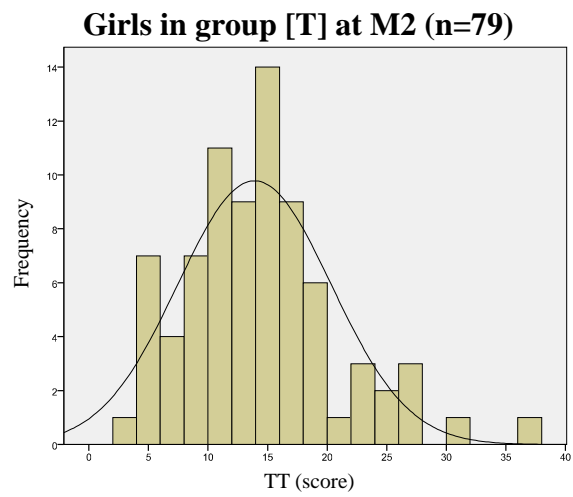
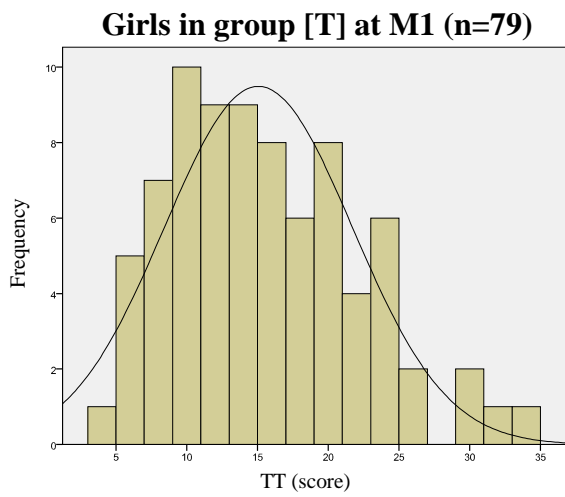
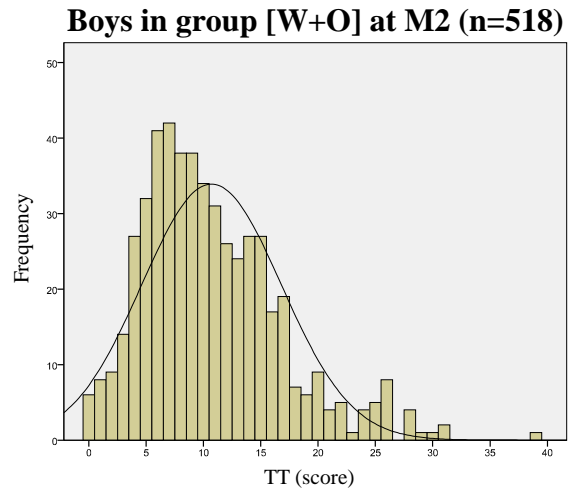
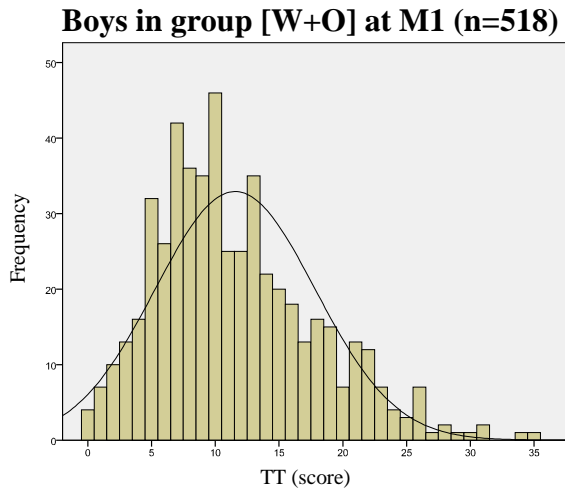
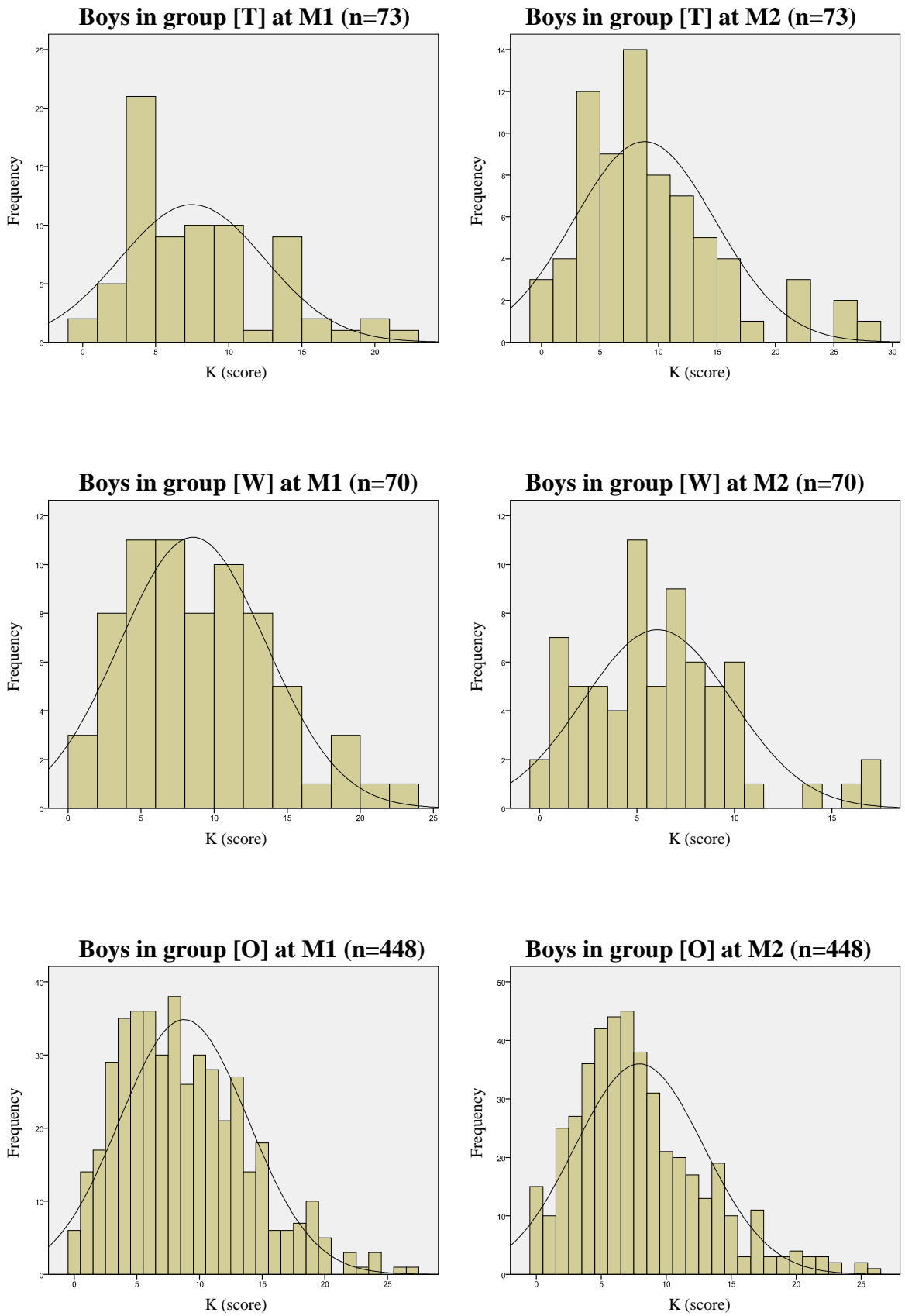


FIGURE A18.5: Histograms of the distribution for the crude K results in the Gulu sample at M1 and M2 – stratified by gender and intervention group



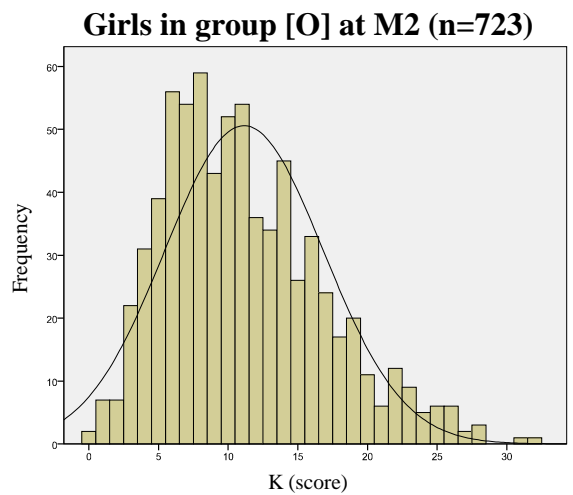
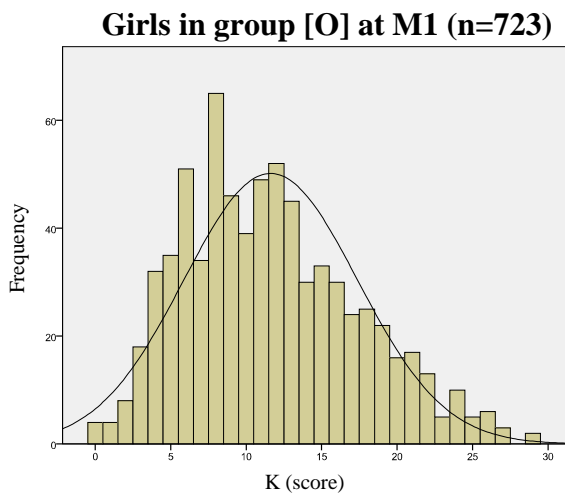
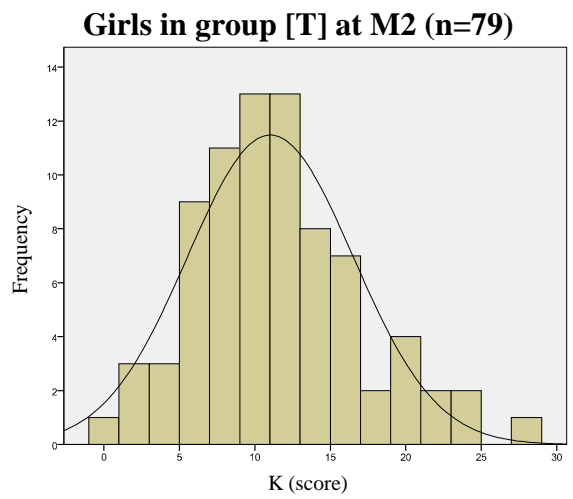
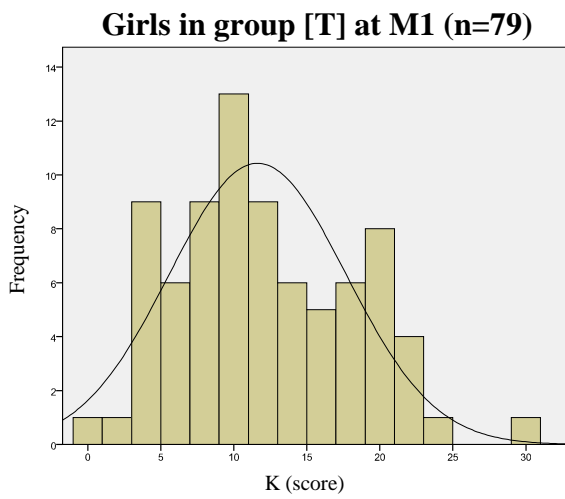
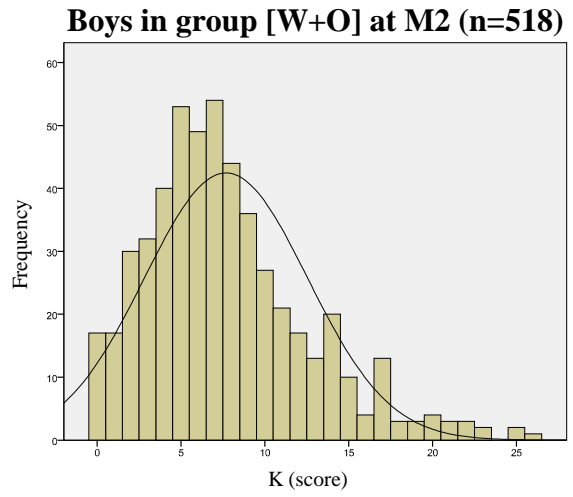
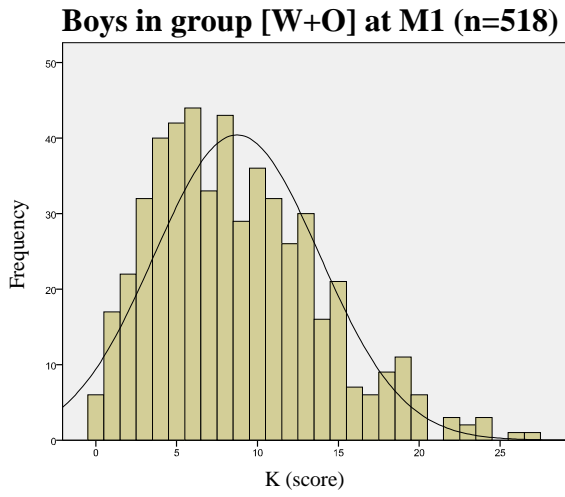
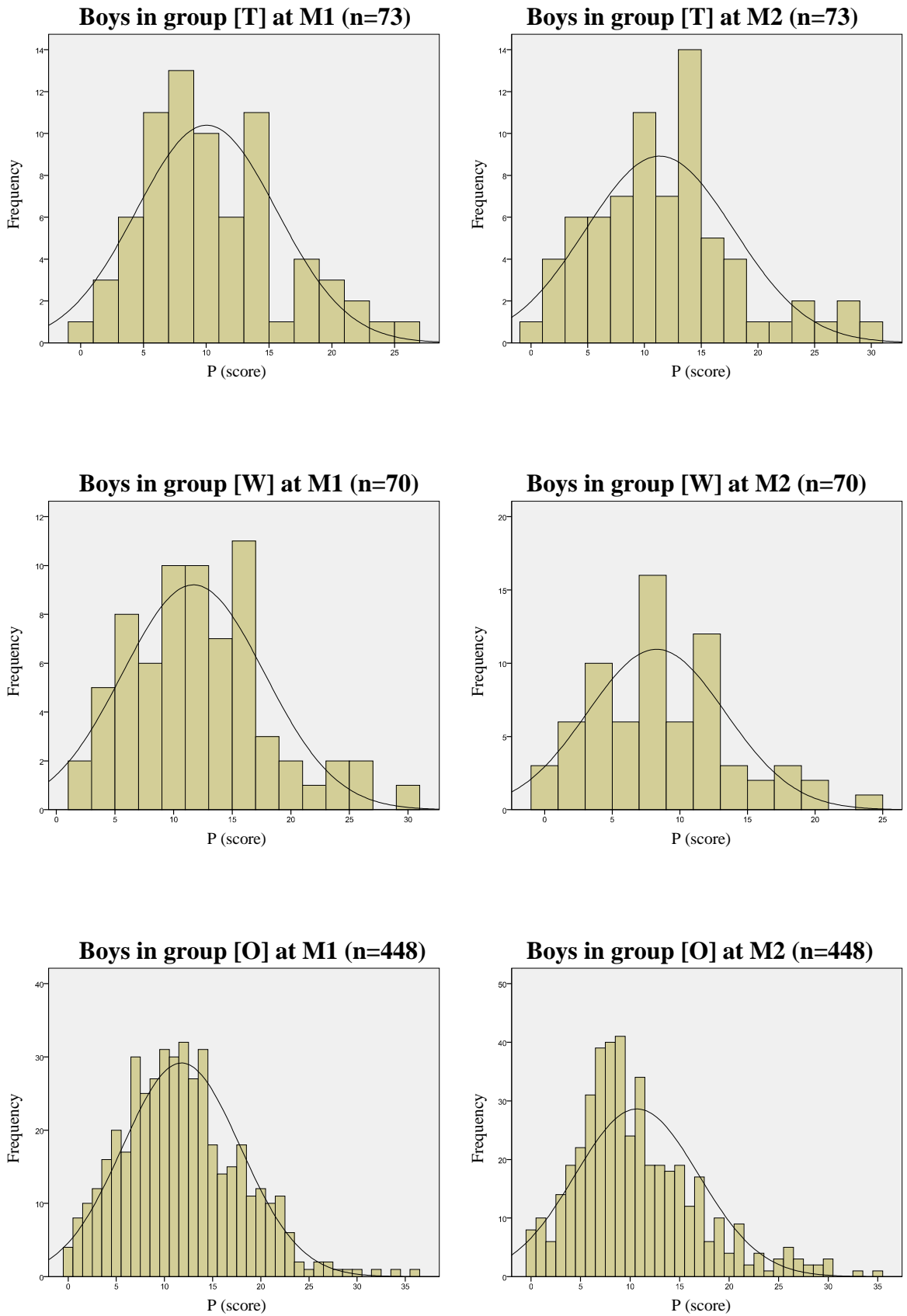
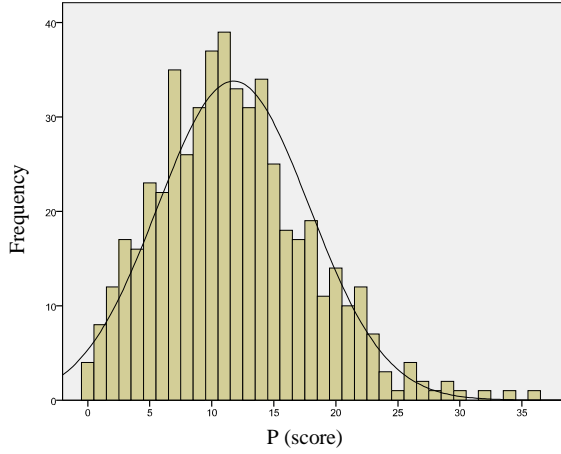


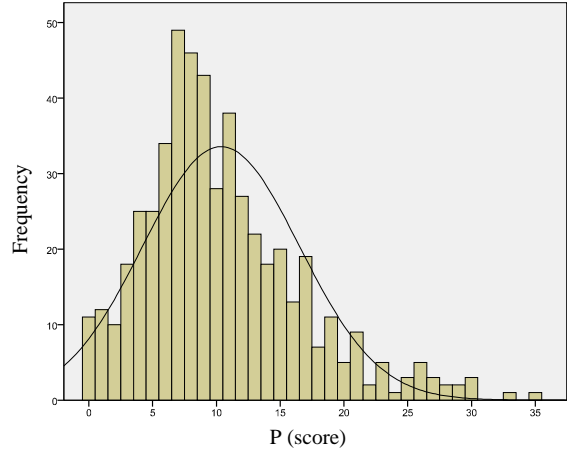
FIGURE A18.6: Histograms of the distribution for the crude P results in the Gulu sample at M1 and M2 – stratified by gender and intervention group



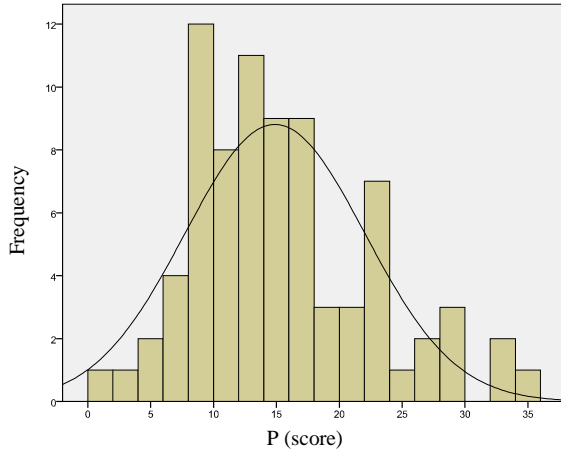
Boys in group [W+O] at M1 (n=518)



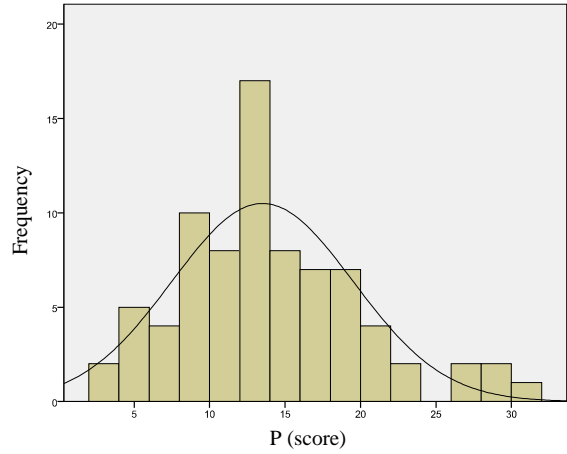
Boys in group [W+O] at M2 (n=518)



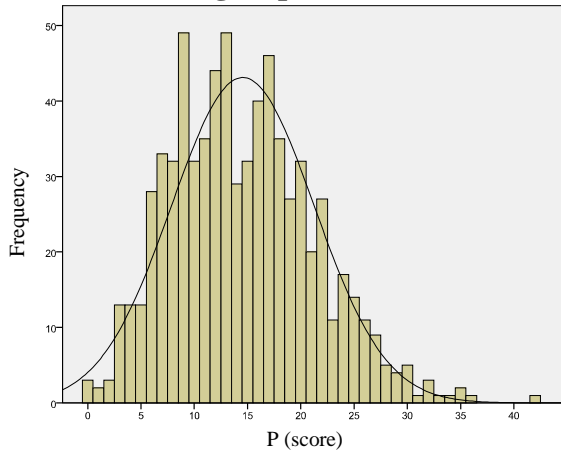
Girls in group [T] at M1 (n=79)



Girls in group [T] at M2 (n=79)



Girls in group [O] at M1 (n=723)



Girls in group [O] at M2 (n=723)

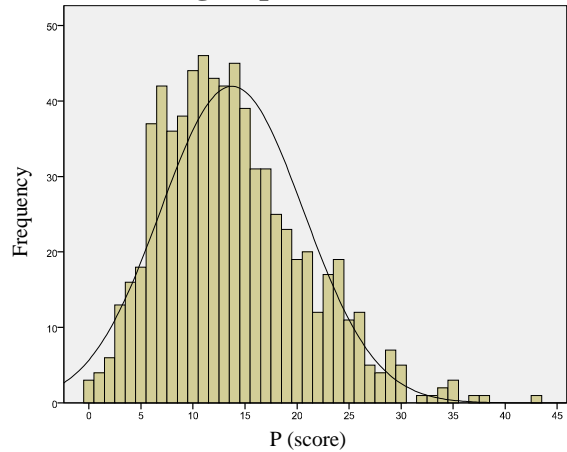
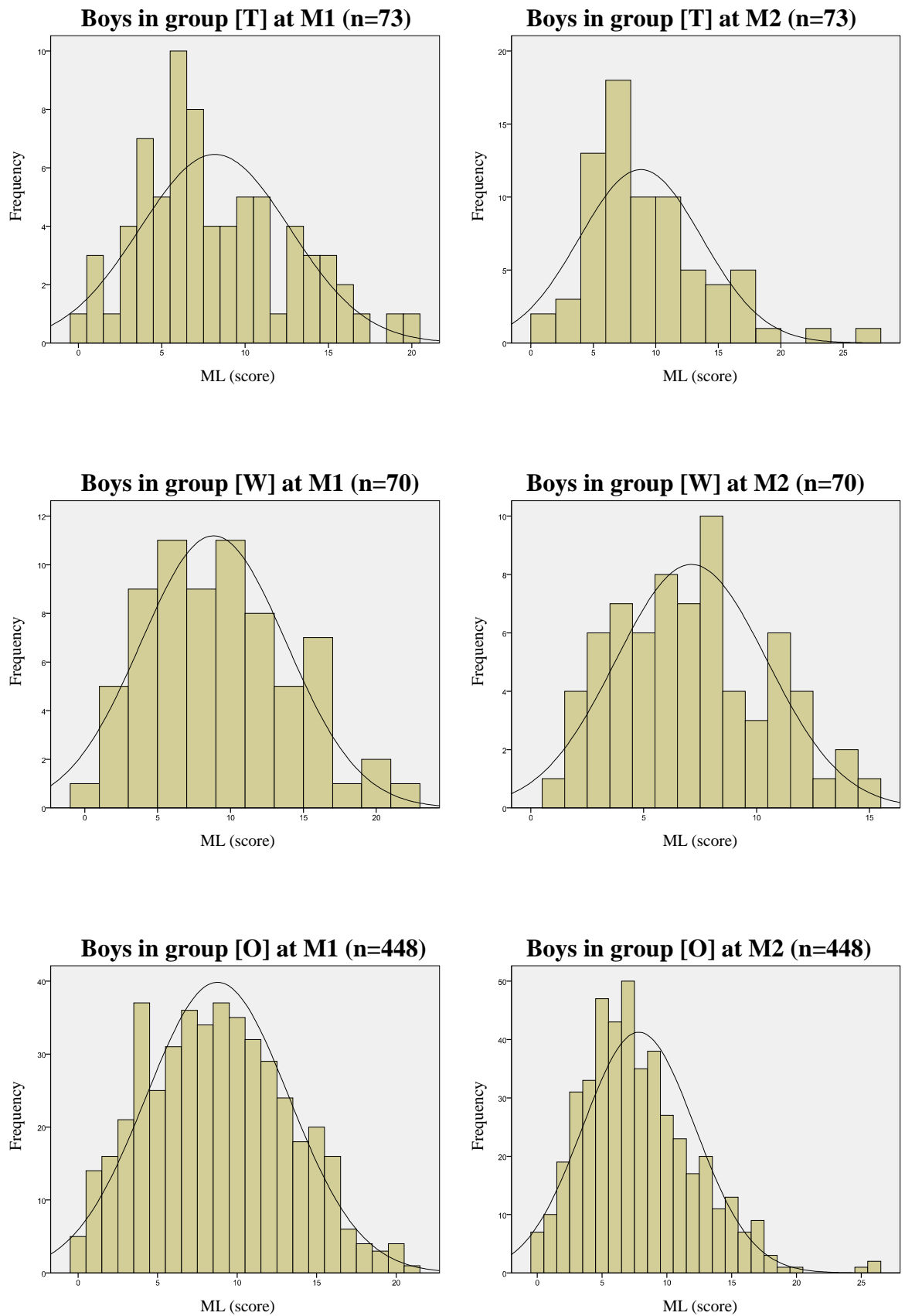
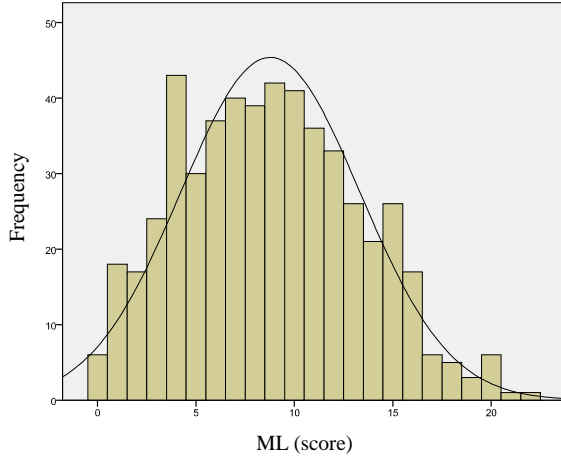


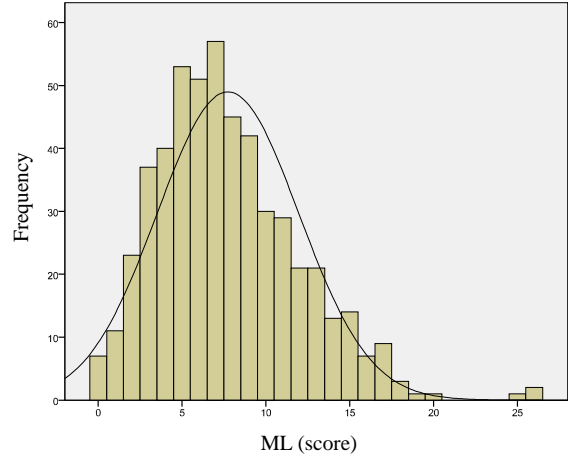
FIGURE A18.7: Histograms of the distribution for the crude ML results in the Gulu sample at M1 and M2 – stratified by gender and intervention group



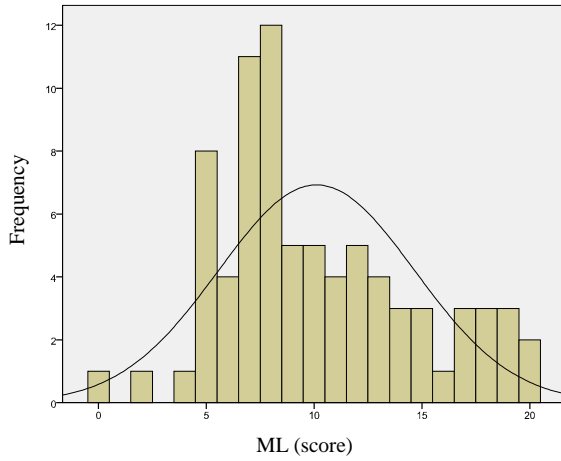
Boys in group [W+O] at M1 (n=518)



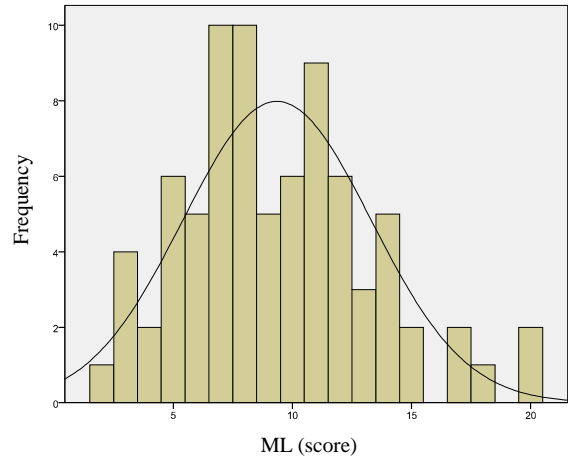
Boys in group [W+O] at M2 (n=518)



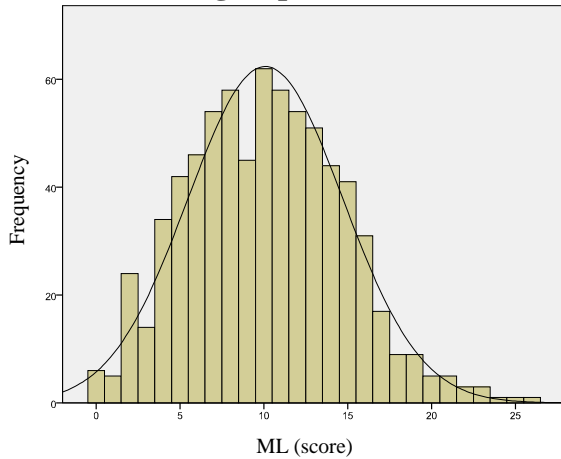
Girls in group [T] at M1 (n=79)



Girls in group [T] at M2 (n=79)



Girls in group [O] at M1 (n=723)



Girls in group [O] at M2 (n=723)

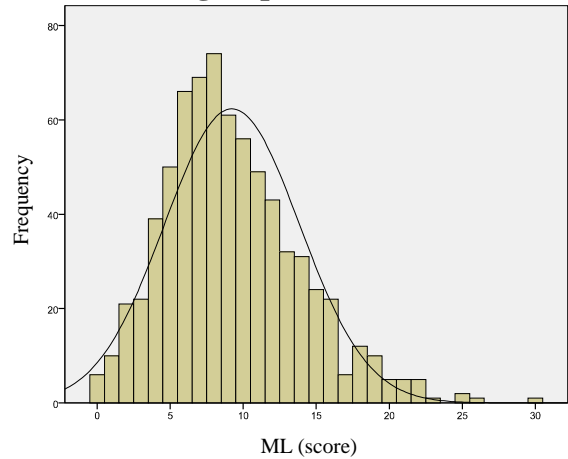
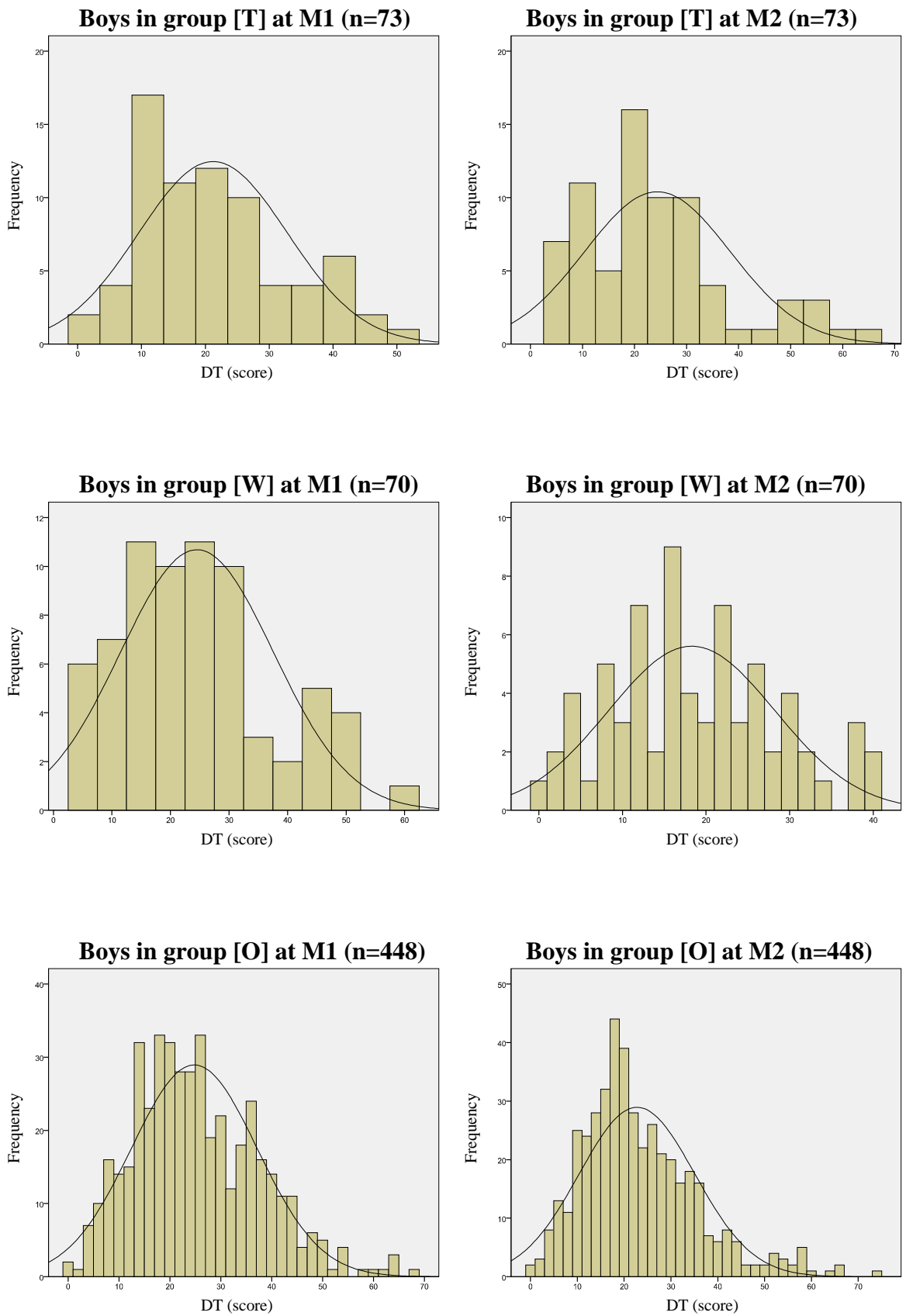


FIGURE A18.8: Histograms of the distribution for the crude DT results in the Gulu sample at M1 and M2 – stratified by gender and intervention group



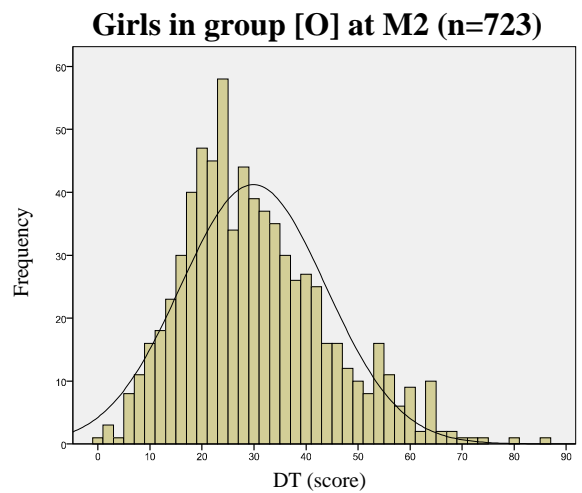
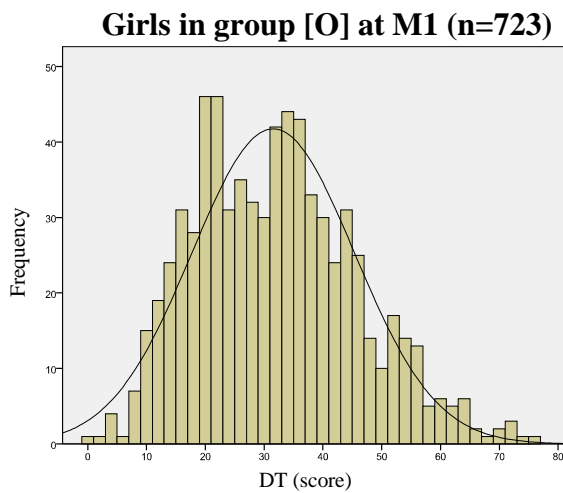
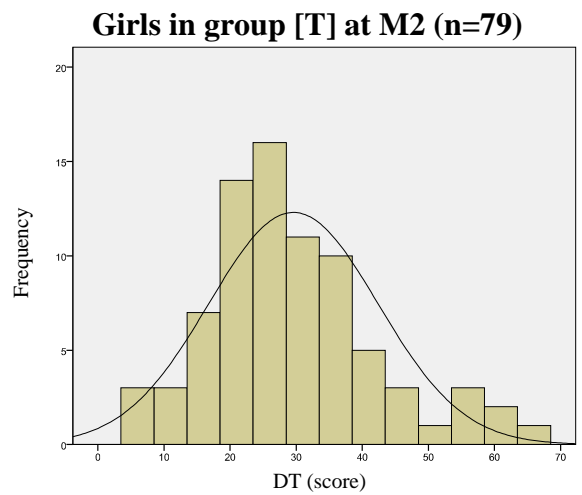
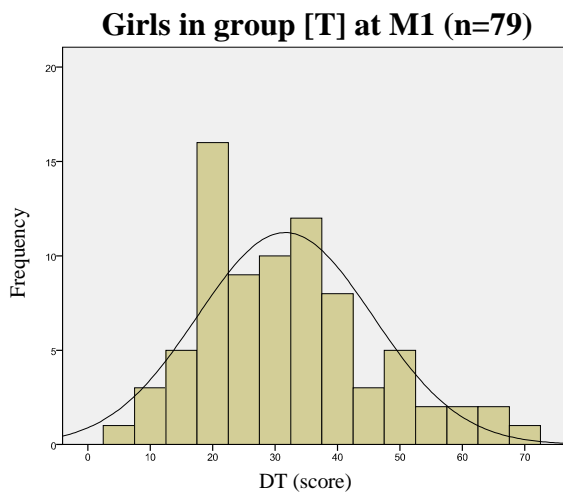
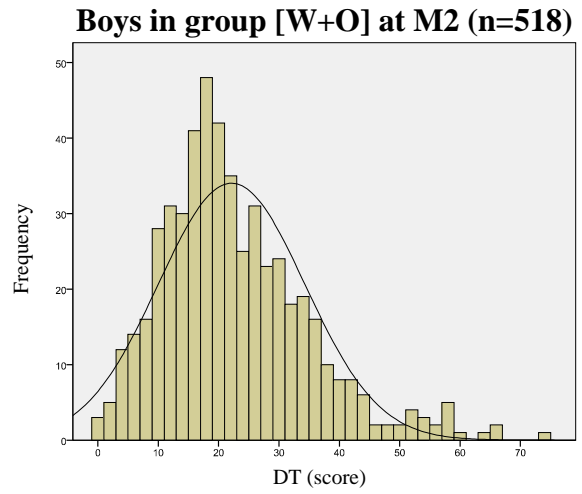
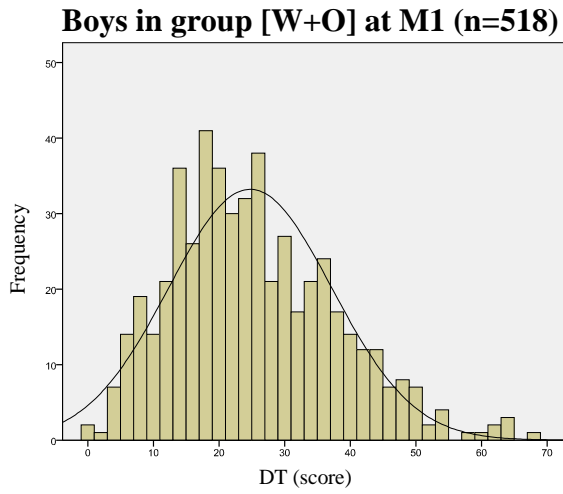
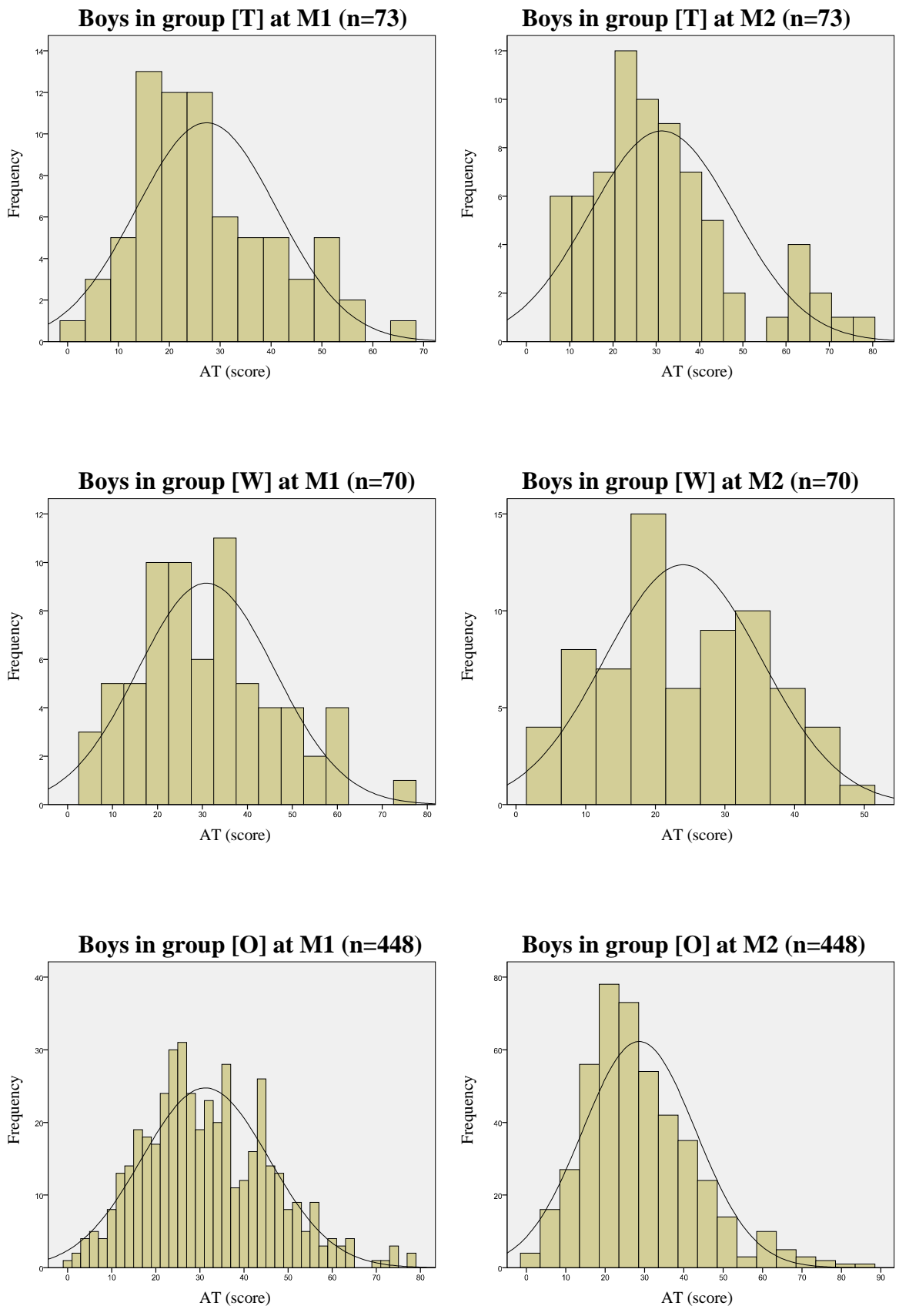
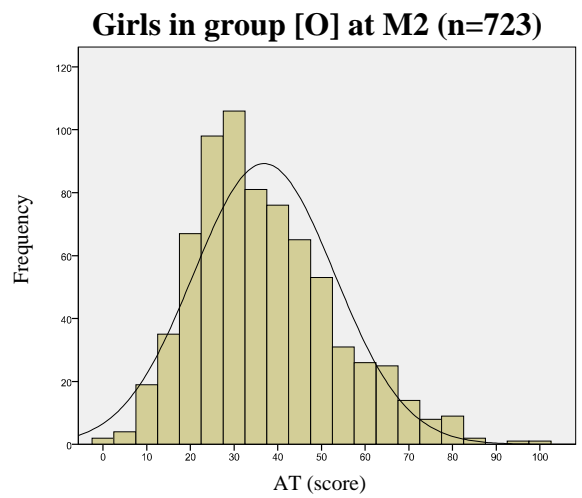
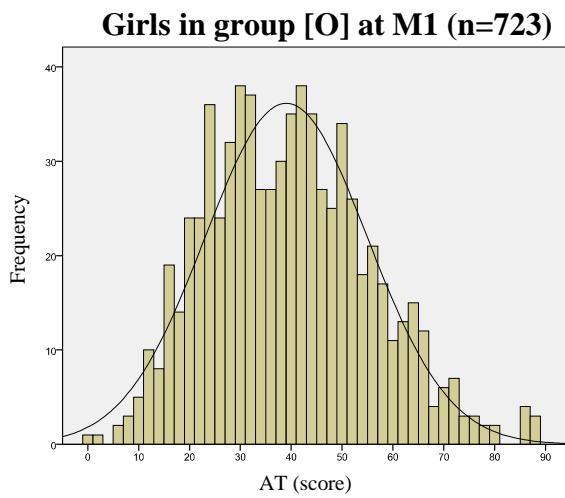
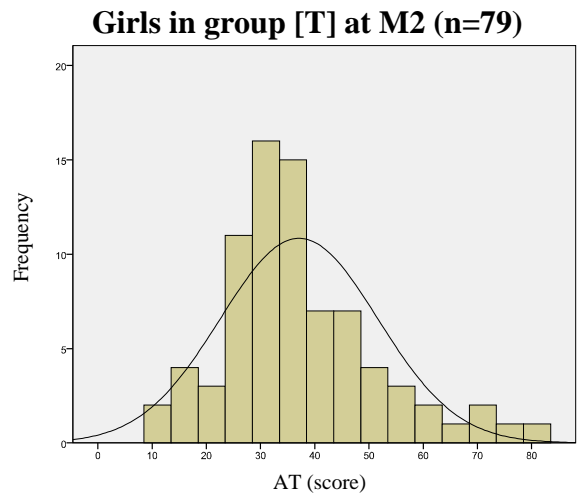
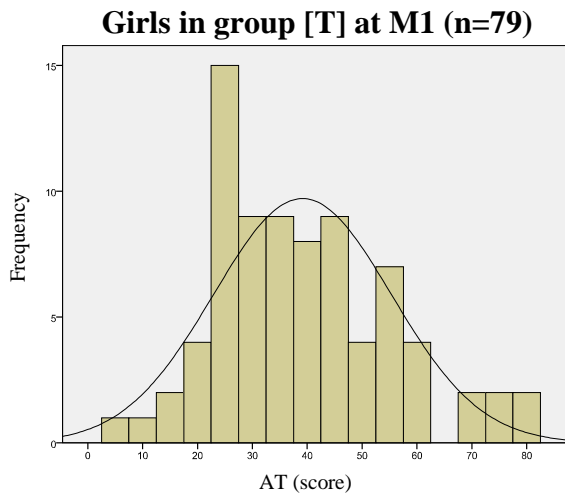
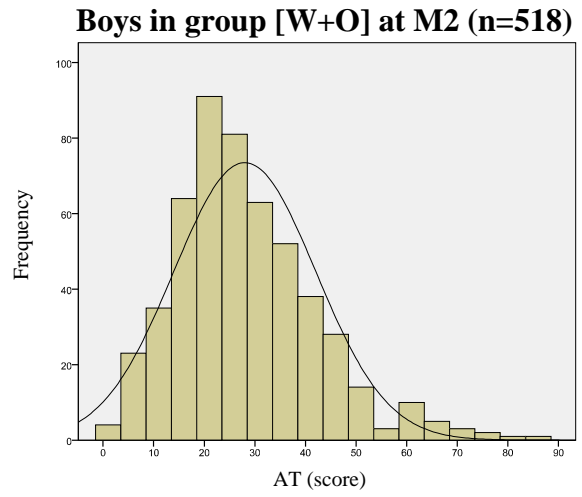
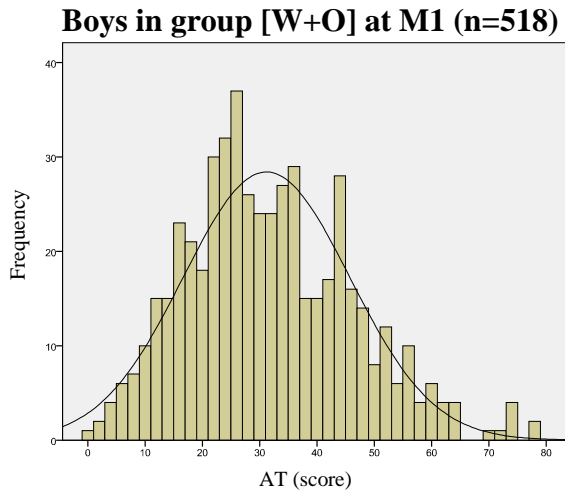


FIGURE A18.9: Histograms of the distribution for the crude AT results in the Gulu sample at M1 and M2 – stratified by gender and intervention group





Inspection of the histograms warranted further assessment of normality and statistical tests of the crude results were also completed (TABLES A18.1 – A18.6).

TABLE A18.1: Statistical tests for normality of the crude results for all of the outcomes for the boys in group [T] of the Gulu sample at M1 and M2

Outcome variable	n	Time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	72	M1	-0.699	0.283	1.191	0.559	0.962	0.028
		M2	-0.411	0.283	0.443	0.559	0.984	0.513
SBJ (cm)	70	M1	0.173	0.287	-0.627	0.566	0.983	0.454
		M2	0.407	0.287	0.122	0.566	0.975	0.186
MSFT (level complete)	70	M1	-0.042	0.287	-0.620	0.566	0.966	0.052
		M2	0.146	0.287	-0.336	0.566	0.963	0.039
TT (score)	73	M1	0.573	0.281	0.007	0.555	0.959	0.019
		M2	1.128	0.281	0.704	0.555	0.888	<0.001
K (score)	73	M1	0.924	0.281	0.321	0.555	0.920	<0.001
		M2	1.056	0.281	1.001	0.555	0.921	<0.001
P (score)	73	M1	0.681	0.281	0.151	0.555	0.960	0.022
		M2	0.661	0.281	0.369	0.555	0.955	0.010
ML (score)	73	M1	0.540	0.281	-0.244	0.555	0.965	0.039
		M2	1.218	0.281	2.085	0.555	0.916	<0.001
DT (score)	73	M1	0.638	0.281	-0.263	0.555	0.953	0.008
		M2	0.903	0.281	0.447	0.555	0.922	<0.001
AT (score)	73	M1	0.571	0.281	-0.132	0.555	0.964	0.035
		M2	0.942	0.281	0.438	0.555	0.921	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A18.2: Statistical tests for normality of the crude results for all of the outcomes for the boys in group [W] of the Gulu sample at M1 and M2

Outcome variable	n	Time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	69	M1	-0.772	0.289	0.980	0.570	0.962	0.034
		M2	-0.651	0.289	-0.185	0.570	0.954	0.013
SBJ (cm)	68	M1	-0.208	0.291	-0.658	0.574	0.973	0.155
		M2	0.010	0.291	-0.855	0.574	0.969	0.092
MSFT (level complete)	69	M1	-0.552	0.289	-0.633	0.570	0.921	<0.001
		M2	-0.498	0.289	0.220	0.570	0.955	0.015
TT (score)	70	M1	0.821	0.287	-0.076	0.566	0.927	<0.001
		M2	0.136	0.287	-0.949	0.566	0.965	0.050
K (score)	70	M1	0.633	0.287	-0.004	0.566	0.958	0.020
		M2	0.779	0.287	0.906	0.566	0.940	0.002
P (score)	70	M1	0.675	0.287	0.359	0.566	0.956	0.016
		M2	0.587	0.287	0.134	0.566	0.964	0.043
ML (score)	70	M1	0.438	0.287	-0.226	0.566	0.973	0.139
		M2	0.323	0.287	-0.601	0.566	0.970	0.096
DT (score)	70	M1	0.685	0.287	-0.082	0.566	0.949	0.006
		M2	0.293	0.287	-0.553	0.566	0.977	0.231
AT (score)	70	M1	0.524	0.287	-0.179	0.566	0.971	0.099
		M2	0.098	0.287	-0.895	0.566	0.973	0.142

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A18.3: Statistical tests for normality of the crude results for all of the outcomes for the boys in group [O] of the Gulu sample at M1 and M2

Outcome variable	n	Time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	436	M1	<i>-0.321</i>	<i>0.117</i>	<i>0.864</i>	<i>0.233</i>	<i>0.989</i>	<i>0.002</i>
		M2	<i>-0.420</i>	<i>0.117</i>	<i>1.121</i>	<i>0.233</i>	<i>0.986</i>	<i><0.001</i>
SBJ (cm)	429	M1	<i>0.339</i>	<i>0.118</i>	<i>0.805</i>	<i>0.235</i>	<i>0.987</i>	<i><0.001</i>
		M2	0.208	0.118	-0.038	0.235	<i>0.993</i>	<i>0.040</i>
MSFT (level complete)	429	M1	-0.178	0.118	-0.143	0.235	<i>0.972</i>	<i><0.001</i>
		M2	-0.205	0.118	-0.383	0.235	<i>0.974</i>	<i><0.001</i>
TT (score)	448	M1	<i>0.744</i>	<i>0.115</i>	<i>0.519</i>	<i>0.230</i>	<i>0.963</i>	<i><0.001</i>
		M2	<i>1.014</i>	<i>0.115</i>	<i>1.243</i>	<i>0.230</i>	<i>0.938</i>	<i><0.001</i>
K (score)	448	M1	<i>0.703</i>	<i>0.115</i>	0.284	0.230	<i>0.961</i>	<i><0.001</i>
		M2	<i>0.921</i>	<i>0.115</i>	<i>0.852</i>	<i>0.230</i>	<i>0.943</i>	<i><0.001</i>
P (score)	448	M1	<i>0.577</i>	<i>0.115</i>	<i>0.513</i>	<i>0.230</i>	<i>0.977</i>	<i><0.001</i>
		M2	<i>0.984</i>	<i>0.115</i>	<i>1.133</i>	<i>0.230</i>	<i>0.941</i>	<i><0.001</i>
ML (score)	448	M1	0.219	0.115	<i>-0.566</i>	<i>0.230</i>	<i>0.983</i>	<i><0.001</i>
		M2	<i>0.815</i>	<i>0.115</i>	<i>1.090</i>	<i>0.230</i>	<i>0.957</i>	<i><0.001</i>
DT (score)	448	M1	<i>0.575</i>	<i>0.115</i>	0.200	0.230	<i>0.976</i>	<i><0.001</i>
		M2	<i>0.995</i>	<i>0.115</i>	<i>1.264</i>	<i>0.230</i>	<i>0.944</i>	<i><0.001</i>
AT (score)	448	M1	<i>0.478</i>	<i>0.115</i>	0.090	0.230	<i>0.983</i>	<i><0.001</i>
		M2	<i>0.916</i>	<i>0.115</i>	<i>1.115</i>	<i>0.230</i>	<i>0.953</i>	<i><0.001</i>

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A18.4: Statistical tests for normality of the crude results for all of the outcomes for the boys in group [W+O] of the Gulu sample at M1 and M2

Outcome variable	n	Time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	505	M1	<i>-0.375</i>	<i>0.109</i>	<i>0.858</i>	<i>0.217</i>	<i>0.988</i>	<i><0.001</i>
		M2	<i>-0.452</i>	<i>0.109</i>	<i>0.924</i>	<i>0.217</i>	<i>0.987</i>	<i><0.001</i>
SBJ (cm)	497	M1	<i>0.266</i>	<i>0.110</i>	<i>0.582</i>	<i>0.219</i>	<i>0.992</i>	<i>0.009</i>
		M2	0.175	0.110	-0.156	0.219	<i>0.993</i>	<i>0.027</i>
MSFT (level complete)	498	M1	<i>-0.226</i>	<i>0.109</i>	-0.225	0.218	<i>0.971</i>	<i><0.001</i>
		M2	<i>-0.248</i>	<i>0.109</i>	-0.344	0.218	<i>0.973</i>	<i><0.001</i>
TT (score)	518	M1	<i>0.752</i>	<i>0.107</i>	0.414	0.214	<i>0.960</i>	<i><0.001</i>
		M2	<i>0.990</i>	<i>0.107</i>	<i>1.315</i>	<i>0.214</i>	<i>0.943</i>	<i><0.001</i>
K (score)	518	M1	<i>0.693</i>	<i>0.107</i>	0.241	0.214	<i>0.961</i>	<i><0.001</i>
		M2	<i>0.949</i>	<i>0.107</i>	<i>0.981</i>	<i>0.214</i>	<i>0.941</i>	<i><0.001</i>
P (score)	518	M1	<i>0.588</i>	<i>0.107</i>	<i>0.479</i>	<i>0.214</i>	<i>0.976</i>	<i><0.001</i>
		M2	<i>0.975</i>	<i>0.107</i>	<i>1.186</i>	<i>0.214</i>	<i>0.943</i>	<i><0.001</i>
ML (score)	518	M1	<i>0.258</i>	<i>0.107</i>	<i>-0.497</i>	<i>0.214</i>	<i>0.982</i>	<i><0.001</i>
		M2	<i>0.810</i>	<i>0.107</i>	<i>1.131</i>	<i>0.214</i>	<i>0.958</i>	<i><0.001</i>
DT (score)	518	M1	<i>0.589</i>	<i>0.107</i>	0.144	0.214	<i>0.975</i>	<i><0.001</i>
		M2	<i>0.969</i>	<i>0.107</i>	<i>1.308</i>	<i>0.214</i>	<i>0.948</i>	<i><0.001</i>
AT (score)	518	M1	<i>0.483</i>	<i>0.107</i>	0.038	0.214	<i>0.982</i>	<i><0.001</i>
		M2	<i>0.893</i>	<i>0.107</i>	<i>1.174</i>	<i>0.214</i>	<i>0.956</i>	<i><0.001</i>

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A18.5: Statistical tests for normality of the crude results for all of the outcomes for the girls in group [T] of the Gulu sample at M1 and M2

Outcome variable	n	Time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	79	M1	-0.697	0.271	0.798	0.535	0.967	0.041
		M2	-0.587	0.271	0.084	0.535	0.970	0.060
SBJ (cm)	79	M1	0.105	0.271	-0.425	0.535	0.990	0.822
		M2	0.225	0.271	0.427	0.535	0.983	0.361
MSFT (level complete)	79	M1	0.372	0.271	-0.552	0.535	0.952	0.005
		M2	0.006	0.271	-0.669	0.535	0.954	0.006
TT (score)	79	M1	0.534	0.271	-0.218	0.535	0.968	0.044
		M2	0.849	0.271	1.125	0.535	0.952	0.005
K (score)	79	M1	0.390	0.271	-0.420	0.535	0.971	0.065
		M2	0.573	0.271	0.374	0.535	0.973	0.093
P (score)	79	M1	0.718	0.271	0.235	0.535	0.955	0.007
		M2	0.672	0.271	0.463	0.535	0.959	0.012
ML (score)	79	M1	0.523	0.271	-0.403	0.535	0.946	0.002
		M2	0.543	0.271	0.161	0.535	0.969	0.054
DT (score)	79	M1	0.595	0.271	0.033	0.535	0.967	0.039
		M2	0.807	0.271	0.670	0.535	0.948	0.003
AT (score)	79	M1	0.599	0.271	0.076	0.535	0.963	0.022
		M2	0.854	0.271	0.898	0.535	0.947	0.003

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A18.6: Statistical tests for normality of the crude results for all of the outcomes for the girls in group [O] of the Gulu sample at M1 and M2

Outcome variable	n	Time	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	694	M1	<i>-0.349</i>	<i>0.093</i>	0.305	0.185	<i>0.993</i>	<i>0.002</i>
		M2	<i>-0.262</i>	<i>0.093</i>	0.055	0.185	<i>0.994</i>	<i>0.010</i>
SBJ (cm)	688	M1	-0.027	0.093	0.188	0.186	0.997	0.267
		M2	<i>0.376</i>	<i>0.093</i>	<i>0.565</i>	<i>0.186</i>	<i>0.990</i>	<i><0.001</i>
MSFT (level complete)	681	M1	<i>0.542</i>	<i>0.094</i>	-0.122	0.187	<i>0.942</i>	<i><0.001</i>
		M2	<i>0.445</i>	<i>0.094</i>	-0.192	0.187	<i>0.961</i>	<i><0.001</i>
TT (score)	723	M1	<i>0.524</i>	<i>0.091</i>	0.002	0.182	<i>0.977</i>	<i><0.001</i>
		M2	<i>0.783</i>	<i>0.091</i>	<i>0.550</i>	<i>0.182</i>	<i>0.960</i>	<i><0.001</i>
K (score)	723	M1	<i>0.500</i>	<i>0.091</i>	-0.280	0.182	<i>0.973</i>	<i><0.001</i>
		M2	<i>0.710</i>	<i>0.091</i>	0.229	0.182	<i>0.962</i>	<i><0.001</i>
P (score)	723	M1	<i>0.463</i>	<i>0.091</i>	0.087	0.182	<i>0.983</i>	<i><0.001</i>
		M2	<i>0.700</i>	<i>0.091</i>	<i>0.488</i>	<i>0.182</i>	<i>0.968</i>	<i><0.001</i>
ML (score)	723	M1	<i>0.262</i>	<i>0.091</i>	-0.111	0.182	<i>0.988</i>	<i><0.001</i>
		M2	<i>0.713</i>	<i>0.091</i>	<i>0.757</i>	<i>0.182</i>	<i>0.968</i>	<i><0.001</i>
DT (score)	723	M1	<i>0.435</i>	<i>0.091</i>	-0.178	0.182	<i>0.983</i>	<i><0.001</i>
		M2	<i>0.675</i>	<i>0.091</i>	0.290	0.182	<i>0.969</i>	<i><0.001</i>
AT (score)	723	M1	<i>0.361</i>	<i>0.091</i>	-0.163	0.182	<i>0.988</i>	<i><0.001</i>
		M2	<i>0.674</i>	<i>0.091</i>	<i>0.376</i>	<i>0.182</i>	<i>0.971</i>	<i><0.001</i>

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12].

The assumption of HOV between M1 vs. M2 for the crude results of the dependent variables in each of the treatment groups was assessed using the Levene and Brown-Forsythe tests (TABLES A18.7 – A18.10). These statistical tests for HOV were robust to violations of the normality assumption ^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$).

TABLE A18.7: Statistical tests of the HOV between M1 and M2 for all of the crude outcomes of group [T] in the Gulu sample - stratified by gender

Outcome variable	Gender	n	Variance		Levene test		Brown-Forsythe test		Summary
			M1	M2	F	p	F	p	
BFA (z score)	Boys	72	0.648	0.710	0.058	0.810	0.066	0.797	√
	Girls	79	0.739	0.557	1.038	0.310	1.133	0.289	√
SBJ (cm)	Boys	70	449.15	463.64	0.232	0.631	0.286	0.594	√
	Girls	79	295.97	284.93	0.142	0.707	0.166	0.684	√
MSFT (level complete)	Boys	70	3.71	3.57	0.219	0.641	0.217	0.642	√
	Girls	79	3.43	3.47	0.017	0.898	0.047	0.828	√
TT (score)	Boys	73	33.94	48.75	0.879	0.350	0.467	0.495	√
	Girls	79	44.14	41.53	0.689	0.408	0.546	0.461	√
K (score)	Boys	73	24.56	36.83	1.492	0.224	0.932	0.336	√
	Girls	79	36.52	30.15	2.184	0.141	1.819	0.179	√
P (score)	Boys	73	31.39	42.67	0.847	0.359	1.123	0.291	√
	Girls	79	51.23	36.10	2.307	0.131	2.161	0.144	√
ML (score)	Boys	73	20.32	24.05	0.033	0.857	0.017	0.896	√
	Girls	79	20.72	15.59	2.083	0.151	1.266	0.262	√
DT (score)	Boys	73	136.58	196.24	0.895	0.346	0.638	0.426	√
	Girls	79	196.93	164.14	1.180	0.279	1.410	0.237	√
AT (score)	Boys	73	190.95	280.60	1.361	0.245	1.036	0.311	√
	Girls	79	263.47	211.30	1.941	0.166	1.831	0.178	√

NOTE: Results in ***BOLD ITALICS*** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

TABLE A18.8: Statistical tests of the HOV between M1 and M2 for all of the crude outcomes for boys in group [W] of the Gulu sample

Outcome variable	n	Variance		Levene test		Brown-Forsythe test		Summary
		M1	M2	F	p	F	p	
BFA (z score)	69	0.727	0.817	1.105	0.295	0.989	0.322	√
SBJ (cm)	68	393.22	432.98	0.685	0.409	0.568	0.452	√
MSFT (level complete)	69	3.59	3.72	0.077	0.782	0.017	0.897	√
TT (score)	70	42.61	23.45	4.691	0.032	2.184	0.142	√
K (score)	70	25.24	14.55	6.722	0.011	5.952	0.016	√
P (score)	70	36.79	26.03	1.399	0.239	1.076	0.301	√
ML (score)	70	24.92	11.20	10.101	0.002	10.129	0.002	(L)
DT (score)	70	170.86	99.28	4.100	0.045	3.611	0.059	√
AT (score)	70	233.17	127.28	4.459	0.037	3.744	0.055	√

TABLE A18.9: Statistical tests of the HOV between M1 and M2 for all of the crude outcomes for boys in group [W+O] of the Gulu sample

Outcome variable	n	Variance		Levene test		Brown-Forsythe test		Summary
		M1	M2	F	p	F	p	
BFA (z score)	505	0.784	0.781	0.015	0.902	0.002	0.961	√
SBJ (cm)	497	410.73	472.47	3.520	0.061	3.196	0.074	√
MSFT (level complete)	498	3.80	4.40	4.585	0.033	3.137	0.077	√
TT (score)	518	39.40	37.13	1.182	0.277	0.858	0.355	√
K (score)	518	26.14	23.69	3.958	0.047	4.146	0.042	√
P (score)	518	37.38	37.90	0.139	0.709	0.449	0.503	√
ML (score)	518	20.73	17.79	6.815	0.009	8.025	0.005	(L)
DT (score)	518	154.78	147.53	1.665	0.197	2.046	0.153	√
AT (score)	518	211.57	197.58	1.866	0.172	2.433	0.119	√

NOTE: Results in **BOLD ITALICS** suggest a statistically significant heterogeneity of variance ($p < 0.01$).
 NOTE: In the summary column - √ denotes satisfactory HOV; L indicates a "liberal" statistical test ^[12,14].

TABLE A18.10: Statistical tests of the HOV between M1 and M2 for all of the crude outcomes for group [O] of the Gulu sample - stratified by gender

Outcome variable	Gender	n	Variance		Levene test		Brown-Forsythe test		Summary
			M1	M2	F	p	F	p	
BFA (z score)	Boys	436	0.795	0.777	0.063	0.802	0.102	0.749	√
	Girls	694	0.820	0.775	0.250	0.617	0.240	0.625	√
SBJ (cm)	Boys	429	413.39	475.87	2.947	0.086	2.765	0.097	√
	Girls	688	303.33	375.04	4.405	0.036	4.376	0.037	√
MSFT (level complete)	Boys	429	3.84	4.48	5.031	0.025	4.098	0.043	√
	Girls	681	2.79	4.15	26.944	<0.001	21.171	<0.001	L
TT (score)	Boys	448	38.98	38.77	0.181	0.670	0.263	0.608	√
	Girls	723	50.40	50.03	0.243	0.622	0.698	0.403	√
K (score)	Boys	448	26.33	24.68	1.965	0.161	2.052	0.152	√
	Girls	723	33.05	32.51	0.354	0.552	0.404	0.525	√
P (score)	Boys	448	37.55	39.03	0.010	0.920	0.202	0.654	√
	Girls	723	44.77	47.29	0.008	0.930	0.003	0.960	√
ML (score)	Boys	448	20.13	18.78	2.683	0.102	3.494	0.062	√
	Girls	723	21.37	21.41	0.518	0.472	0.817	0.366	√
DT (score)	Boys	448	152.64	152.61	0.481	0.488	0.849	0.357	√
	Girls	723	191.00	195.76	0.054	0.815	0.225	0.635	√
AT (score)	Boys	448	208.69	205.92	0.512	0.475	0.998	0.318	√
	Girls	723	254.64	261.18	0.056	0.813	0.233	0.630	√

NOTE: Results in ***BOLD ITALICS*** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

NOTE: In the summary column - √ denotes satisfactory HOV; L indicates a "liberal" statistical test ^[12,14].

With the exception of the MSFT in the girls sample for group [O] and ML in the boys sample for both groups [W] and [W+O], the assumption of HOV held for all of the dependent variables in each treatment group when comparing M1 vs. M2. The statistical analysis could proceed with little concern when the HOV assumption was satisfied ^[12,14].

The violation of the HOV assumption for ML in the boys sample for both groups [W] and [W+O] was offset by the large and balanced sample size ^[12,14]. Despite a tendency towards a “liberal” statistical test that increased the risk of a type I error, the results of ANOVA analysis in these circumstances were minimally affected.

The violation of the HOV assumption for the MSFT in the girls sample for group [O] was also offset by the large and balanced sample size ^[12,14]. However, the F-statistic of the Levene and Brown-Forsythe tests suggested a severe violation of the assumption of HOV in this case. This may have significantly compromised the outcomes of the subsequent statistical testing. The analysis was referred to as a “liberal” statistical test and the risk of a type I error was increased ^[12,14].

A consistent statistical approach was applied for all outcome variables when comparing M1 vs. M2 for each treatment group. The interpretation of any significant statistical findings was treated with caution for the “liberal” statistical test of the MSFT for girls in group [O] ^[12,14].

APPENDIX A19: Impact of the Intervention – Between-Group Differences (Assumptions)

The statistical methods used to assess whether there was any difference between the treatment groups in the change from M1 to M2 assumed an approximately normal distribution for each of the outcome variables. This was initially assessed visually using a histogram of the eligible sample for the crude results of each of the dependent variables (FIGURES A19.1 – A19.9). Tests for normality assessing skew, kurtosis and a Shapiro-Wilk statistic of the crude data were also completed (TABLES A19.1 & A19.2).

FIGURE A19.1: Histograms of the distribution of the crude change in BFA from M1 to M2 in the Gulu sample – stratified by gender and intervention group

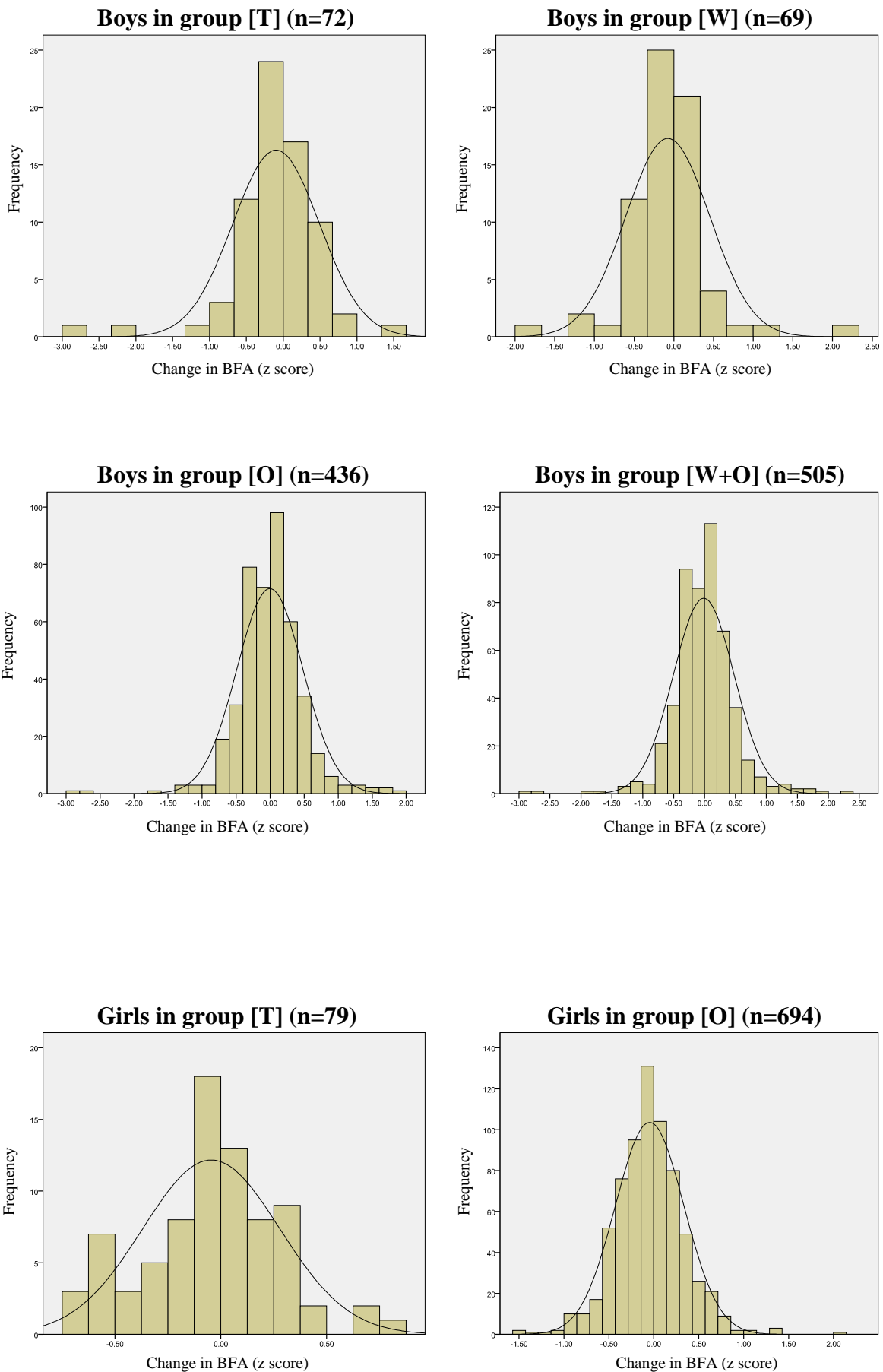


FIGURE A19.2: Histograms of the distribution of the crude change in SBJ from M1 to M2 in the Gulu sample – stratified by gender and intervention group

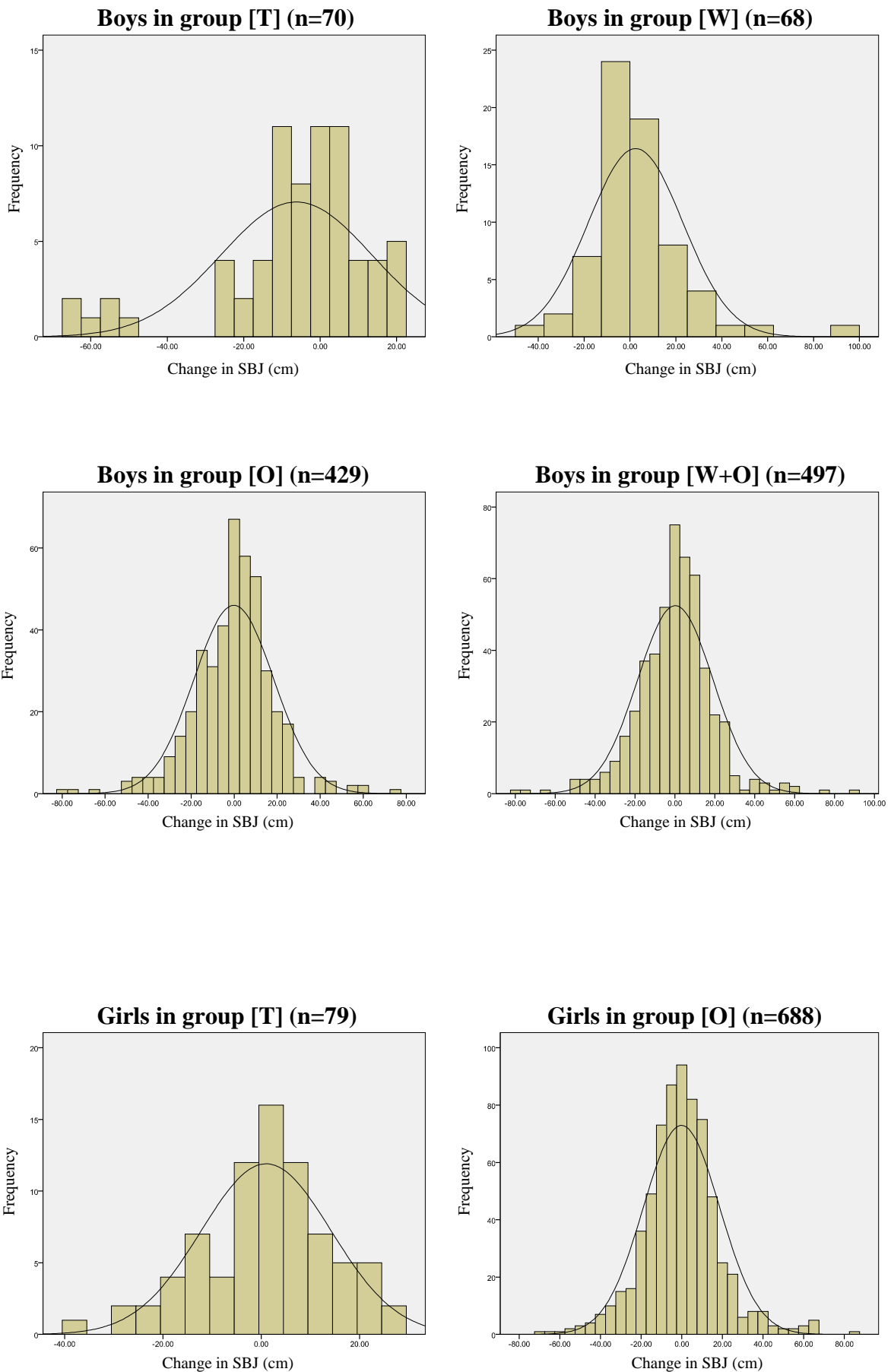


FIGURE A19.3: Histograms of the distribution of the crude change in MSFT from M1 to M2 in the Gulu sample – stratified by gender and intervention group

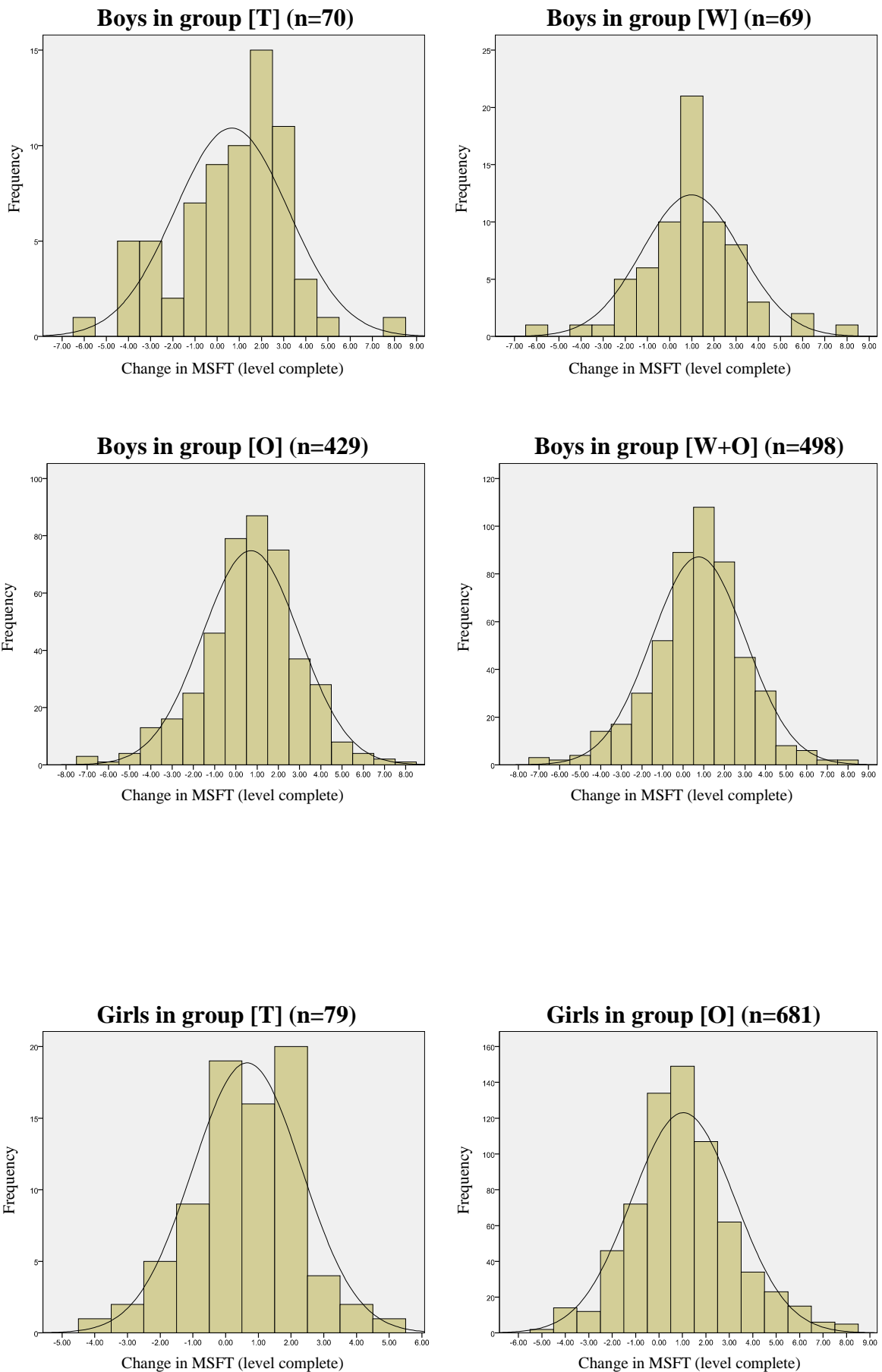


FIGURE A19.4: Histograms of the distribution of the crude change in TT from M1 to M2 in the Gulu sample – stratified by gender and intervention group

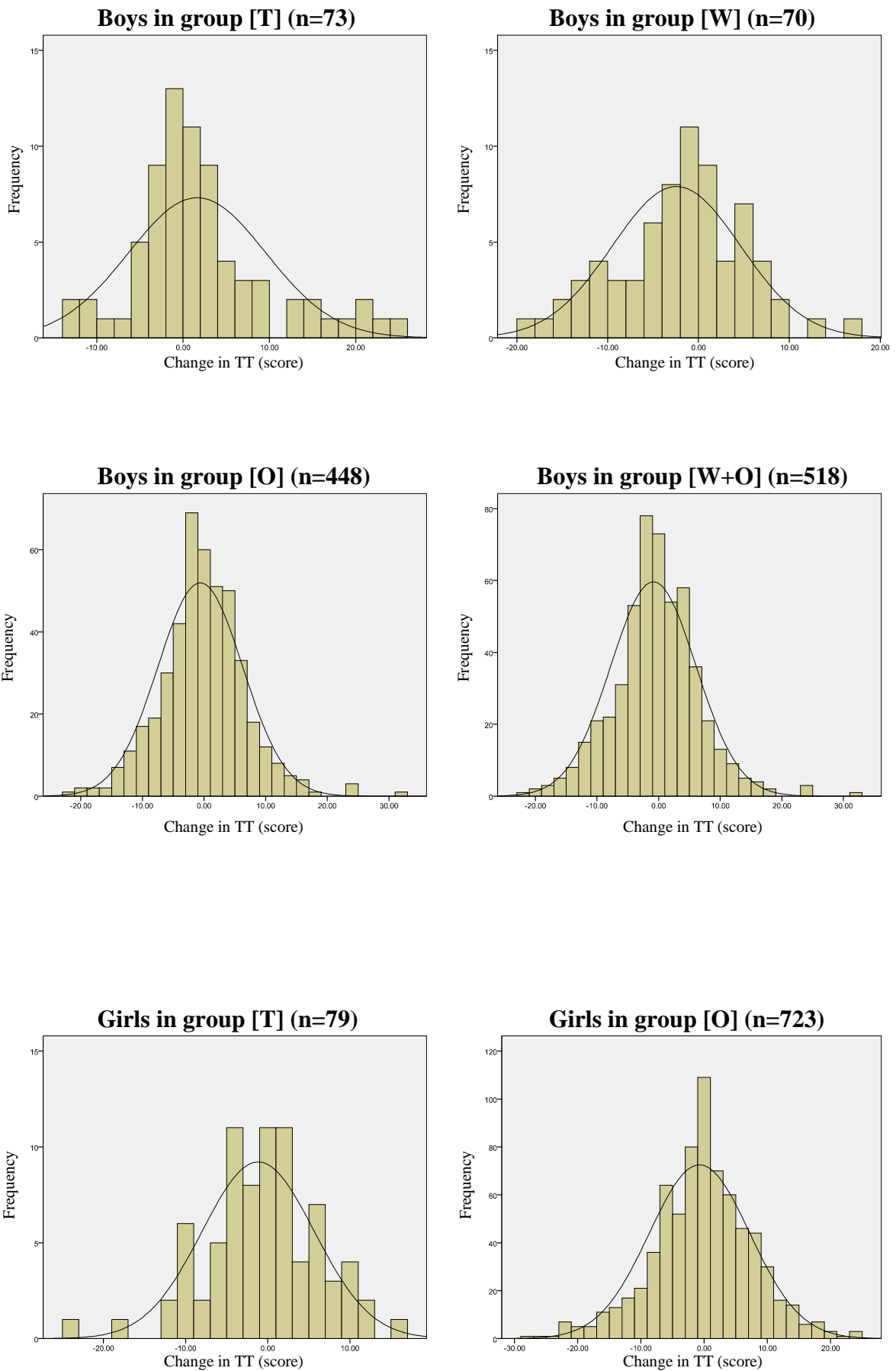


FIGURE A19.5: Histograms of the distribution of the crude change in K from M1 to M2 in the Gulu sample – stratified by gender and intervention group

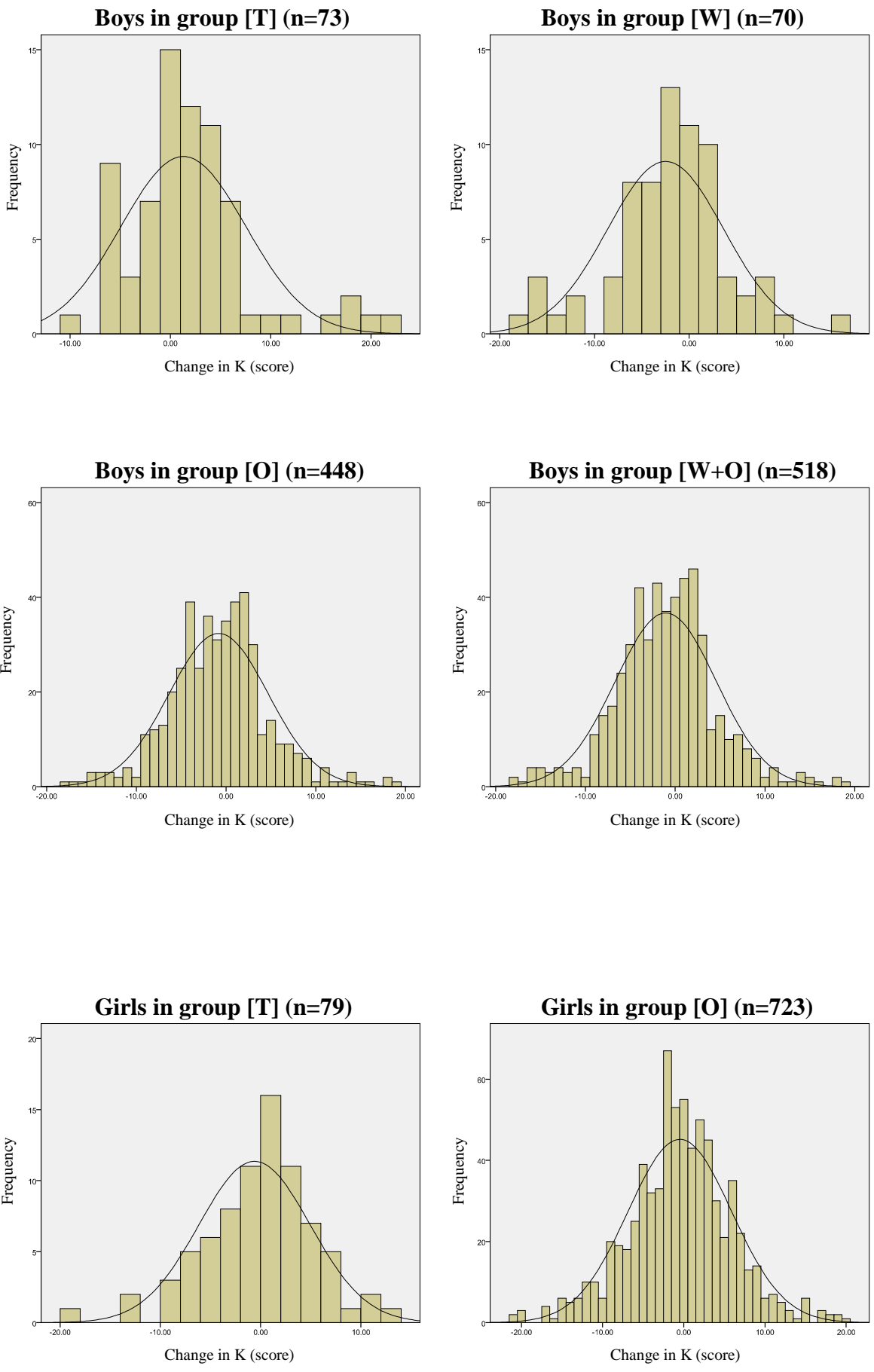


FIGURE A19.6: Histograms of the distribution of the crude change in P from M1 to M2 in the Gulu sample – stratified by gender and intervention group

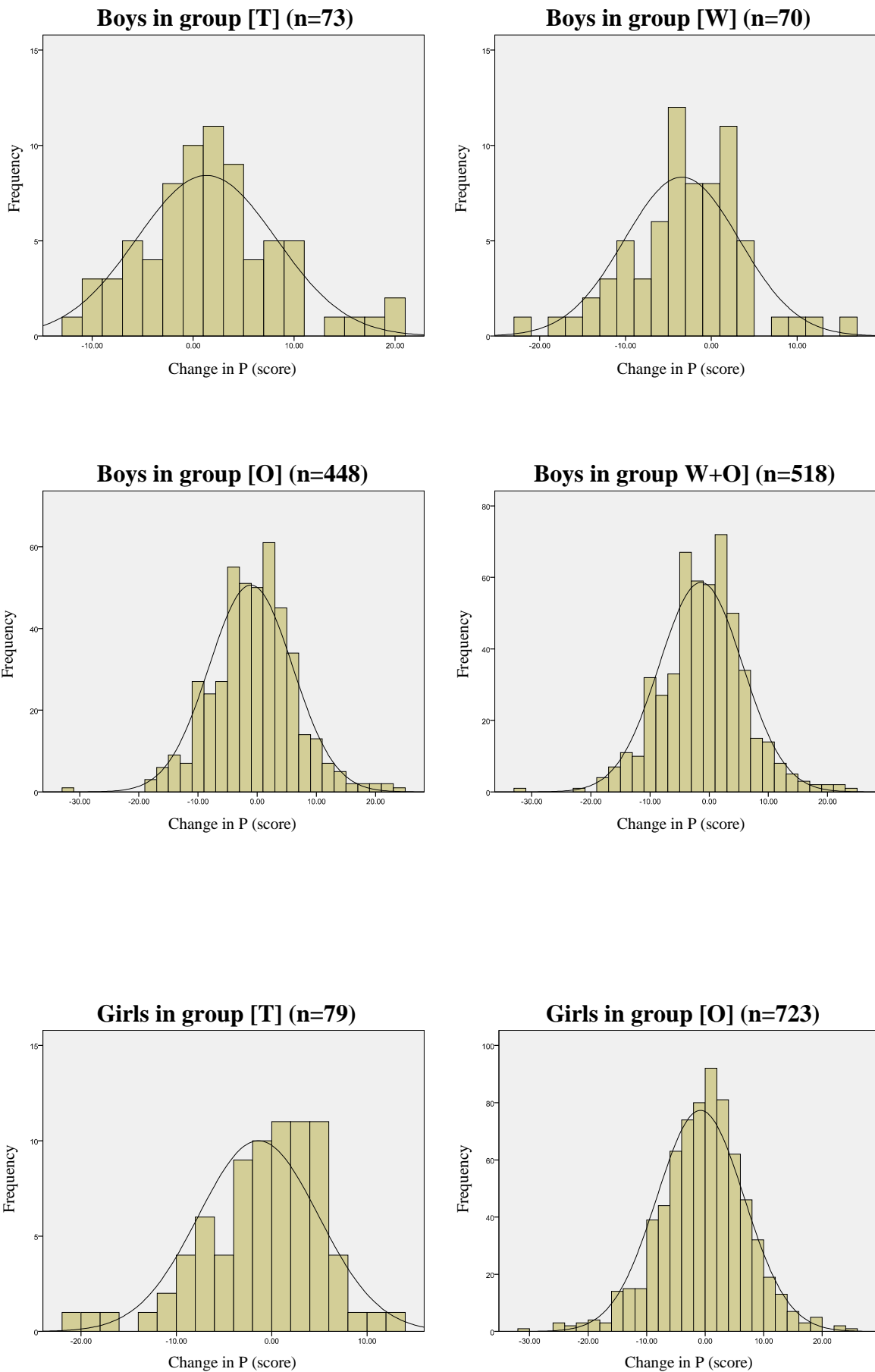


FIGURE A19.7: Histograms of the distribution of the crude change in ML from M1 to M2 in the Gulu sample – stratified by gender and intervention group

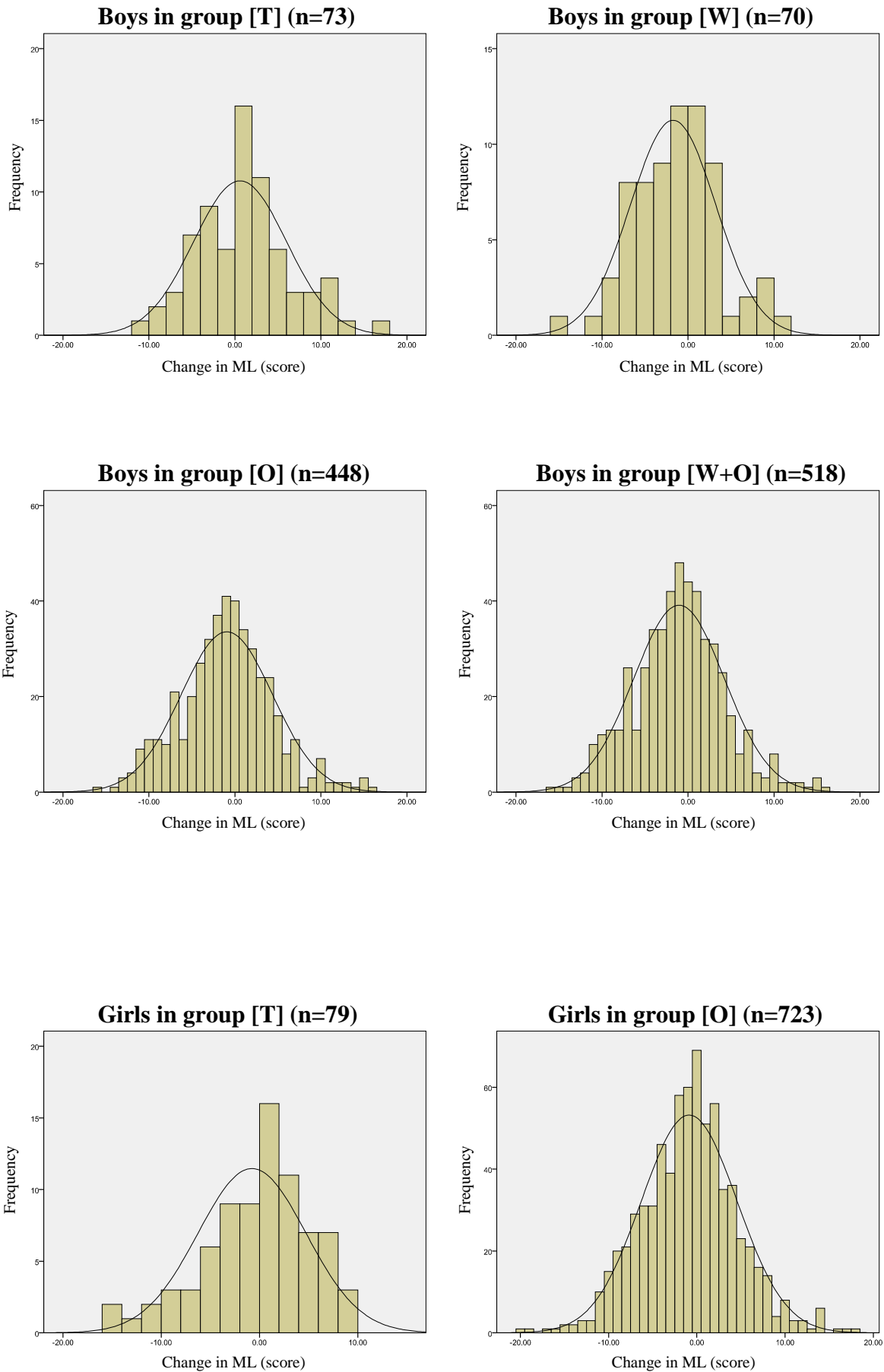


FIGURE A19.8: Histograms of the distribution of the crude change in DT from M1 to M2 in the Gulu sample – stratified by gender and intervention group

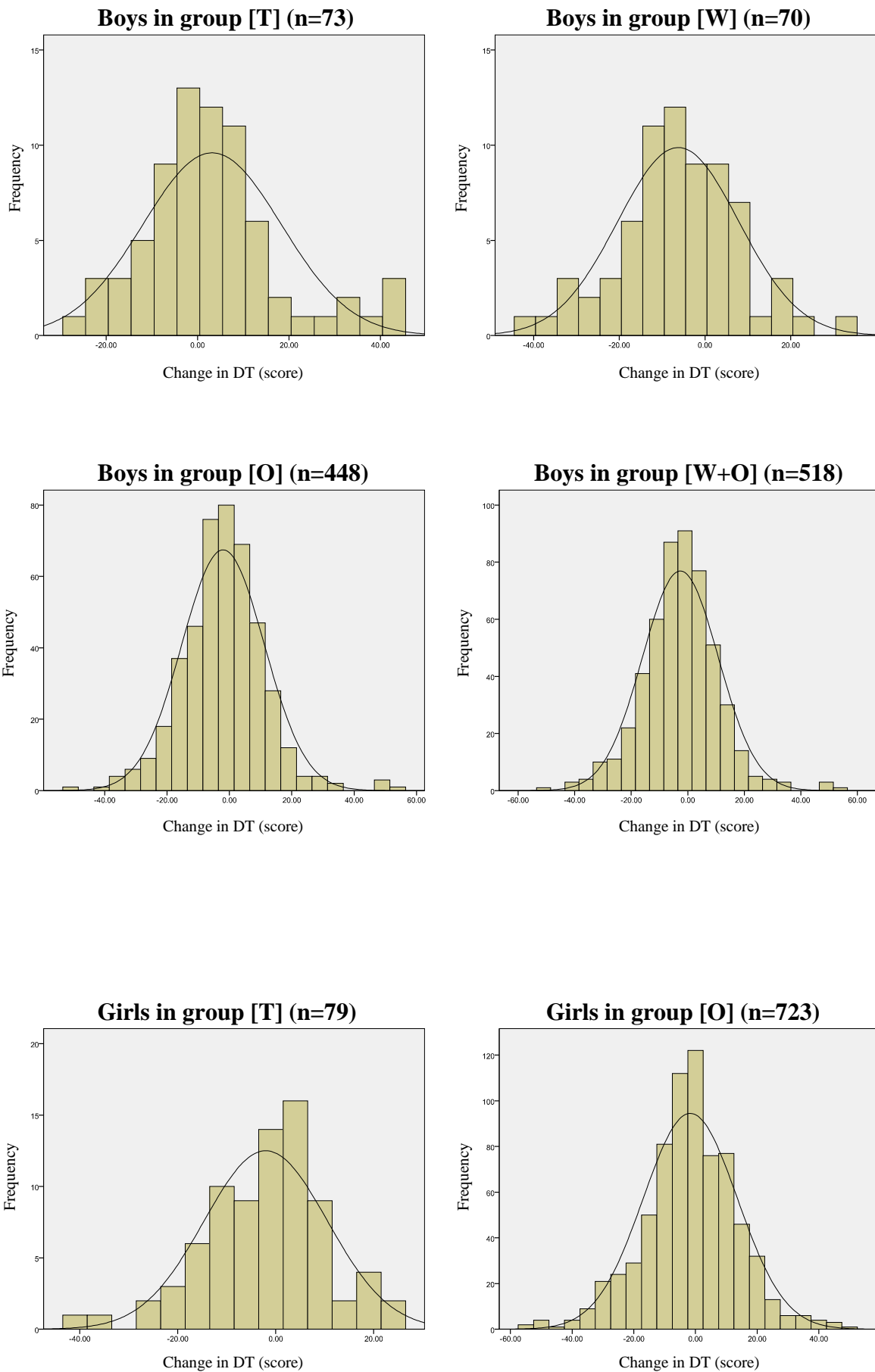


FIGURE A19.9: Histograms of the distribution of the crude change in AT from M1 to M2 in the Gulu sample – stratified by gender and intervention group

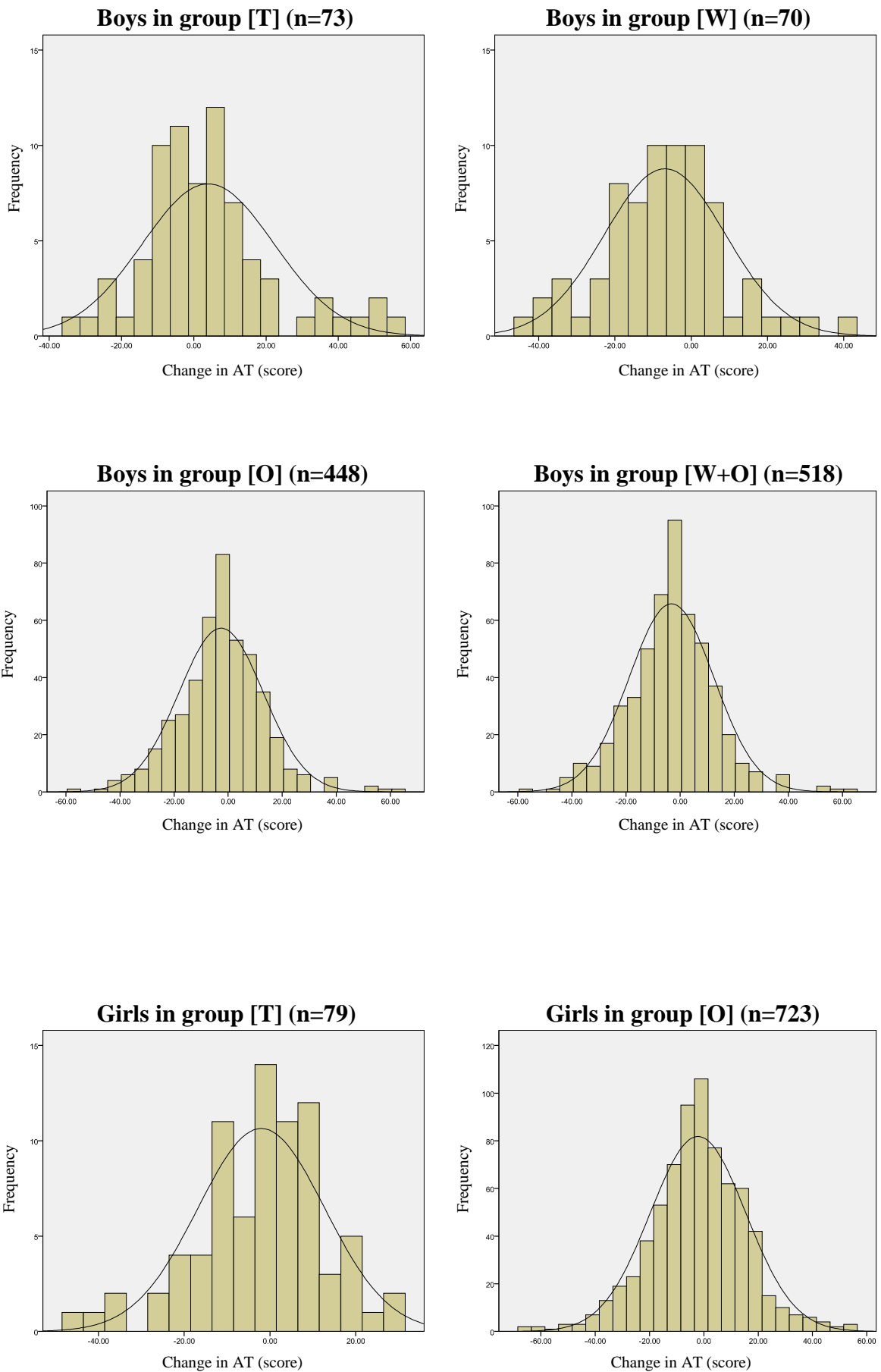


TABLE A19.1: Statistical tests for normality of the crude change in all of the outcomes for the boys from M1 to M2 in the Gulu sample – stratified by intervention group

Outcome variable	Δ in group	n	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	[T]	72	<i>-1.793</i>	<i>0.283</i>	<i>7.661</i>	<i>0.559</i>	<i>0.845</i>	<i><0.001</i>
	[W]	69	<i>0.718</i>	<i>0.289</i>	<i>7.026</i>	<i>0.570</i>	<i>0.874</i>	<i><0.001</i>
	[O]	436	<i>-0.574</i>	<i>0.117</i>	<i>6.443</i>	<i>0.233</i>	<i>0.925</i>	<i><0.001</i>
	[W+O]	505	<i>-0.367</i>	<i>0.109</i>	<i>6.350</i>	<i>0.217</i>	<i>0.921</i>	<i><0.001</i>
SBJ (cm)	[T]	70	<i>-1.444</i>	<i>0.287</i>	<i>2.226</i>	<i>0.566</i>	<i>0.864</i>	<i><0.001</i>
	[W]	68	<i>1.140</i>	<i>0.291</i>	<i>4.429</i>	<i>0.574</i>	<i>0.918</i>	<i><0.001</i>
	[O]	429	-0.226	0.118	<i>2.443</i>	<i>0.235</i>	<i>0.966</i>	<i><0.001</i>
	[W+O]	497	0.020	0.110	<i>2.876</i>	<i>0.219</i>	<i>0.962</i>	<i><0.001</i>
MSFT (level complete)	[T]	70	-0.310	0.287	0.325	0.566	<i>0.953</i>	<i>0.011</i>
	[W]	69	0.052	0.289	<i>1.982</i>	<i>0.570</i>	<i>0.950</i>	<i>0.007</i>
	[O]	429	<i>-0.316</i>	<i>0.118</i>	<i>0.812</i>	<i>0.235</i>	<i>0.971</i>	<i><0.001</i>
	[W+O]	498	<i>-0.272</i>	<i>0.109</i>	<i>0.949</i>	<i>0.218</i>	<i>0.970</i>	<i><0.001</i>
TT (score)	[T]	73	<i>0.876</i>	<i>0.281</i>	0.905	0.555	<i>0.923</i>	<i><0.001</i>
	[W]	70	-0.117	0.287	0.190	0.566	0.983	0.458
	[O]	448	<i>0.296</i>	<i>0.115</i>	<i>1.927</i>	<i>0.230</i>	<i>0.976</i>	<i><0.001</i>
	[W+O]	518	<i>0.229</i>	<i>0.107</i>	<i>1.685</i>	<i>0.214</i>	<i>0.979</i>	<i><0.001</i>
K (score)	[T]	73	<i>1.156</i>	<i>0.281</i>	<i>2.044</i>	<i>0.555</i>	<i>0.901</i>	<i><0.001</i>
	[W]	70	-0.240	0.287	0.983	0.566	<i>0.965</i>	<i>0.048</i>
	[O]	448	0.201	0.115	1.198	0.230	<i>0.981</i>	<i><0.001</i>
	[W+O]	518	0.100	0.107	1.206	0.214	<i>0.980</i>	<i><0.001</i>
P (score)	[T]	73	0.512	0.281	0.492	0.555	0.971	0.096
	[W]	70	-0.165	0.287	0.718	0.566	0.977	0.230
	[O]	448	0.059	0.115	<i>1.215</i>	<i>0.230</i>	<i>0.986</i>	<i><0.001</i>
	[W+O]	518	0.043	0.107	<i>1.145</i>	<i>0.214</i>	<i>0.986</i>	<i><0.001</i>
ML (score)	[T]	73	0.471	0.281	0.476	0.555	0.977	0.194
	[W]	70	0.065	0.287	0.032	0.566	0.986	0.627
	[O]	448	0.171	0.115	0.428	0.230	<i>0.988</i>	<i><0.001</i>
	[W+O]	518	0.166	0.107	0.391	0.214	<i>0.990</i>	<i>0.001</i>
DT (score)	[T]	73	<i>0.849</i>	<i>0.281</i>	1.010	0.555	<i>0.941</i>	<i>0.002</i>
	[W]	70	-0.008	0.287	0.762	0.566	0.985	0.548
	[O]	448	<i>0.252</i>	<i>0.115</i>	<i>1.898</i>	<i>0.230</i>	<i>0.978</i>	<i><0.001</i>
	[W+O]	518	0.191	0.107	<i>1.700</i>	<i>0.214</i>	<i>0.981</i>	<i><0.001</i>
AT (score)	[T]	73	<i>0.827</i>	<i>0.281</i>	1.024	0.555	<i>0.941</i>	<i>0.002</i>
	[W]	70	0.165	0.287	0.651	0.566	0.984	0.538
	[O]	448	<i>0.249</i>	<i>0.115</i>	<i>1.667</i>	<i>0.230</i>	<i>0.980</i>	<i><0.001</i>
	[W+O]	518	<i>0.229</i>	<i>0.107</i>	<i>1.489</i>	<i>0.214</i>	<i>0.982</i>	<i><0.001</i>

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

TABLE A19.2: Statistical tests for normality of the crude change in all of the outcomes for the girls from M1 to M2 in the Gulu sample – stratified by intervention group

Outcome variable	Δ in group	n	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
BFA (z score)	[T]	79	0.005	0.271	0.160	0.535	0.981	0.285
	[O]	694	0.224	0.093	2.205	0.185	0.980	<0.001
SBJ (cm)	[T]	79	-0.413	0.271	0.121	0.535	0.981	0.270
	[O]	688	0.315	0.093	2.179	0.186	0.968	<0.001
MSFT (level complete)	[T]	79	-0.233	0.271	0.328	0.535	0.960	0.015
	[O]	681	0.358	0.094	0.604	0.187	0.968	<0.001
TT (score)	[T]	79	-0.352	0.271	0.925	0.535	0.982	0.336
	[O]	723	-0.158	0.091	0.689	0.182	0.989	<0.001
K (score)	[T]	79	-0.514	0.271	0.981	0.535	0.977	0.159
	[O]	723	-0.062	0.091	0.722	0.182	0.989	<0.001
P (score)	[T]	79	-0.750	0.271	0.903	0.535	0.963	0.021
	[O]	723	-0.230	0.091	1.060	0.182	0.988	<0.001
ML (score)	[T]	79	-0.635	0.271	0.158	0.535	0.963	0.021
	[O]	723	0.055	0.091	0.586	0.182	0.992	<0.001
DT (score)	[T]	79	-0.562	0.271	0.901	0.535	0.975	0.125
	[O]	723	-0.135	0.091	0.966	0.182	0.988	<0.001
AT (score)	[T]	79	-0.601	0.271	0.671	0.535	0.973	0.092
	[O]	723	-0.116	0.091	1.059	0.182	0.988	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12]. The assumption of HOV between the treatment groups for the crude results of each of the dependent variables was assessed using the Levene and Brown-Forsythe tests (TABLES A19.3 – A19.6). These statistical tests for HOV were robust to violations of the normality assumption ^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$).

TABLE A19.3: Statistical tests of the HOV between the crude change in groups [T] and [W] for all of the boys outcomes in the Gulu sample

Outcome variable	Δ in group [T]		Δ in group [W]		Levene test		Brown-Forsythe test		Summary
	n	Variance	n	Variance	F	p	F	p	
BFA (z score)	72	0.346	69	0.281	0.264	0.608	0.246	0.621	√
SBJ (cm)	70	391.68	68	426.73	0.039	0.845	0.072	0.790	√
MSFT (level complete)	70	6.55	69	4.96	3.562	0.061	2.636	0.107	√
TT (score)	73	63.49	70	50.05	0.240	0.625	0.069	0.794	√
K (score)	73	38.69	70	37.65	0.022	0.883	0.020	0.889	√
P (score)	73	47.80	70	44.92	0.073	0.787	0.065	0.798	√
ML (score)	73	29.22	70	24.64	0.089	0.766	0.016	0.899	√
DT (score)	73	230.19	70	199.87	0.151	0.698	0.084	0.772	√
AT (score)	73	332.97	70	253.38	0.437	0.510	0.387	0.535	√

TABLE A19.4: Statistical tests of the HOV between the crude change in groups [W] and [O] for all of the boys outcomes in the Gulu sample

Outcome variable	Δ in group [W]		Δ in group [O]		Levene test		Brown-Forsythe test		Summary
	n	Variance	n	Variance	F	p	F	p	
BFA (z score)	69	0.281	436	0.236	0.012	0.914	0.010	0.919	√
SBJ (cm)	68	426.73	429	346.09	0.280	0.597	0.220	0.639	√
MSFT (level complete)	69	4.96	429	5.24	1.320	0.251	0.880	0.349	√
TT (score)	70	50.05	448	47.41	0.267	0.605	0.211	0.646	√
K (score)	70	37.65	448	30.52	0.436	0.509	0.338	0.561	√
P (score)	70	44.92	448	49.83	0.320	0.572	0.333	0.564	√
ML (score)	70	24.64	448	28.37	0.081	0.775	0.075	0.784	√
DT (score)	70	199.87	448	175.52	0.374	0.541	0.352	0.553	√
AT (score)	70	253.38	448	243.72	0.161	0.689	0.162	0.687	√

NOTE: In the summary column - √ denotes satisfactory HOV^[12,14].

TABLE 19.5: Statistical tests of the HOV between the crude change in groups [T] and [W+O] for all of the boys outcomes in the Gulu sample

Outcome variable	Δ in group [T]		Δ in group [W+O]		Levene test		Brown-Forsythe test		Summary
	n	Variance	n	Variance	F	p	F	p	
BFA (z score)	72	0.346	505	0.243	0.754	0.386	0.698	0.404	√
SBJ (cm)	70	391.68	497	357.07	0.038	0.845	0.000	0.994	√
MSFT (level complete)	70	6.55	498	5.20	2.578	0.109	2.055	0.152	√
TT (score)	73	63.49	518	48.05	1.306	0.254	0.585	0.445	√
K (score)	73	38.69	518	31.73	0.108	0.743	0.066	0.797	√
P (score)	73	47.80	518	49.73	0.052	0.820	0.060	0.806	√
ML (score)	73	29.22	518	27.89	0.017	0.898	0.006	0.939	√
DT (score)	73	230.19	518	180.50	1.052	0.305	0.756	0.385	√
AT (score)	73	332.97	518	246.69	1.561	0.212	1.448	0.229	√

TABLE A19.6: Statistical tests of the HOV between the crude change in groups [T] and [O] for all of the girls outcomes in the Gulu sample

Outcome variable	Δ in group [T]		Δ in group [O]		Levene test		Brown-Forsythe test		Summary
	n	Variance	n	Variance	F	p	F	p	
BFA (z score)	79	0.105	694	0.146	1.598	0.207	1.553	0.213	√
SBJ (cm)	79	175.24	688	354.02	5.530	0.019	5.554	0.019	√
MSFT (level complete)	79	2.79	681	4.87	3.478	0.063	3.855	0.050	√
TT (score)	79	46.78	723	63.31	1.652	0.199	1.685	0.195	√
K (score)	79	30.81	723	40.80	1.700	0.193	1.946	0.163	√
P (score)	79	39.73	723	55.69	2.077	0.150	2.269	0.132	√
ML (score)	79	30.21	723	29.40	0.071	0.790	0.004	0.948	√
DT (score)	79	158.88	723	232.85	2.530	0.112	2.638	0.105	√
AT (score)	79	219.13	723	310.88	1.894	0.169	2.106	0.147	√

NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

The assumption of HOV held for all of the dependent variables when considering between-group differences in the cohort study and RCT. The statistical analysis could proceed with little concern when the HOV assumption was satisfied ^[12,14].

APPENDIX A20: Accelerometry – MVPA Between-Group Differences (Assumptions)

The statistical methods used to assess whether there was any difference between the accelerometer sub-samples of groups [T] vs. [W] for MVPA assumed an approximately normal distribution. This was initially assessed visually using a histogram of the crude results for each of the time frames of interest (FIGURES A20.1 – A20.3).

FIGURE A20.1: Histograms of the crude daily average MVPA on valid days for the boys participating in the accelerometer study in Gulu – stratified by intervention group

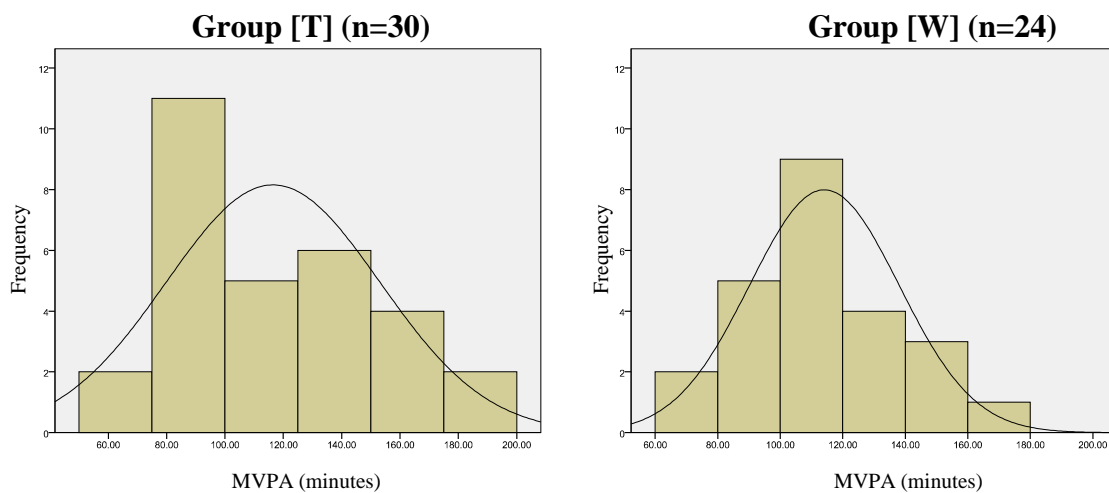


FIGURE A20.2: Histograms of the crude daily average MVPA during the GMKL training times (4-7pm) for the boys participating in the accelerometer study in Gulu – stratified by intervention group

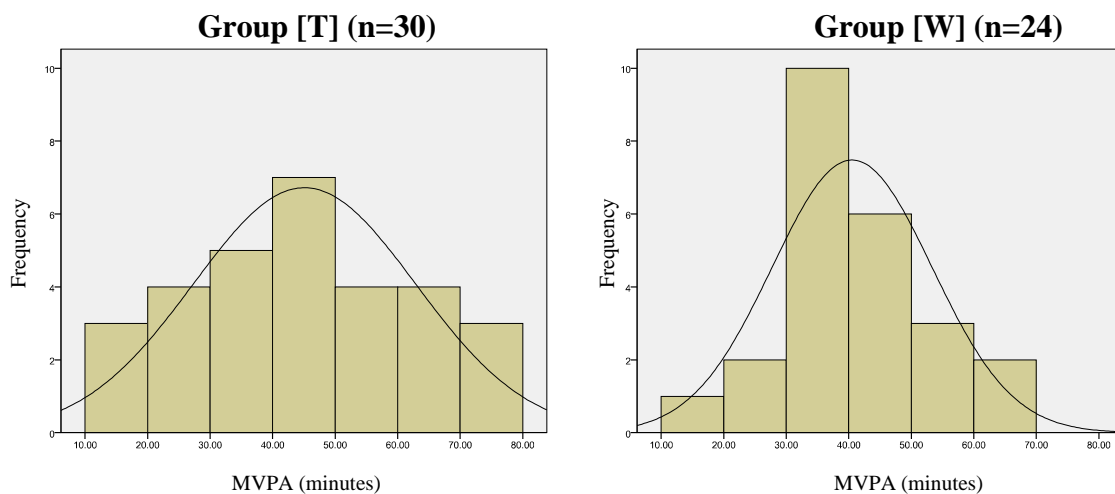
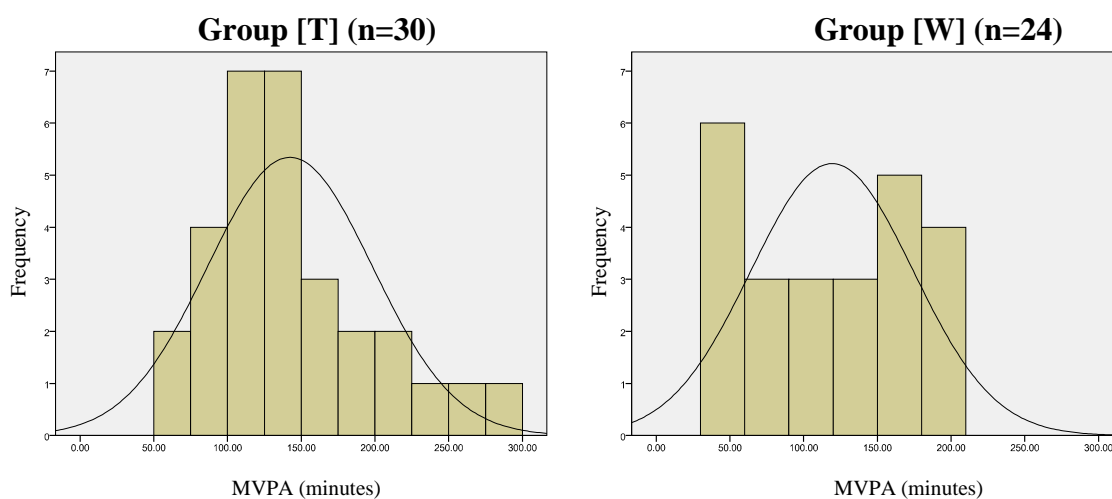


FIGURE A20.3: Histograms of the crude average MVPA on the GMKL activity day for the boys participating in the accelerometer study in Gulu – stratified by intervention group



Inspection of the histograms warranted further assessment of normality and statistical tests of the crude results were also completed (TABLE A20.1)

TABLE A20.1: Statistical tests for normality of the crude MVPA measures during the GMKL in Gulu - stratified by intervention group

Group	n	MVPA (minutes)	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
[T]	30	Daily average on valid days	0.341	0.427	-0.791	0.833	0.960	0.317
		Daily average during the GMKL training times (4-7pm)	0.155	0.427	-0.689	0.833	0.974	0.643
		Average on the GMKL activity day	0.915	0.427	0.481	0.833	0.929	0.045
[W]	24	Daily average on valid days	0.337	0.472	-0.314	0.918	0.974	0.758
		Daily average during the GMKL training times (4-7pm)	0.547	0.472	0.038	0.918	0.959	0.421
		Average on the GMKL activity day	0.040	0.472	-1.530	0.918	0.907	0.030

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12]. The assumption of HOV between the treatment groups for the crude results of each of the dependent variables was assessed using the Levene and Brown-Forsythe tests (TABLE A20.2). These statistical tests for HOV were robust to violations of the normality assumption ^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$).

TABLE A20.2: Statistical tests for the HOV between groups [T] and [W] for the crude MVPA results during the GMKL in Gulu

MVPA	Group [T]		Group [W]		Levene test		Brown-Forsythe		Summary
	n	Var.	n	Var.	F	p	F	p	
Daily average on valid days	30	1345.8	24	573.7	7.804	0.007	6.664	0.013	(√)
Daily average during the GMKL training times (4-7pm)	30	317.1	24	163.8	3.095	0.084	2.764	0.102	√
Average on the GMKL activity day	30	3137.0	24	3025.8	0.391	0.535	0.646	0.425	√

NOTE: Results in **BOLD ITALICS** suggest a statistically significant heterogeneity of variance ($p < 0.01$).

NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

The Levene test for the daily average of total MVPA suggested a statistically significant heterogeneity of variance between groups [T] vs. [W]. However, the Brown-Forsythe test is more robust when the data from both samples have a similar skew and indicated that the assumption of HOV held for all of the time frames of interest ^[14]. The statistical analysis could proceed with little concern when the HOV assumption was satisfied ^[12,14].

APPENDIX A21: Accelerometry – VPA Between-Group Differences (Assumptions)

The statistical methods used to assess whether there was any difference between the accelerometer sub-samples of groups [T] vs. [W] for VPA assumed an approximately normal distribution. This was initially assessed visually using a histogram for the crude results of each of the time frames of interest (FIGURES A21.1 – A21.3).

FIGURE A21.1: Histograms of the crude daily average VPA on valid days for the boys participating in the accelerometer study in Gulu – stratified by intervention group

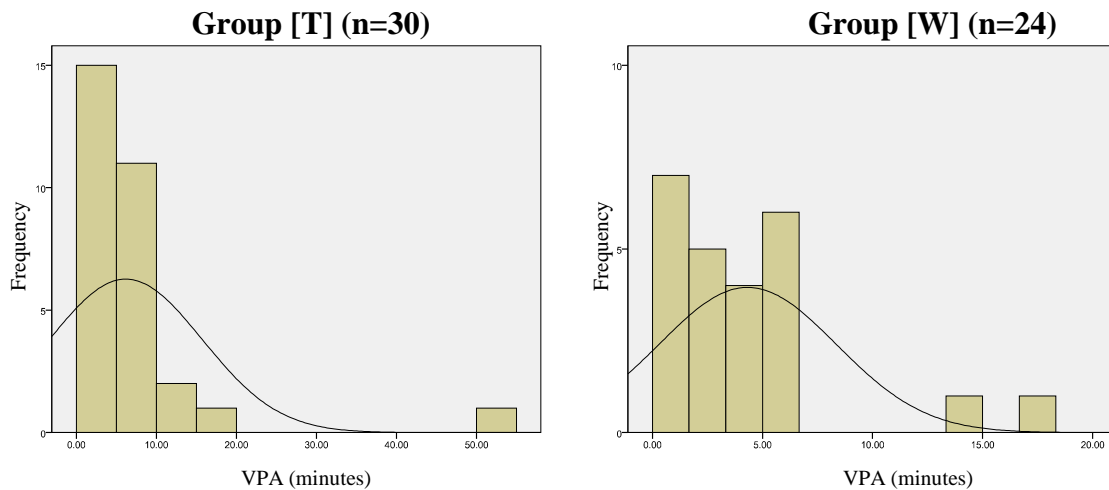


FIGURE A21.2: Histograms of the crude daily average VPA during the GMKL training times (4-7pm) for the boys participating in the accelerometer study in Gulu – stratified by intervention group

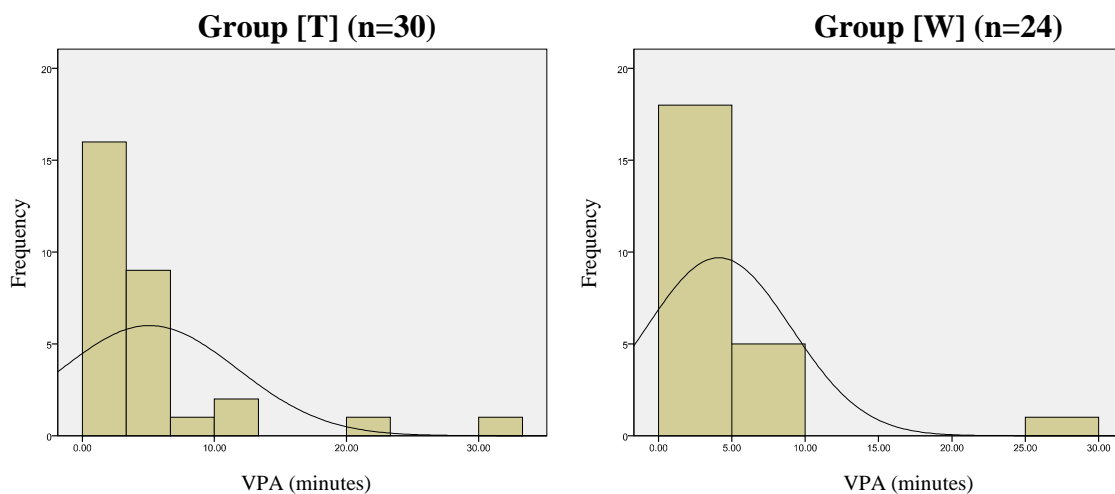
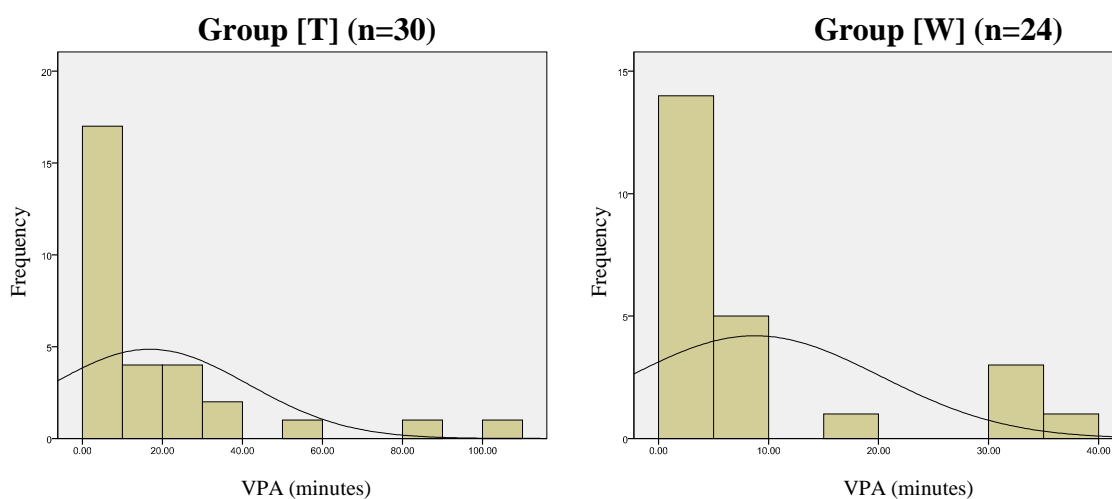


FIGURE A21.3: Histograms of the crude average VPA on the GMKL activity day for the boys participating in the accelerometer study in Gulu – stratified by intervention group



Inspection of the histograms indicated that the distributions were not normally distributed.

This was supported by statistical tests for normality of the crude results (TABLE A21.1).

TABLE A21.1: Statistical tests for normality of the crude VPA measures during the GMKL in Gulu - stratified by intervention group

Group	n	VPA (minutes)	Skew		Kurtosis		Shapiro Wilk	
			Test stat	SES	Test stat	SEK	Test stat	p
[T]	30	Daily average on valid days	4.022	0.427	18.692	0.833	0.533	<0.001
		Daily average during the GMKL training times (4-7pm)	2.830	0.427	9.289	0.833	0.670	<0.001
		Average on the GMKL activity day	2.390	0.427	5.788	0.833	0.677	<0.001
[W]	24	Daily average on valid days	2.018	0.472	4.561	0.918	0.761	<0.001
		Daily average during the GMKL training times (4-7pm)	3.526	0.472	14.736	0.918	0.604	<0.001
		Average on the GMKL activity day	1.627	0.472	1.267	0.918	0.707	<0.001

NOTE: If a sample is normally distributed the skew/kurtosis statistic should be within $\pm 2.0 \times$ SE of skew/kurtosis (SES or SEK) ^[10].

NOTE: The Shapiro-Wilk test has been used because $n < 2000$ (significance value $p < 0.05$) ^[11].

NOTE: Results in **BOLD ITALICS** suggest a statistically significant non-normal distribution of the sample.

For the statistical analyses performed, violation of the assumption of normality was not of major concern provided the groups being compared had approximately equal variance ^[12]. The assumption of HOV between the treatment groups for the crude results of each of the dependent variables was assessed using the Levene and Brown-Forsythe tests (TABLE A21.2). These statistical tests for HOV were robust to violations of the normality assumption ^[14]. Moderate violations of the HOV assumption were of minimal concern and the α value was adjusted accordingly ($p < 0.01$).

TABLE A21.2: Statistical tests for the HOV between groups [T] and [W] for the crude VPA results during the GMKL in Gulu

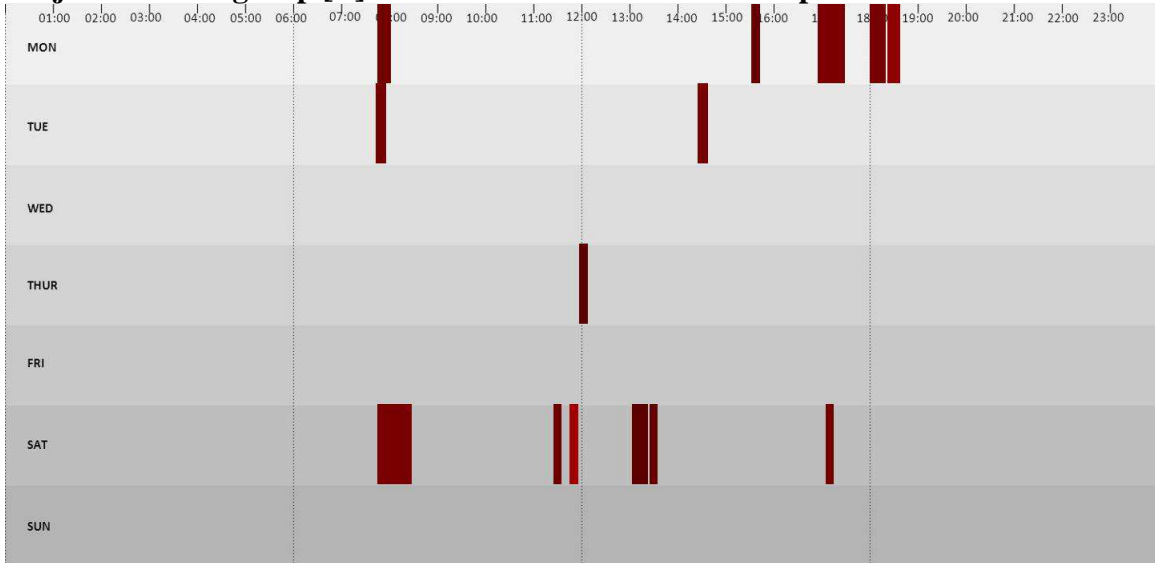
VPA (minutes)	Group [T]		Group [W]		Levene test		Brown-Forsythe		Summary
	n	Var.	n	Var.	F	p	F	p	
Daily average on valid days	30	91.14	24	16.29	1.694	0.199	0.795	0.377	√
Daily average during the GMKL training times (4-7pm)	30	44.27	24	24.42	0.935	0.338	0.364	0.549	√
Average on the GMKL activity day	30	607.10	24	130.24	4.557	0.038	2.208	0.143	√

NOTE: In the summary column - √ denotes satisfactory HOV ^[12,14].

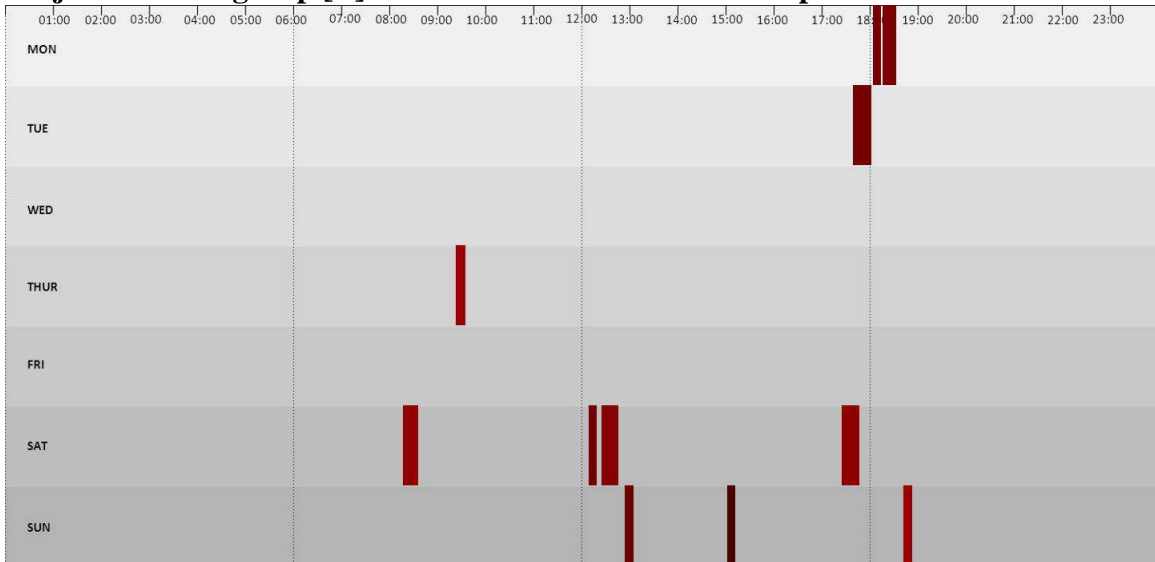
The assumption of HOV held for all of the time frames when considering between-group differences for VPA using accelerometry. The statistical analysis could proceed with little concern when the HOV assumption was satisfied ^[12,14].

APPENDIX A22: “Heat Maps” for 10-Minute Bouts of MVPA (Results)

Subject T1 from group [T] in Gulu accelerometer sub-sample: week 1 – device 1



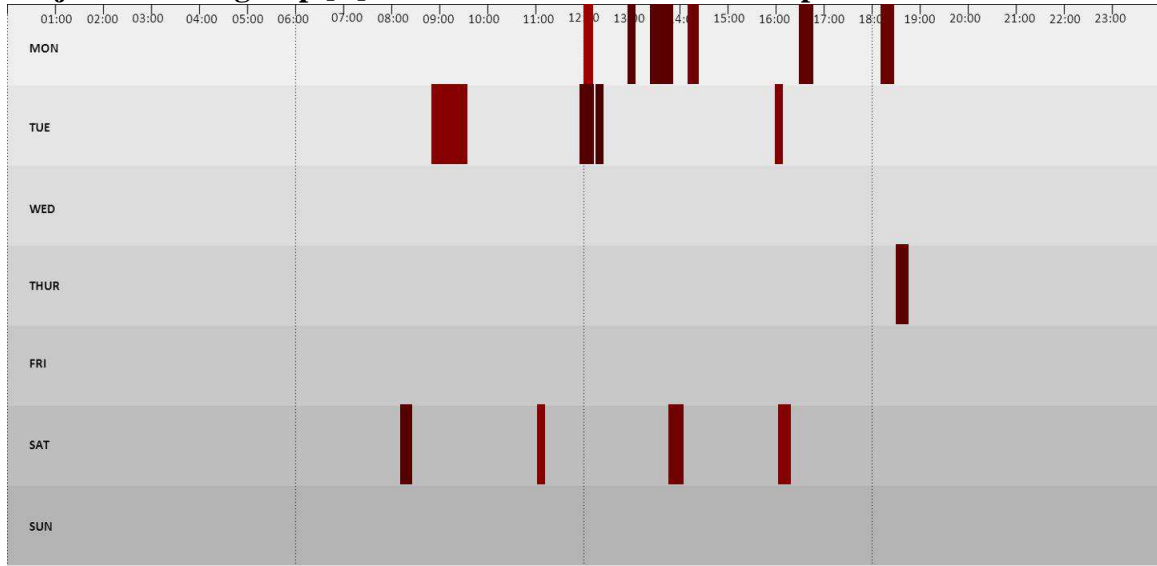
Subject T2 from group [T] in Gulu accelerometer sub-sample: week 1 – device 2



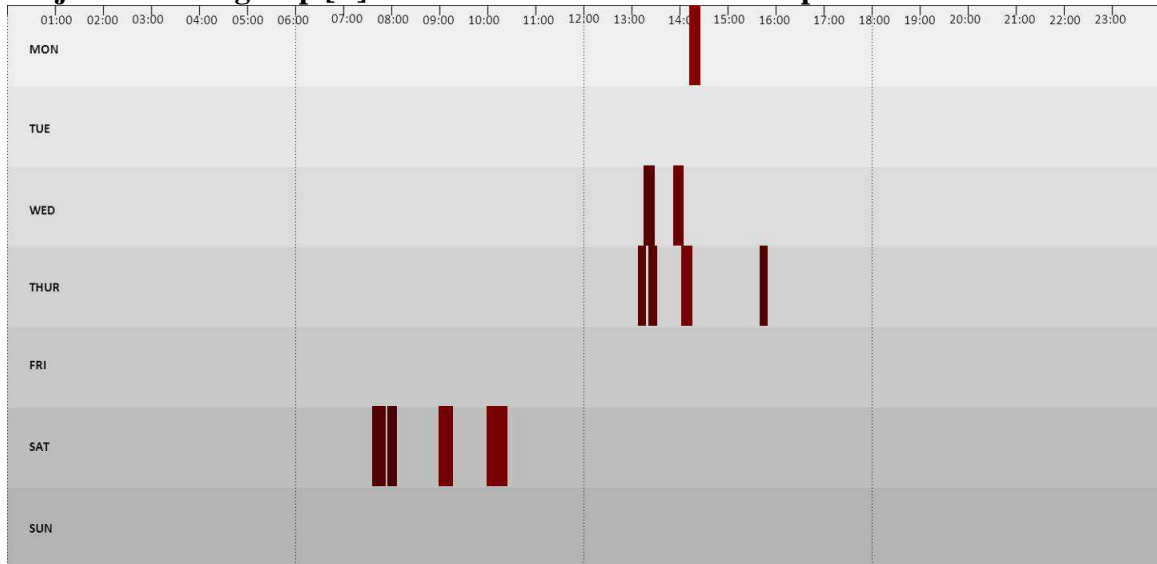
Subject T3 from group [T] in Gulu accelerometer sub-sample: week 1 – device 3



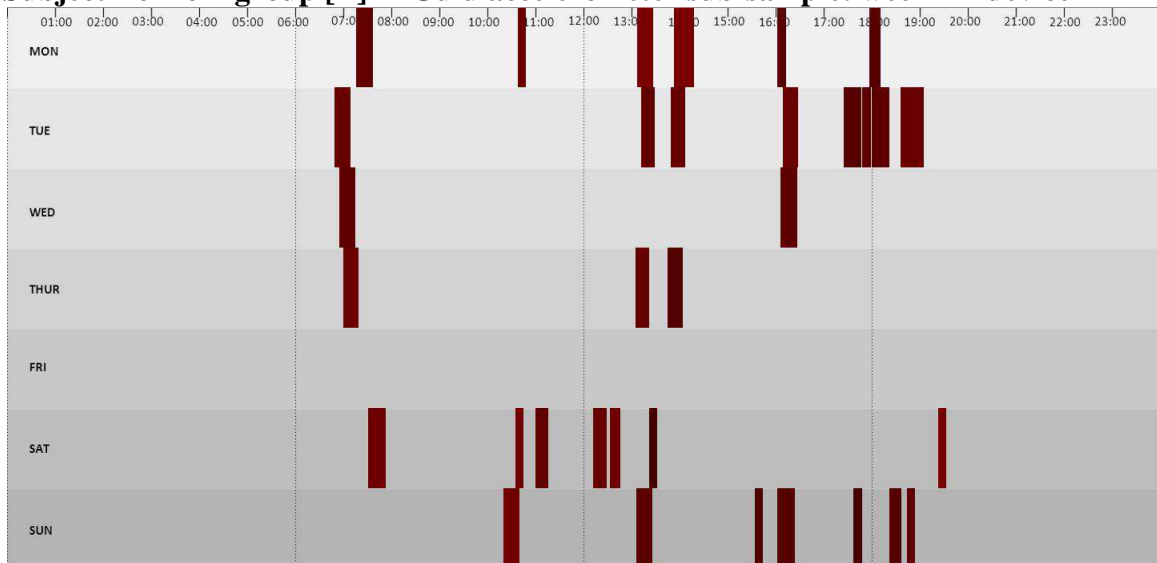
Subject T4 from group [T] in Gulu accelerometer sub-sample: week 1 – device 4



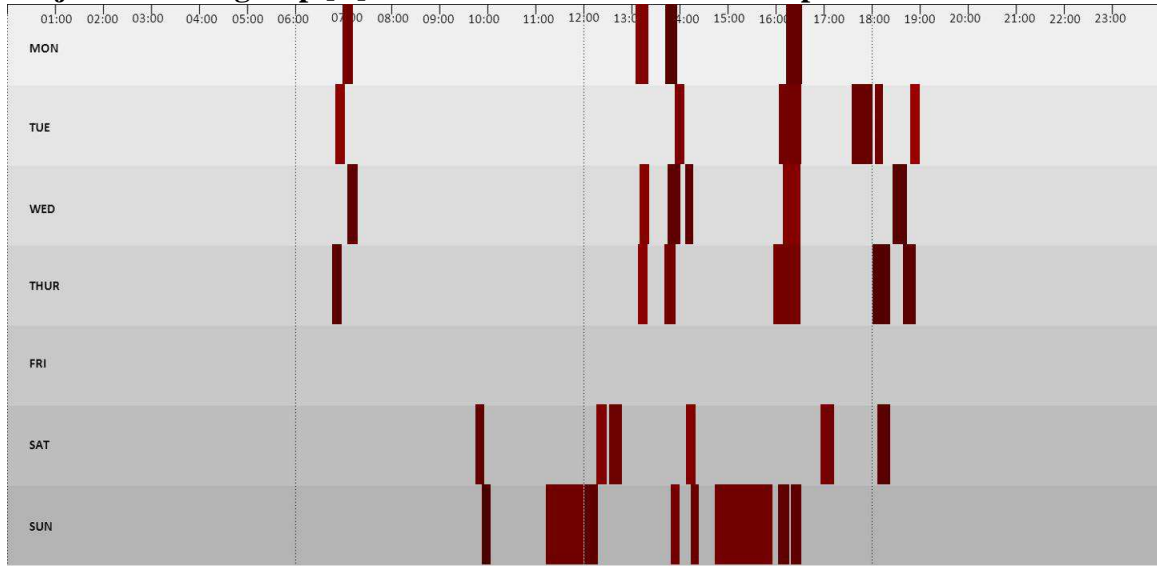
Subject T5 from group [T] in Gulu accelerometer sub-sample: week 2 – device 1



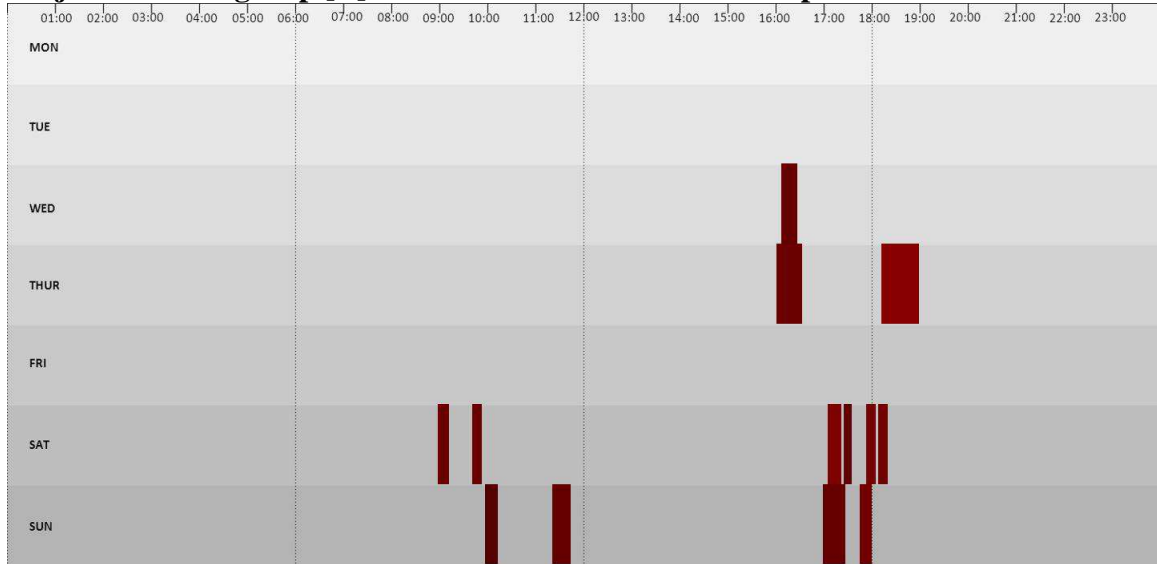
Subject T6 from group [T] in Gulu accelerometer sub-sample: week 2 – device 2



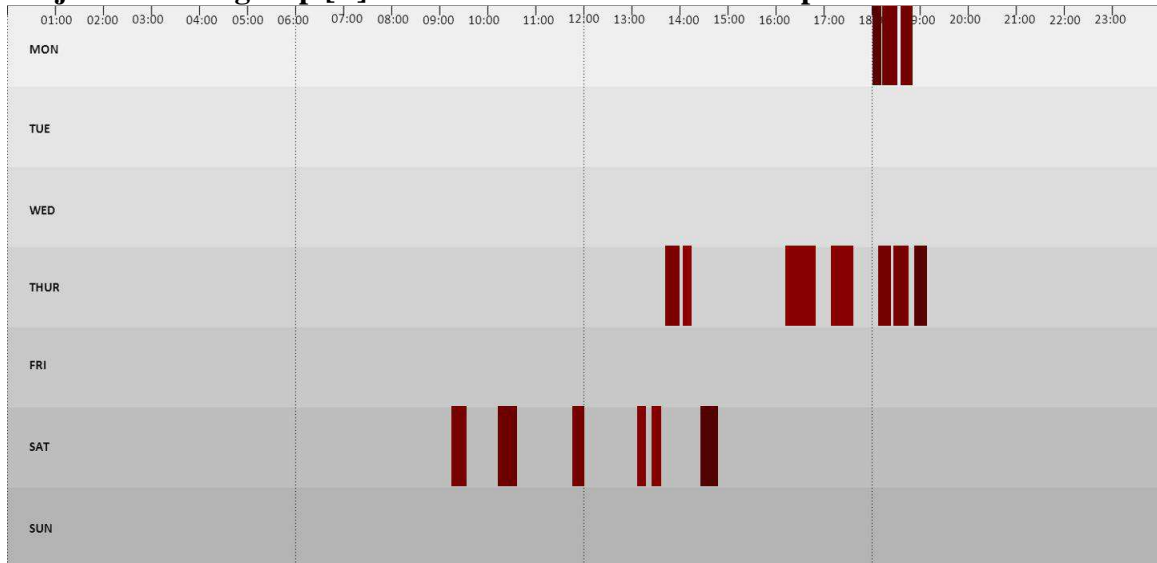
Subject T7 from group [T] in Gulu accelerometer sub-sample: week 2 – device 3



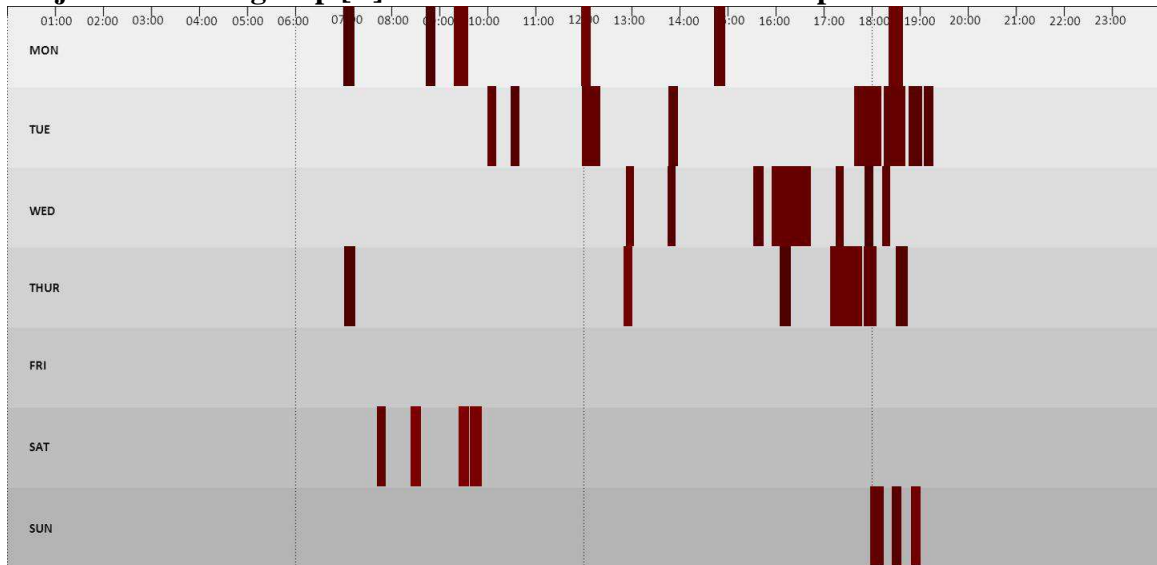
Subject T8 from group [T] in Gulu accelerometer sub-sample: week 2 – device 5



Subject T9 from group [T] in Gulu accelerometer sub-sample: week 2 – device 6



Subject T10 from group [T] in Gulu accelerometer sub-sample: week 2 – device 7



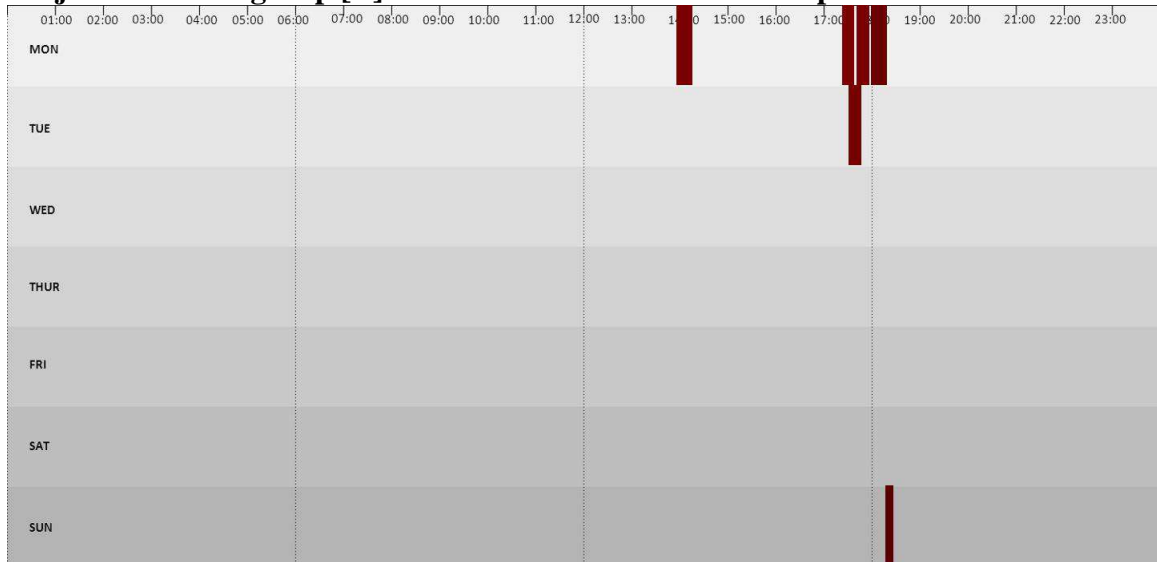
Subject T11 from group [T] in Gulu accelerometer sub-sample: week 3 – device 2



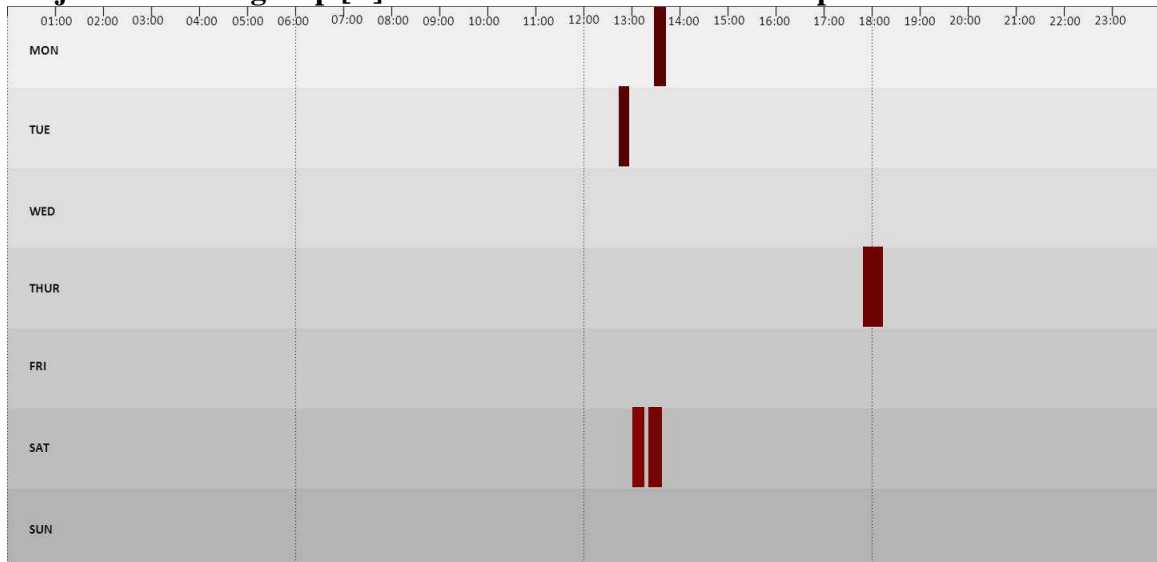
Subject T12 from group [T] in Gulu accelerometer sub-sample: week 3 – device 3



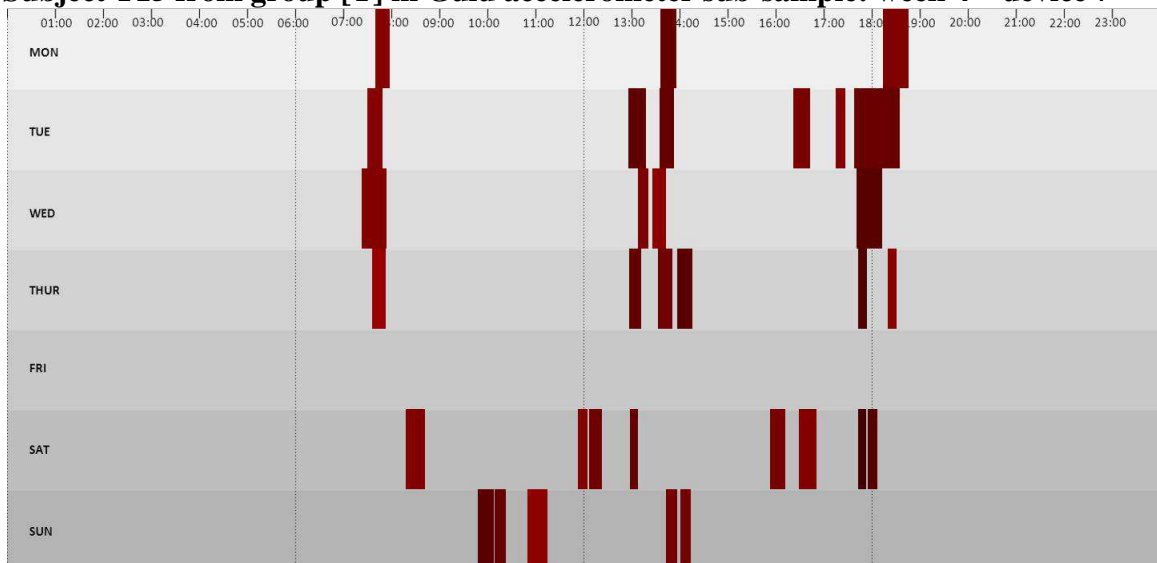
Subject T13 from group [T] in Gulu accelerometer sub-sample: week 3 – device 5



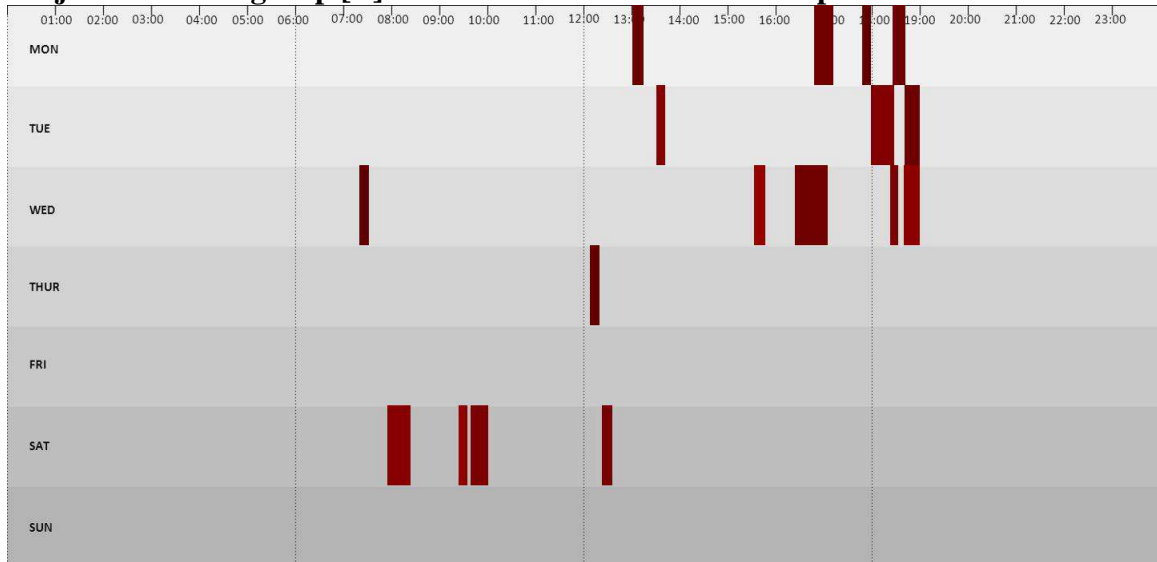
Subject T14 from group [T] in Gulu accelerometer sub-sample: week 4 – device 3



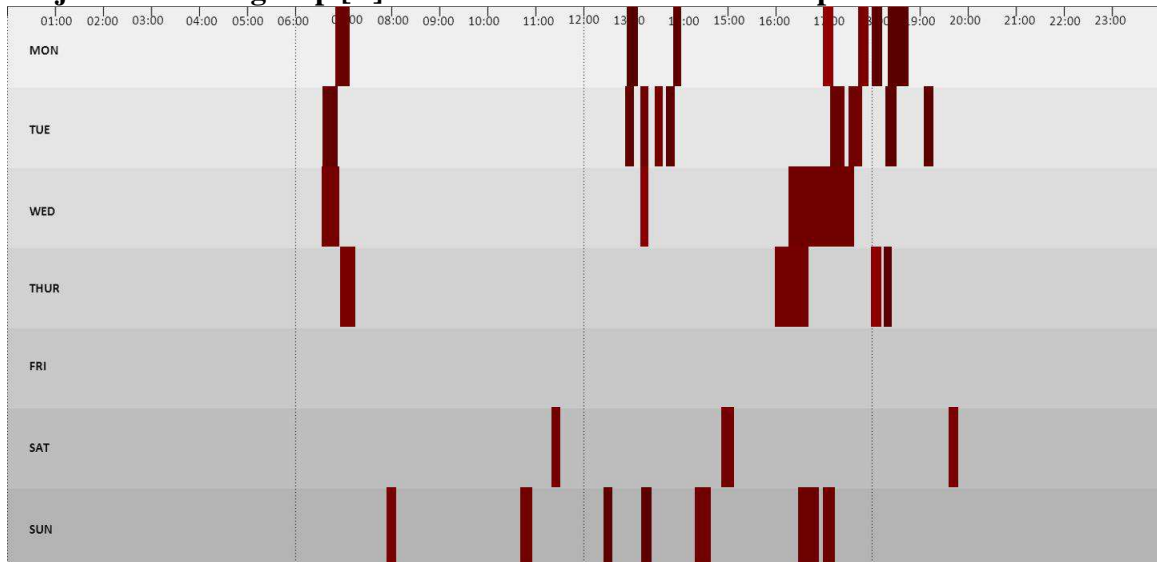
Subject T15 from group [T] in Gulu accelerometer sub-sample: week 4 – device 7



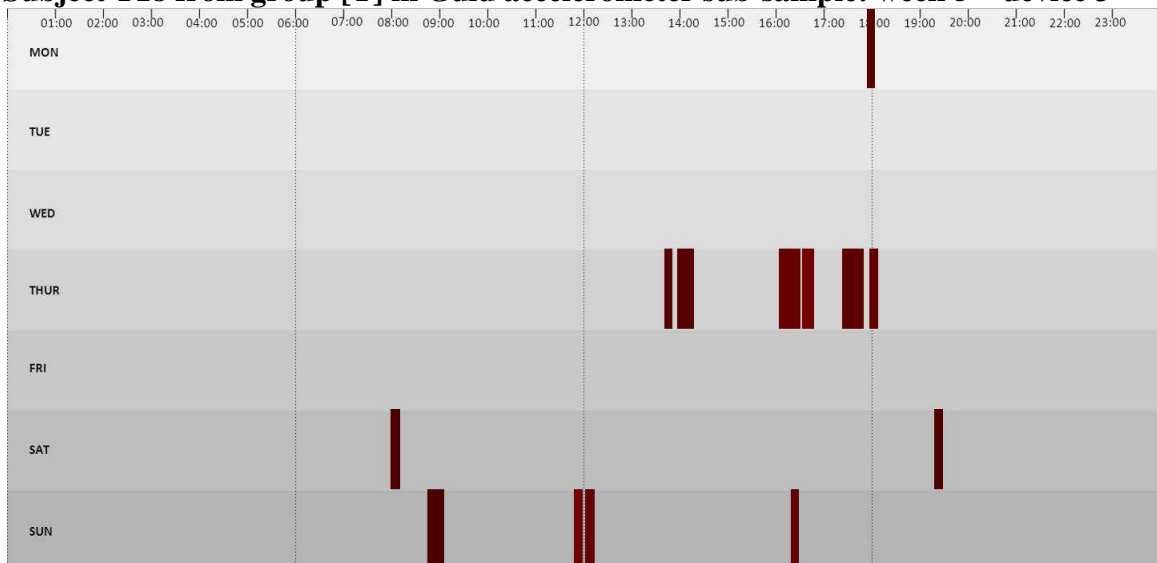
Subject T16 from group [T] in Gulu accelerometer sub-sample: week 5 – device 1



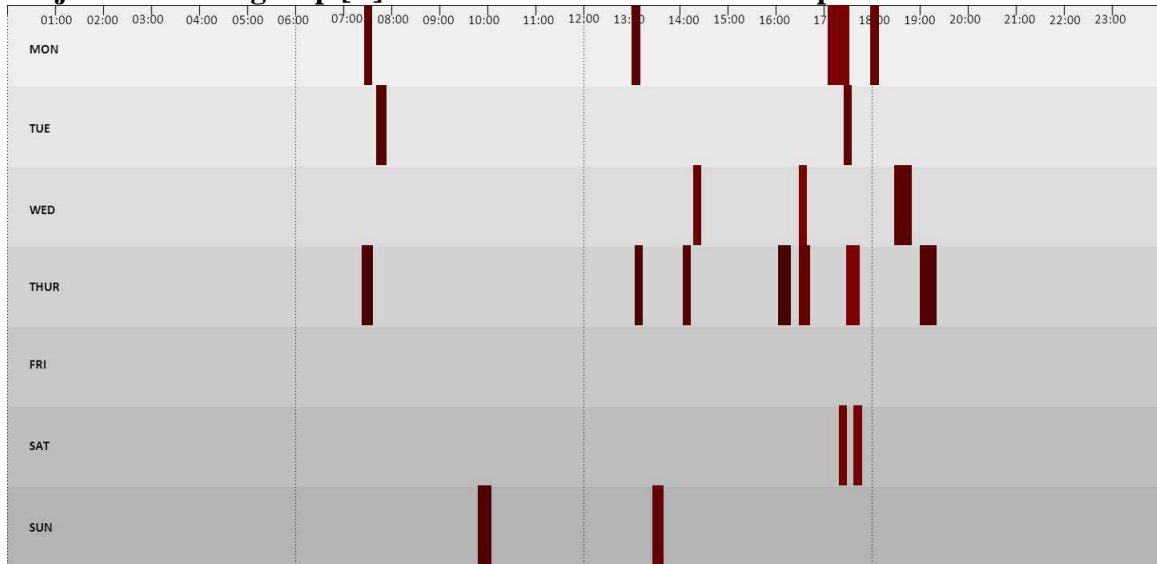
Subject T17 from group [T] in Gulu accelerometer sub-sample: week 5 – device 2



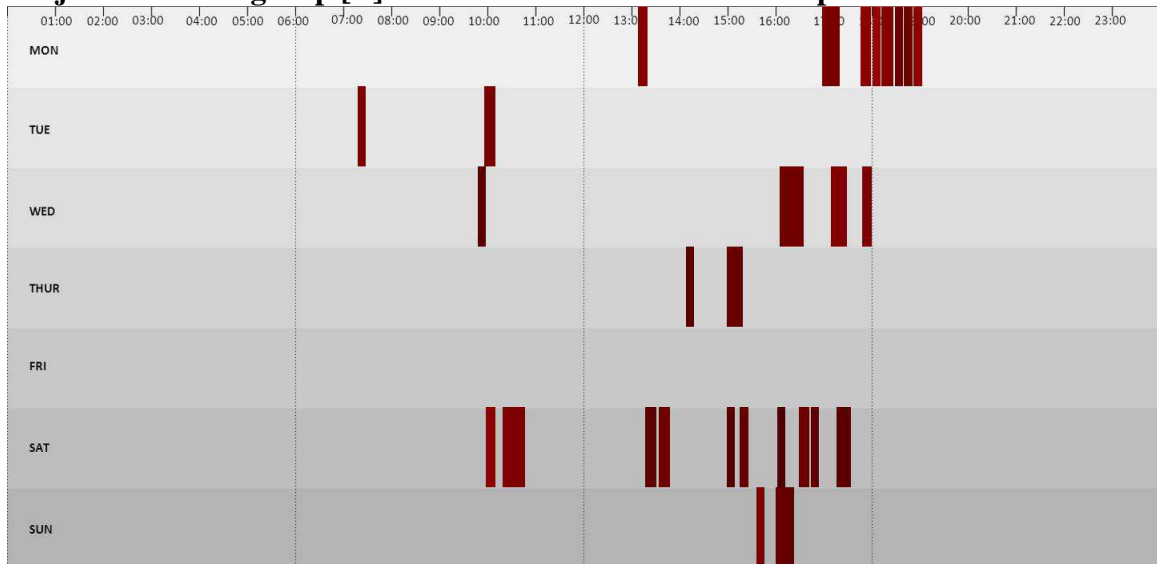
Subject T18 from group [T] in Gulu accelerometer sub-sample: week 5 – device 3



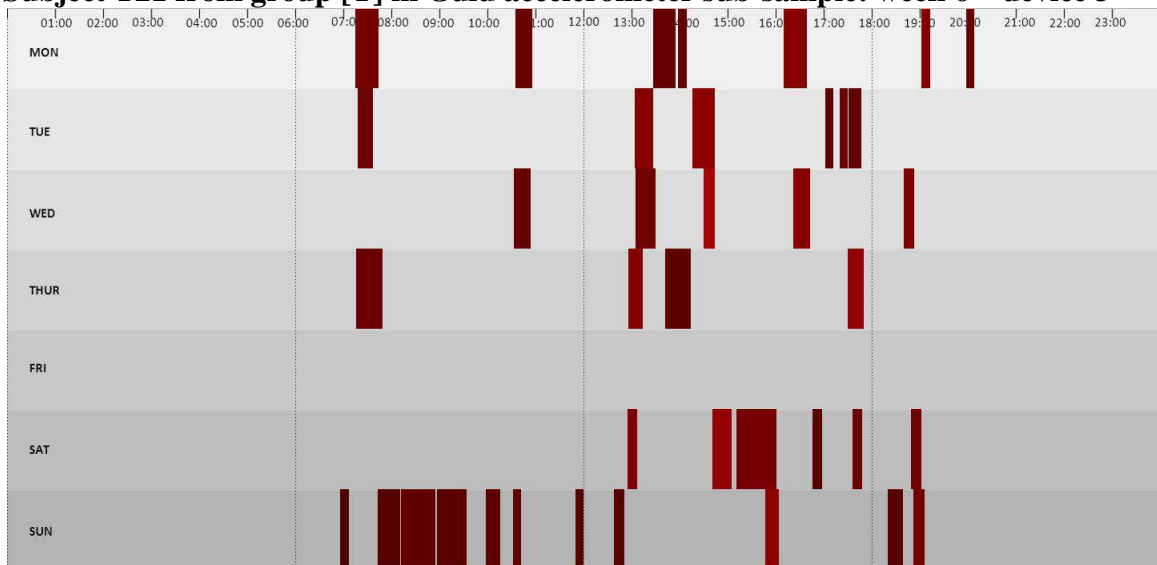
Subject T19 from group [T] in Gulu accelerometer sub-sample: week 6 – device 1



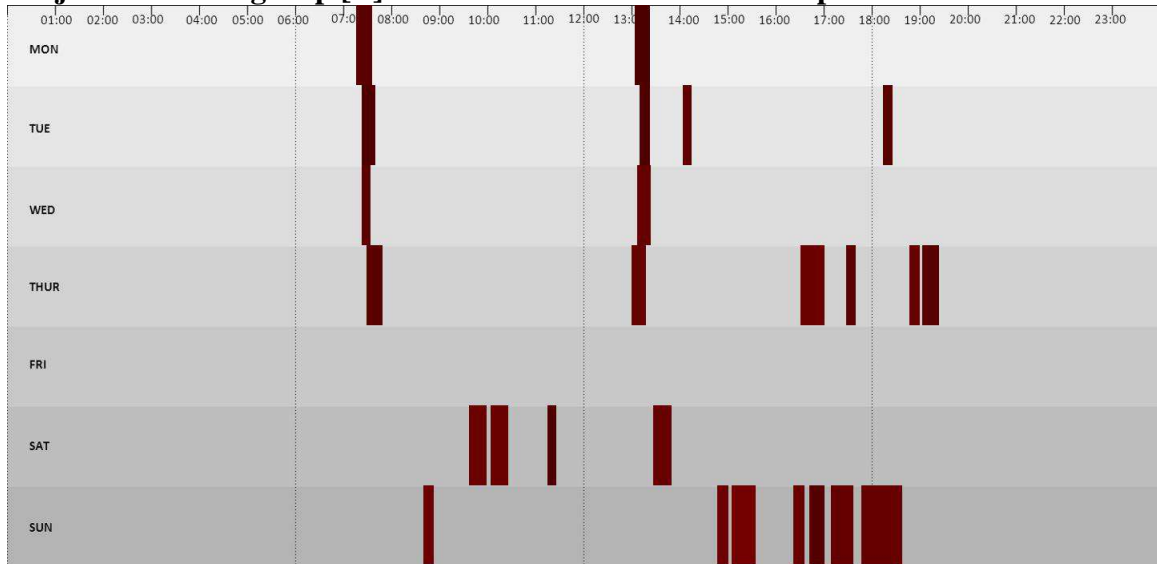
Subject T20 from group [T] in Gulu accelerometer sub-sample: week 6 – device 2



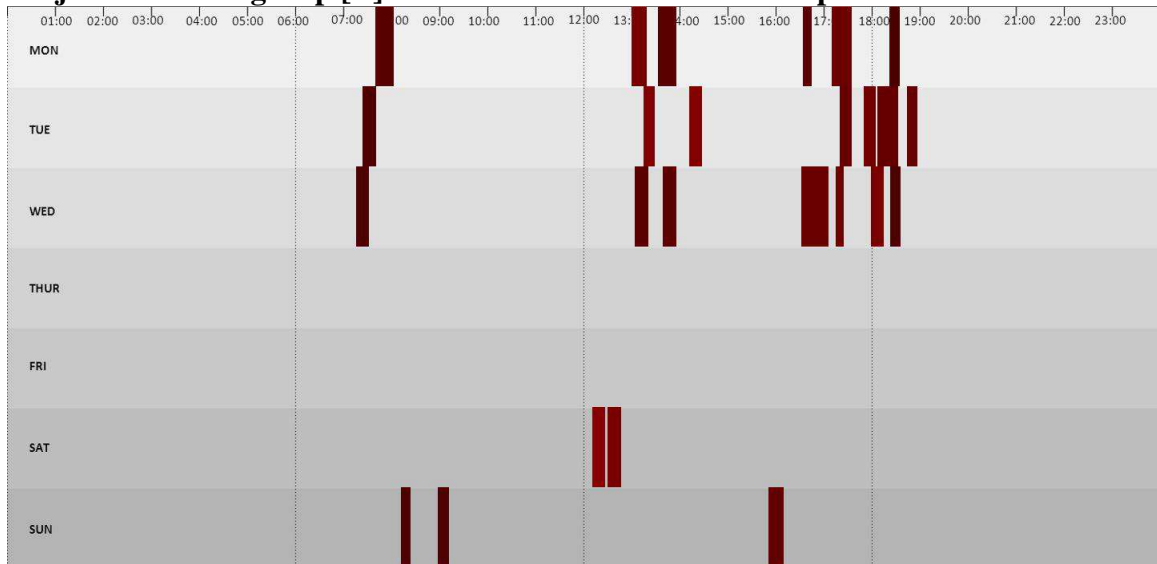
Subject T21 from group [T] in Gulu accelerometer sub-sample: week 6 – device 3



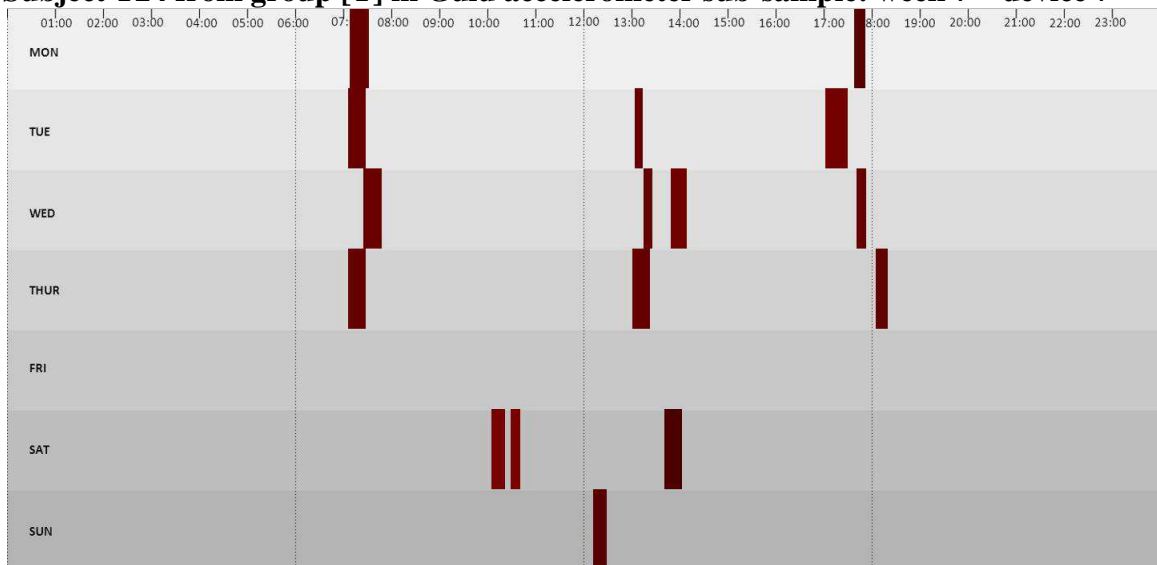
Subject T22 from group [T] in Gulu accelerometer sub-sample: week 7 – device 3



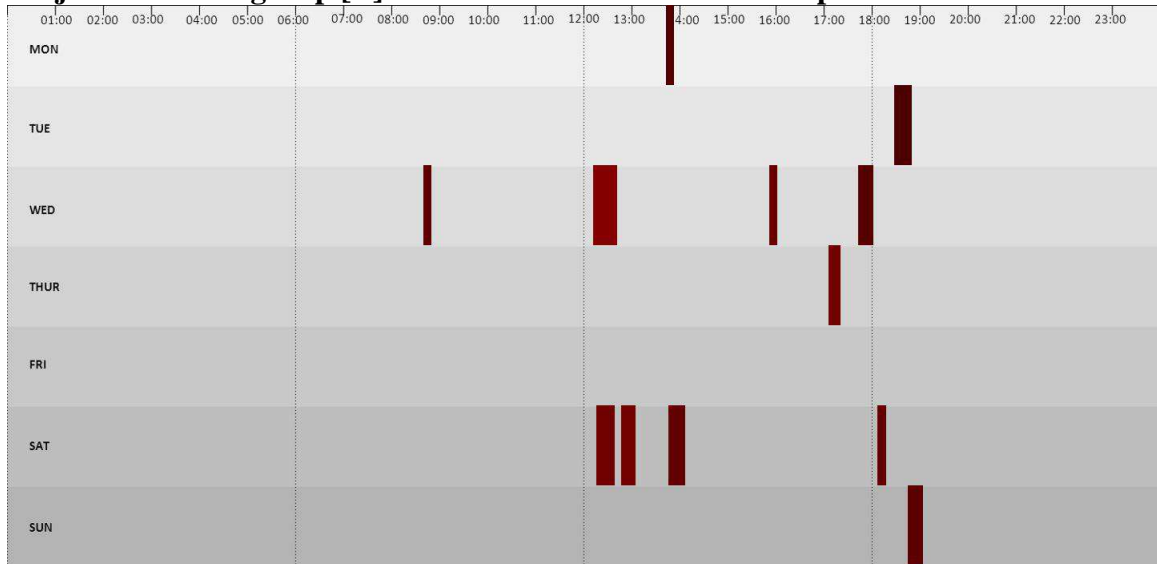
Subject T23 from group [T] in Gulu accelerometer sub-sample: week 7 – device 4



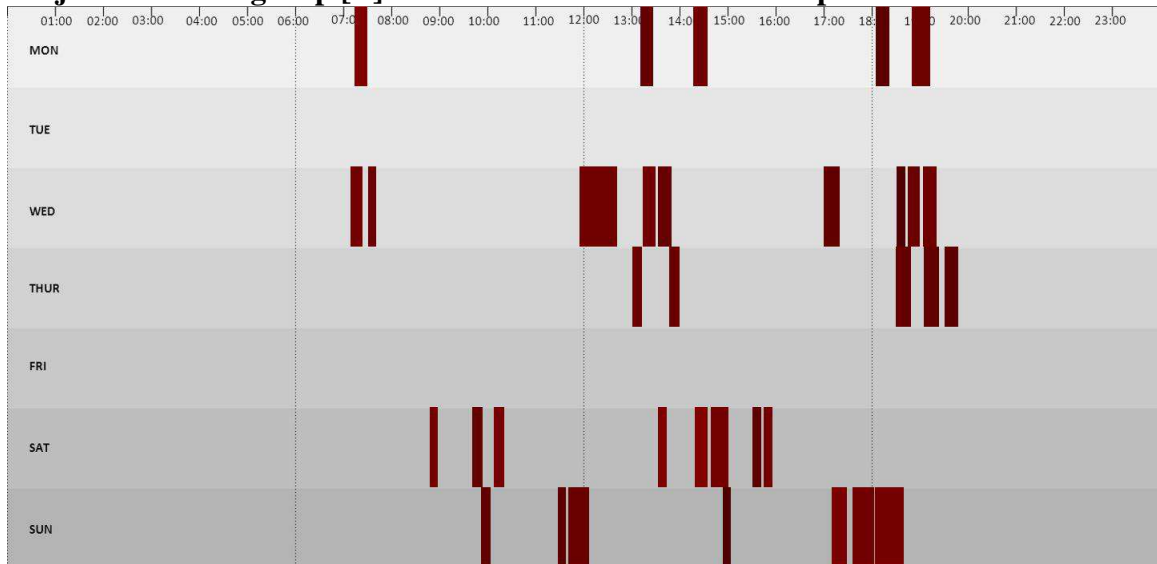
Subject T24 from group [T] in Gulu accelerometer sub-sample: week 7 – device 7



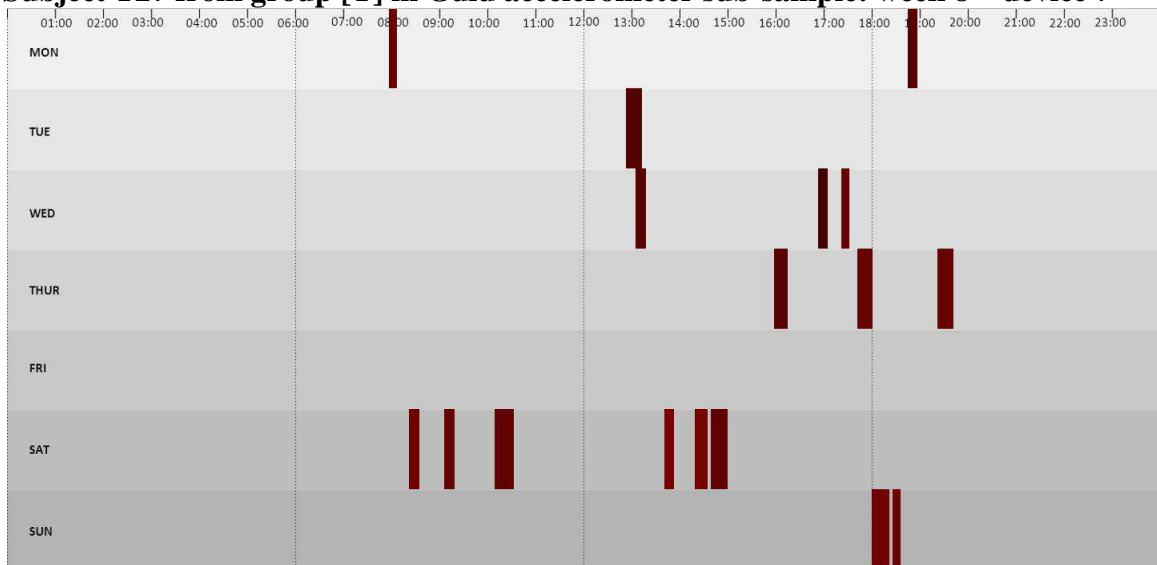
Subject T25 from group [T] in Gulu accelerometer sub-sample: week 8 – device 3



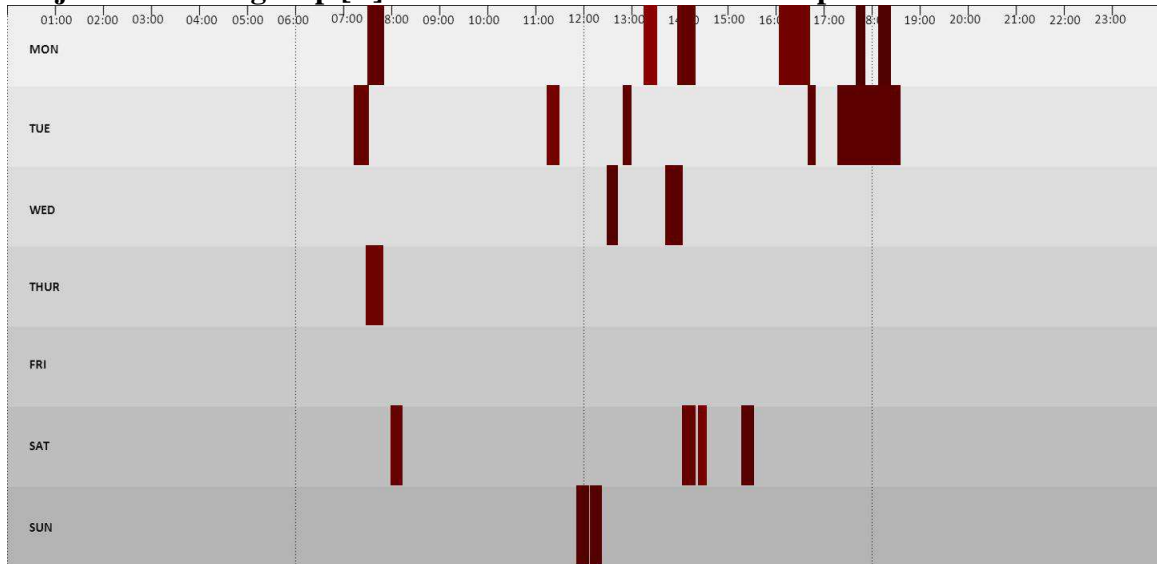
Subject T26 from group [T] in Gulu accelerometer sub-sample: week 8 – device 6



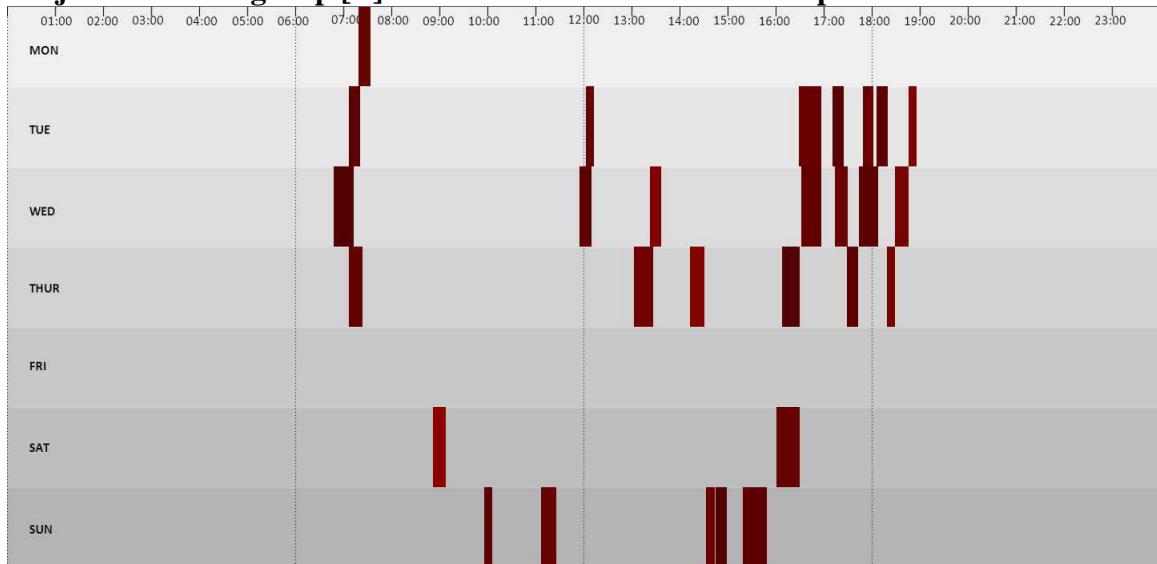
Subject T27 from group [T] in Gulu accelerometer sub-sample: week 8 – device 7



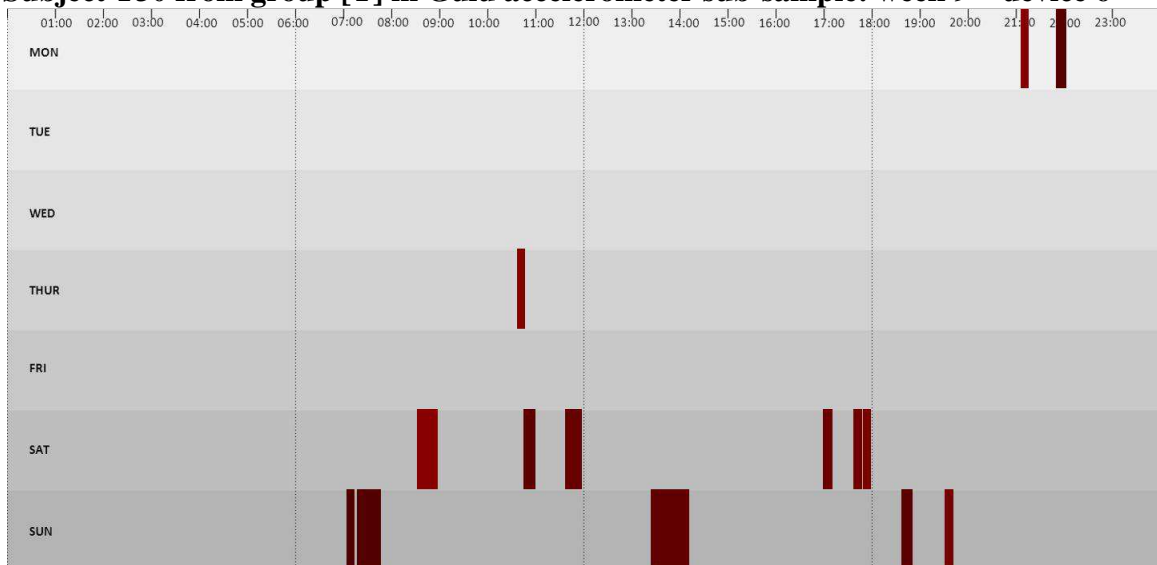
Subject T28 from group [T] in Gulu accelerometer sub-sample: week 9 – device 1



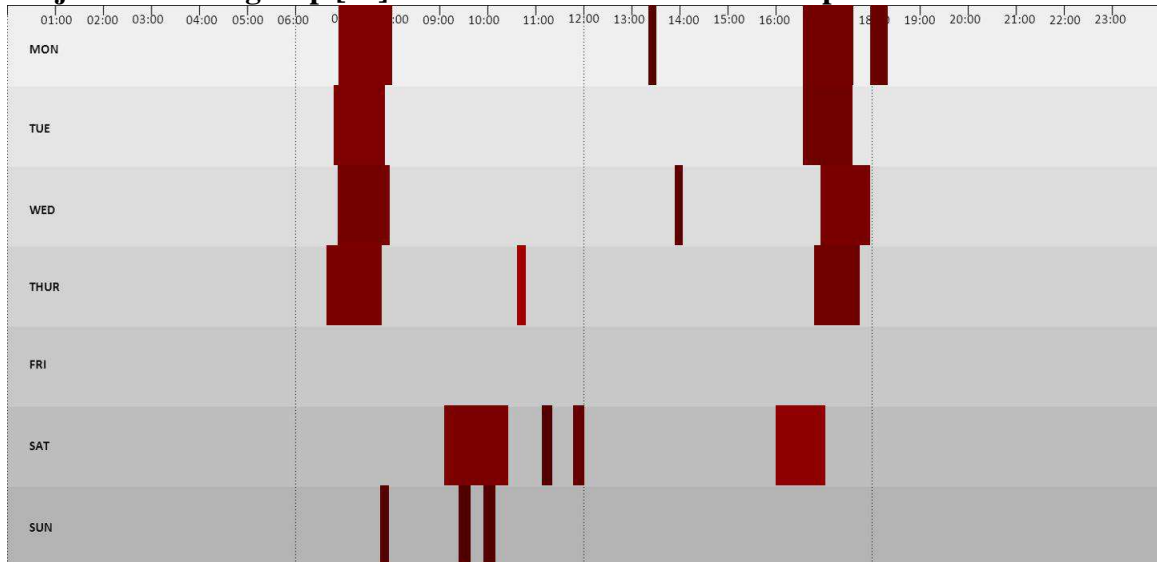
Subject T29 from group [T] in Gulu accelerometer sub-sample: week 9 – device 3



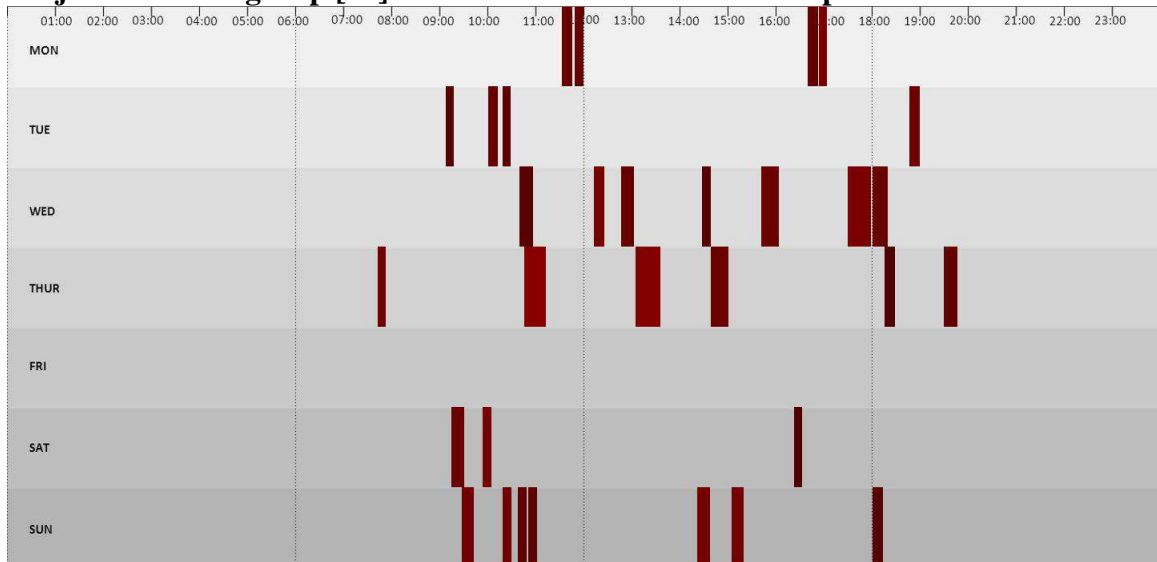
Subject T30 from group [T] in Gulu accelerometer sub-sample: week 9 – device 6



Subject W1 from group [W] in Gulu accelerometer sub-sample: week 1 – device 6



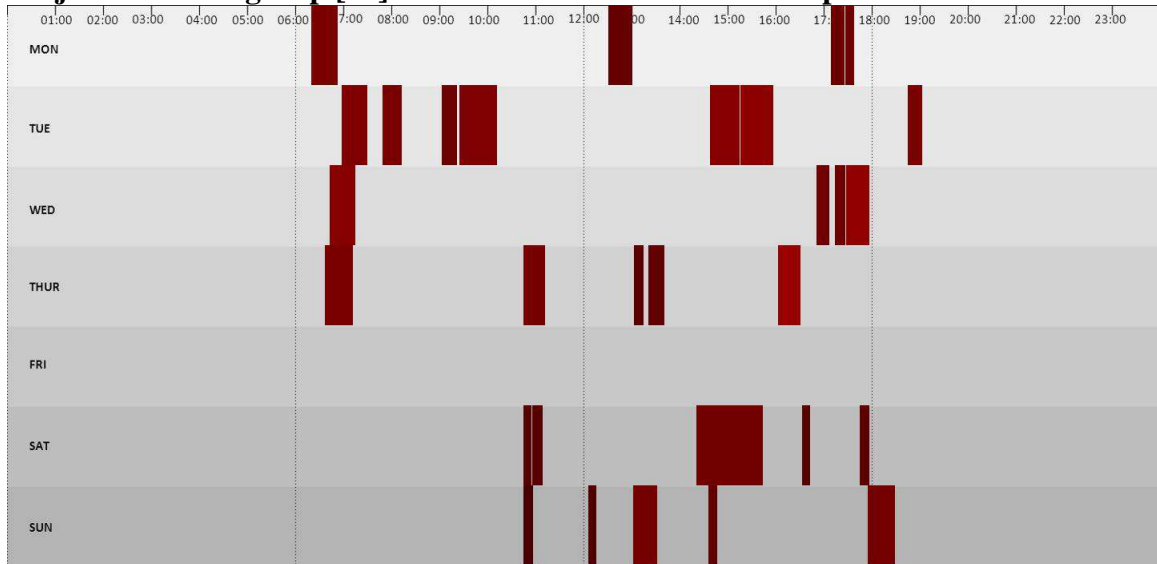
Subject W2 from group [W] in Gulu accelerometer sub-sample: week 1 – device 7



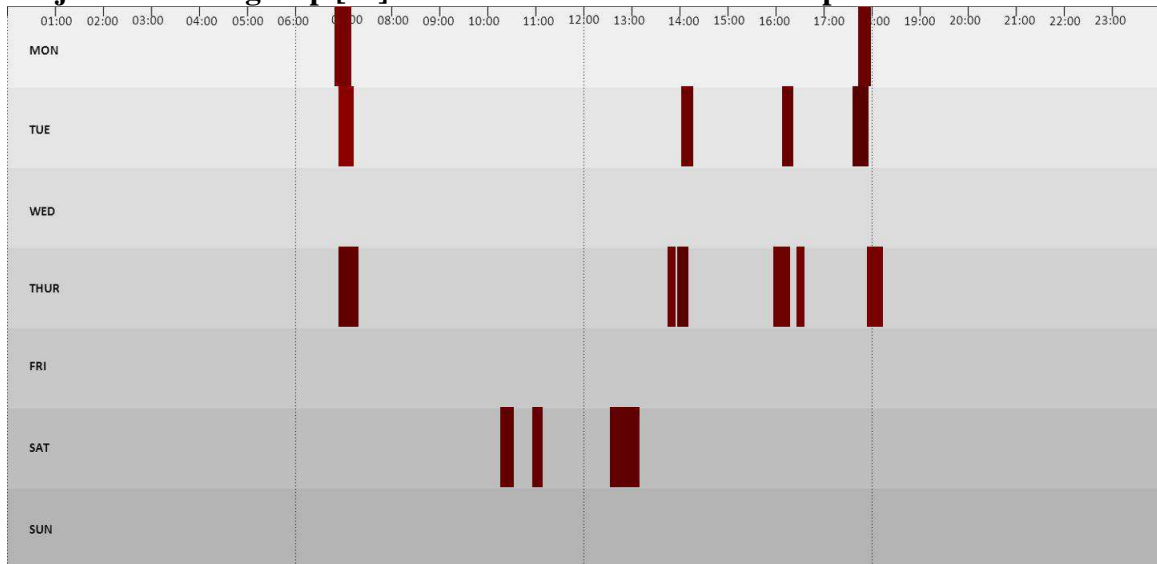
Subject W3 from group [W] in Gulu accelerometer sub-sample: week 2 – device 4



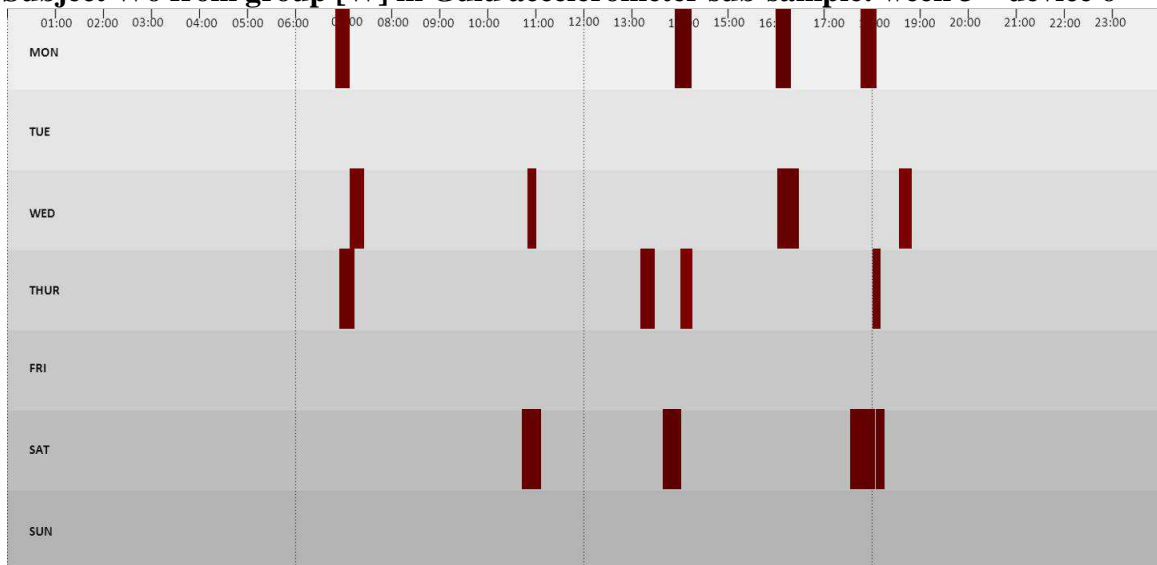
Subject W4 from group [W] in Gulu accelerometer sub-sample: week 3 – device 1



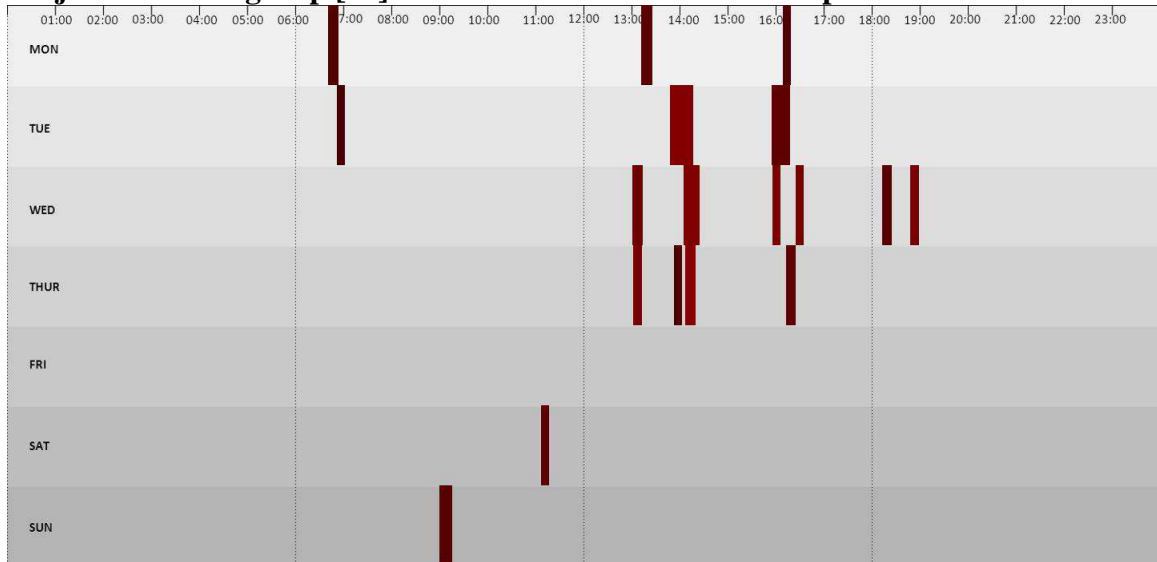
Subject W5 from group [W] in Gulu accelerometer sub-sample: week 3 – device 4



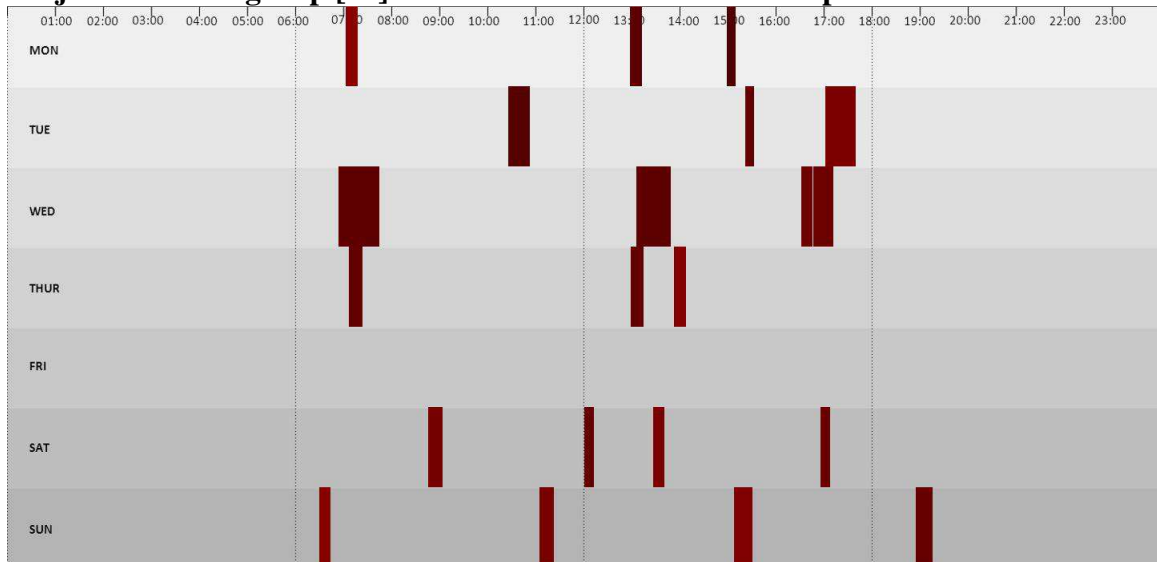
Subject W6 from group [W] in Gulu accelerometer sub-sample: week 3 – device 6



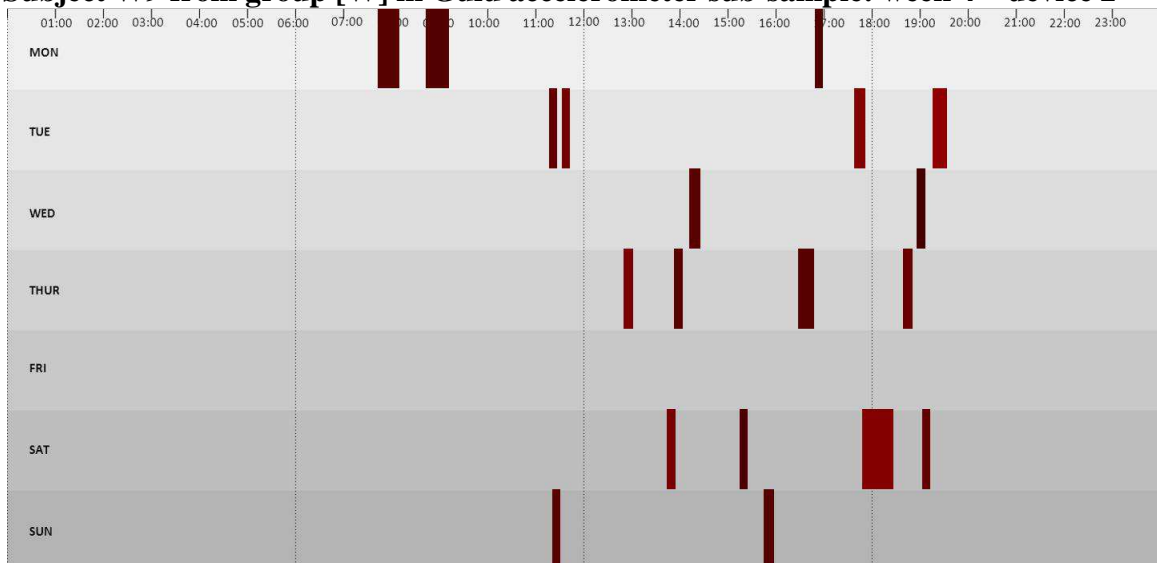
Subject W7 from group [W] in Gulu accelerometer sub-sample: week 3 – device 7



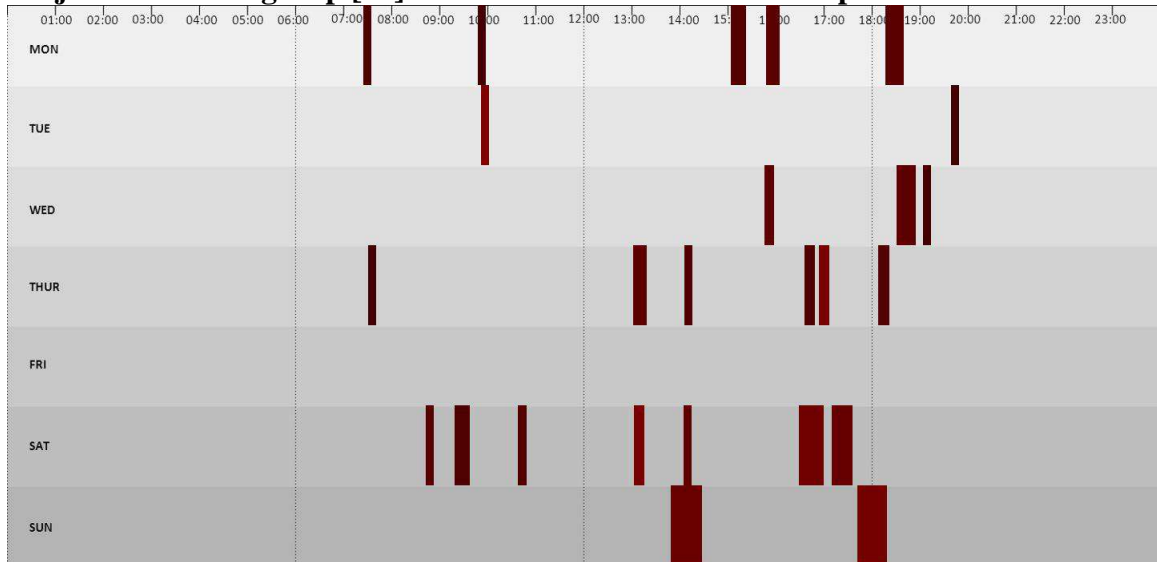
Subject W8 from group [W] in Gulu accelerometer sub-sample: week 4 – device 1



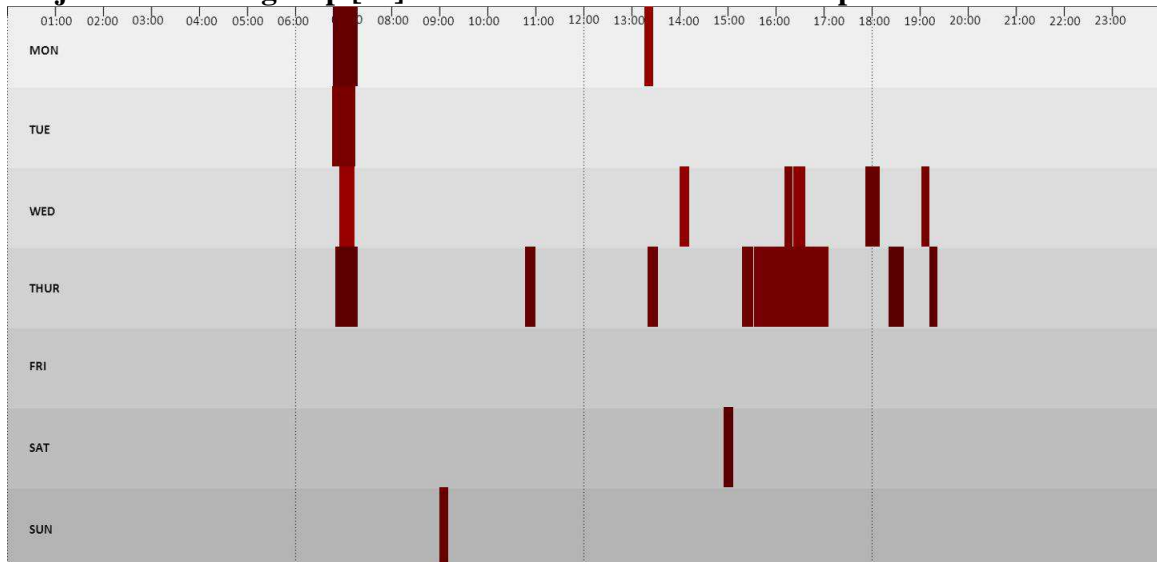
Subject W9 from group [W] in Gulu accelerometer sub-sample: week 4 – device 2



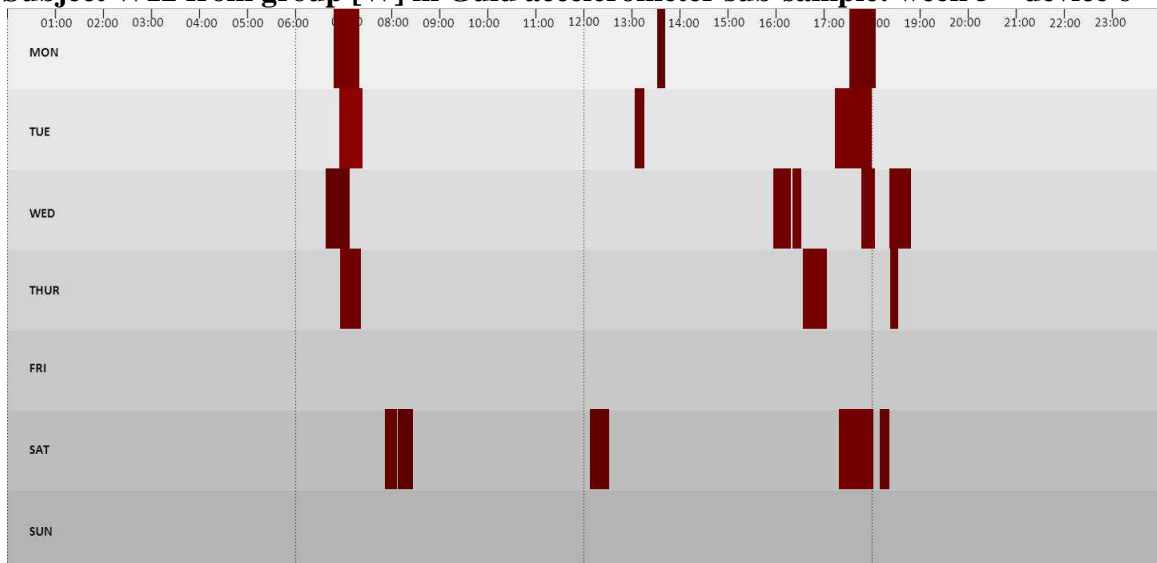
Subject W10 from group [W] in Gulu accelerometer sub-sample: week 4 – device 6



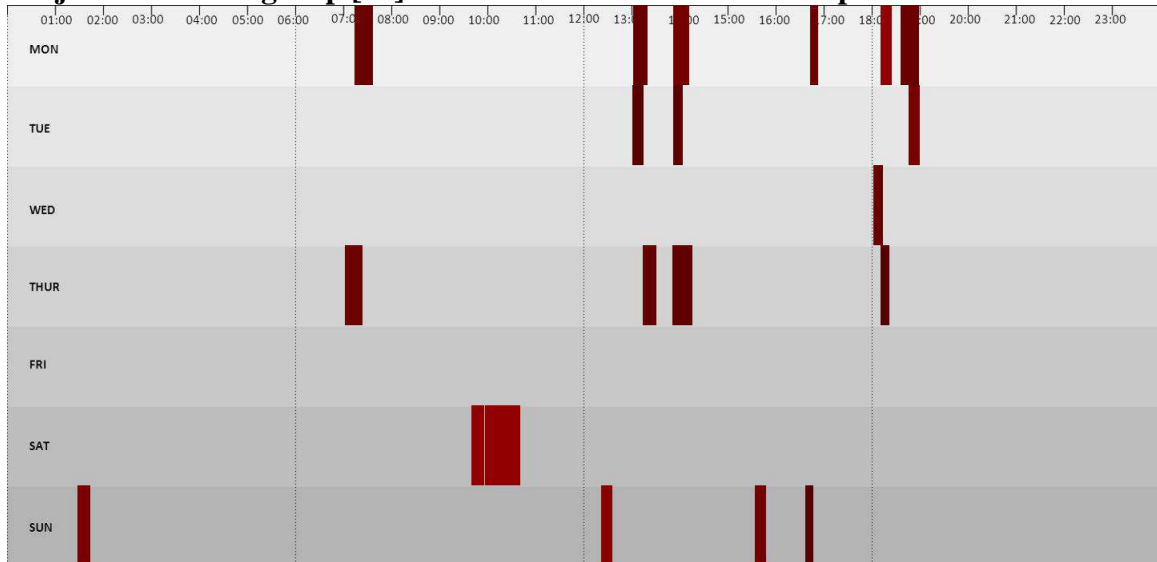
Subject W11 from group [W] in Gulu accelerometer sub-sample: week 5 – device 4



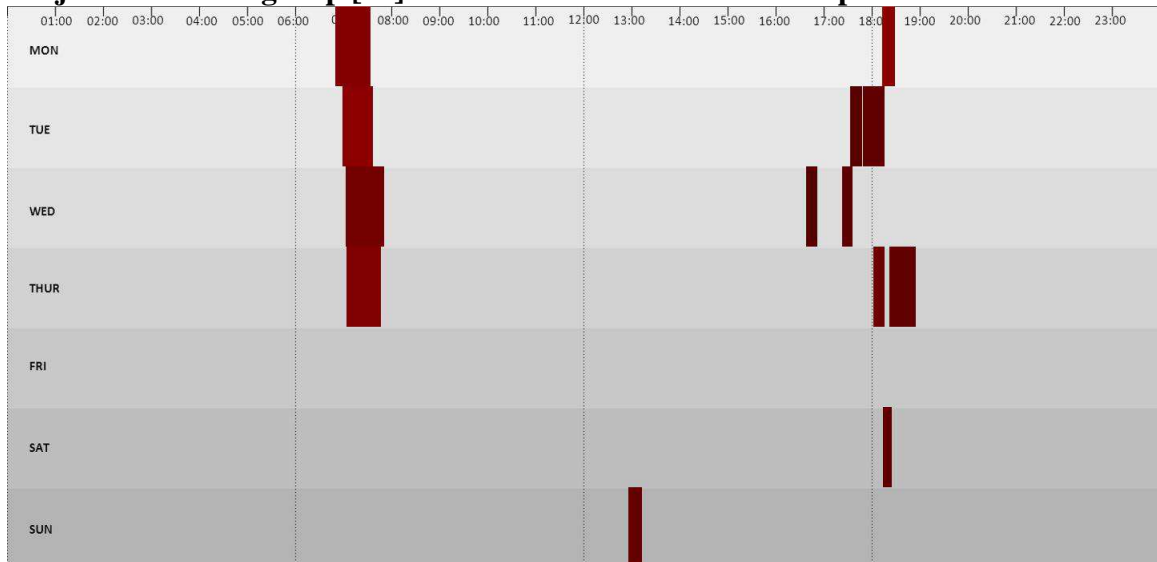
Subject W12 from group [W] in Gulu accelerometer sub-sample: week 5 – device 6



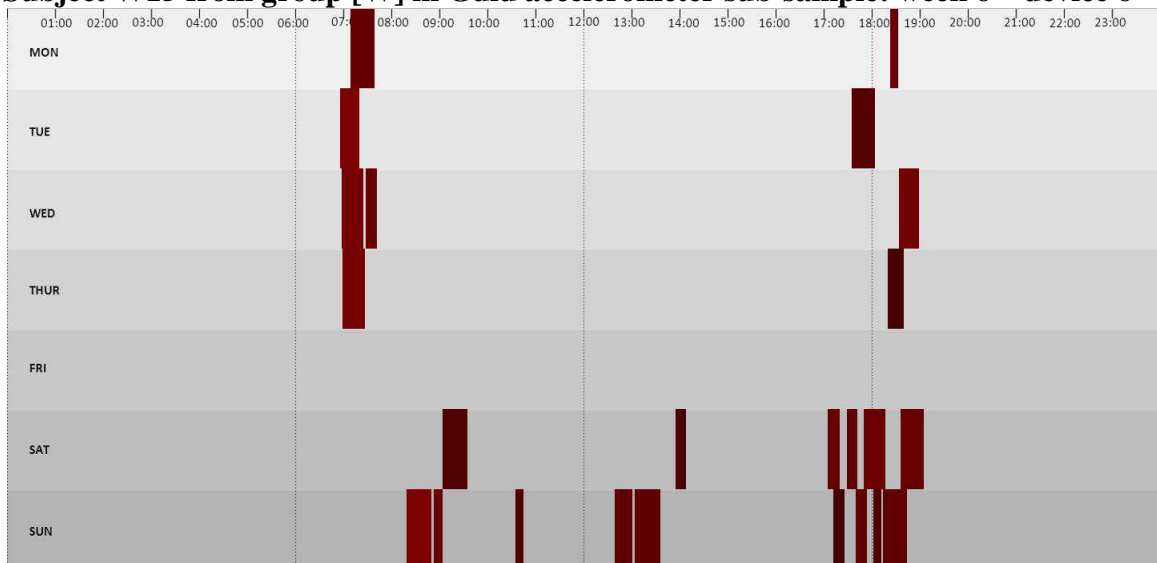
Subject W13 from group [W] in Gulu accelerometer sub-sample: week 5 – device 7



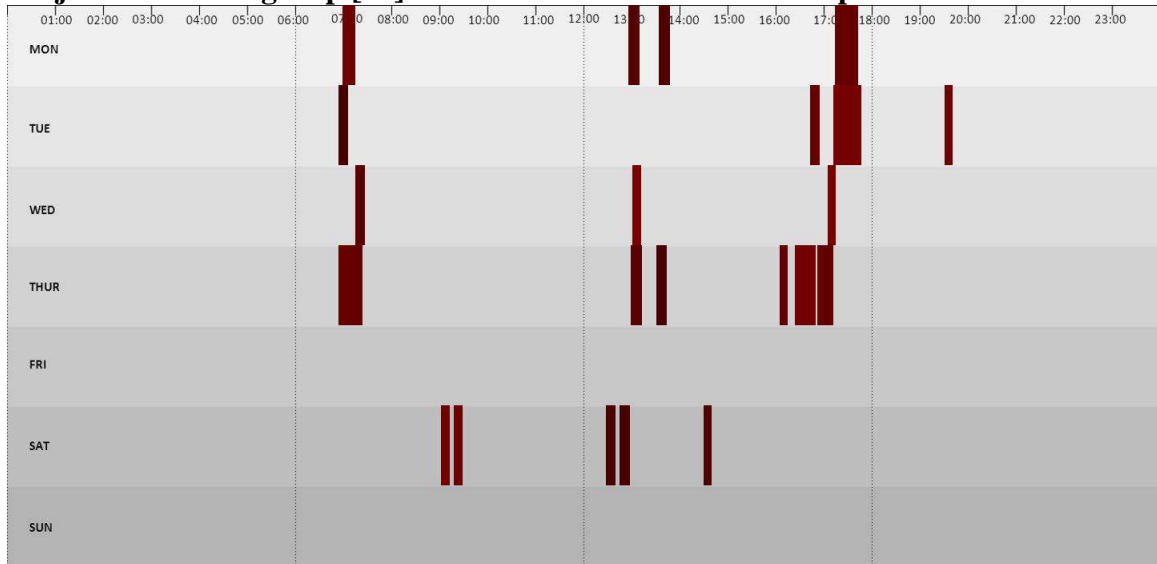
Subject W14 from group [W] in Gulu accelerometer sub-sample: week 6 – device 4



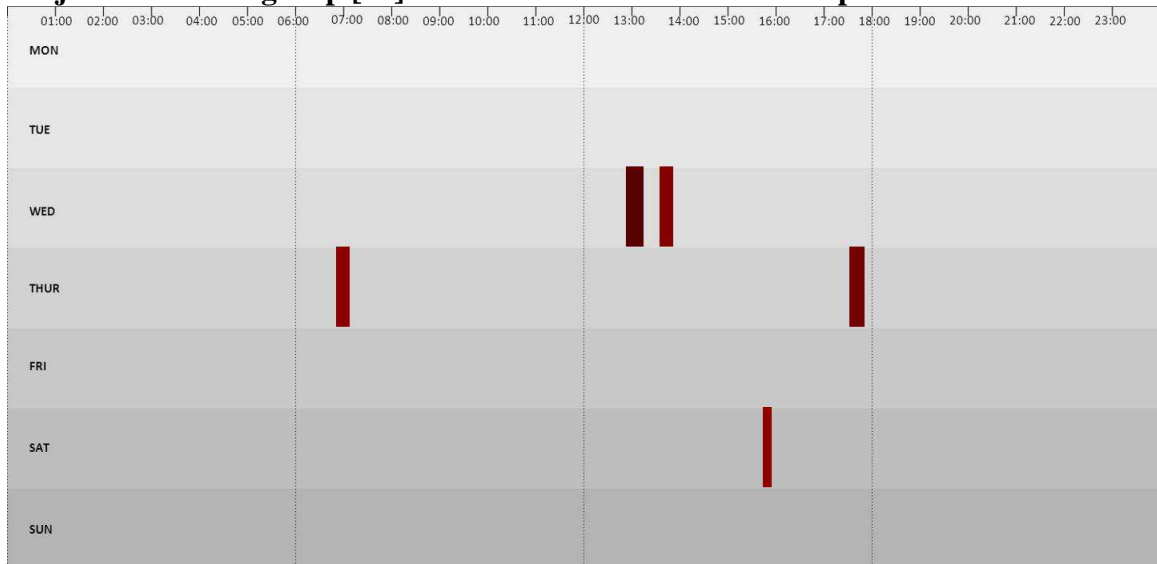
Subject W15 from group [W] in Gulu accelerometer sub-sample: week 6 – device 6



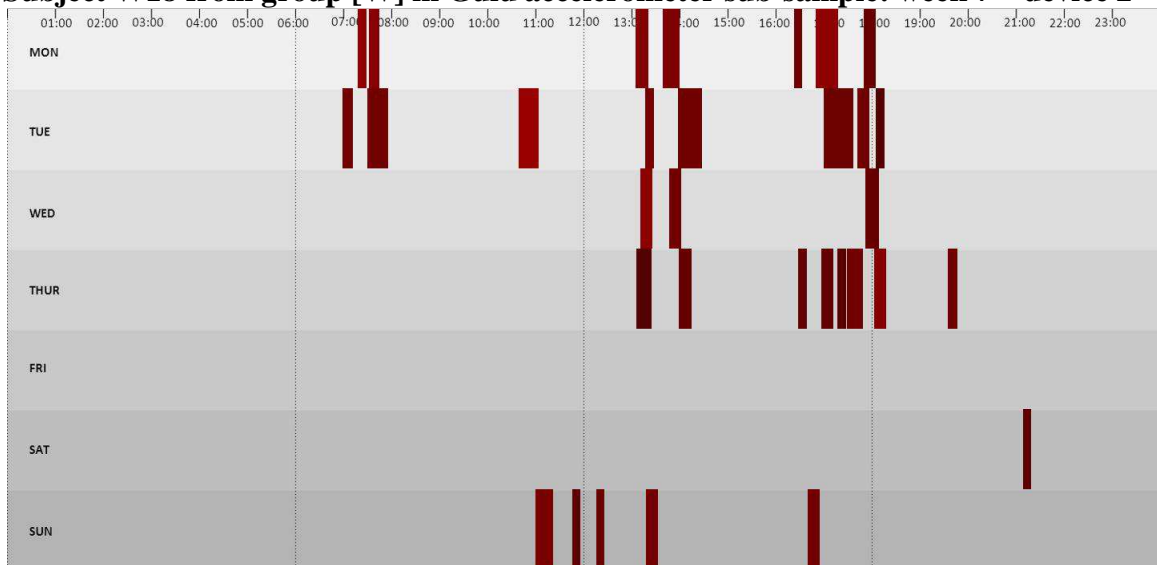
Subject W16 from group [W] in Gulu accelerometer sub-sample: week 6 – device 7



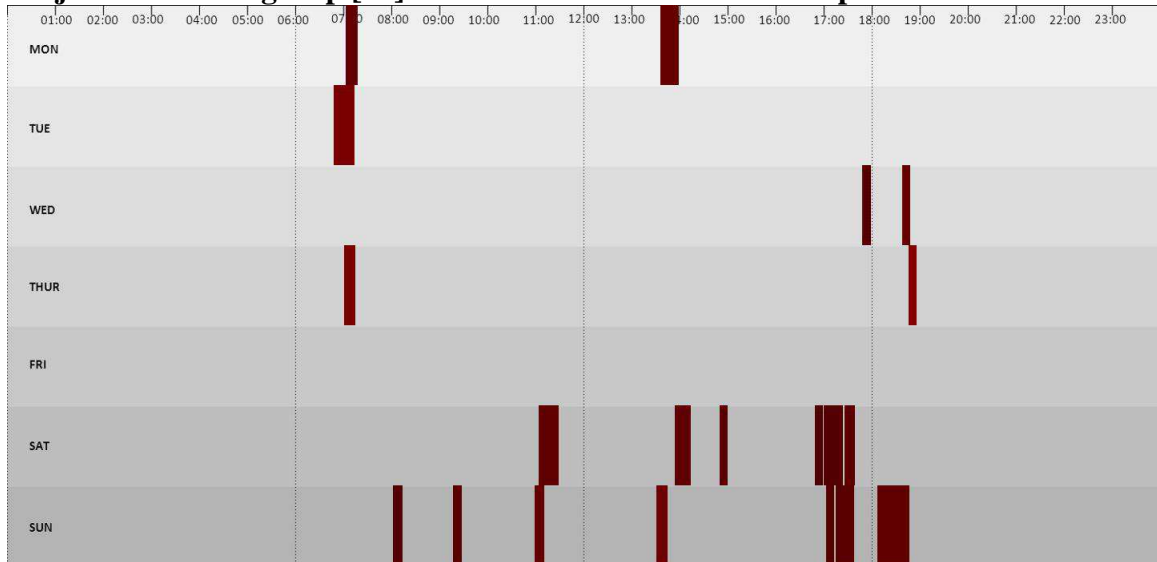
Subject W17 from group [W] in Gulu accelerometer sub-sample: week 7 – device 1



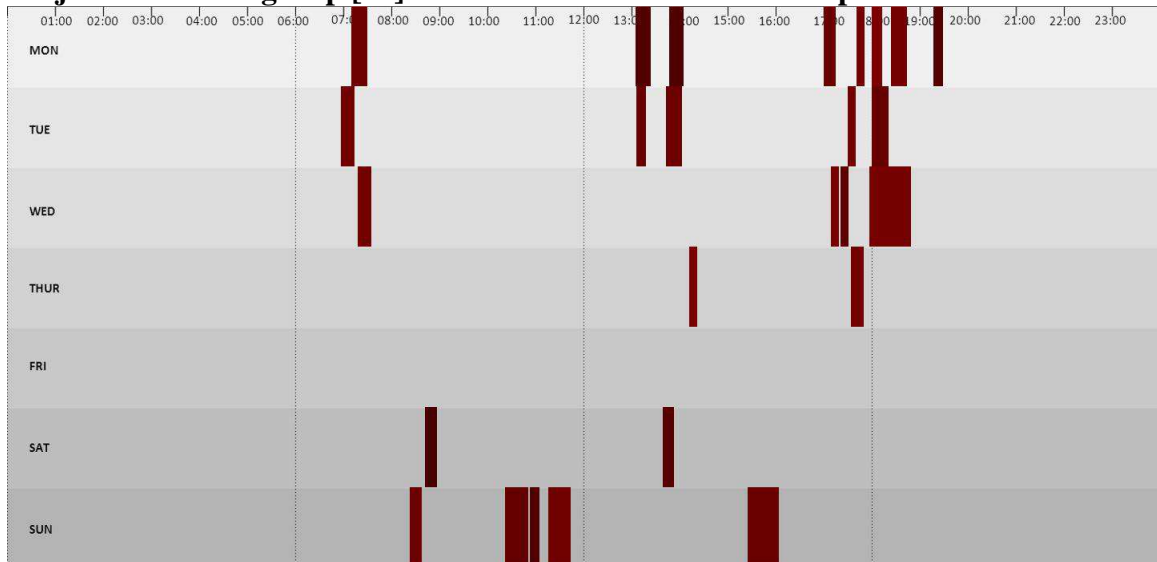
Subject W18 from group [W] in Gulu accelerometer sub-sample: week 7 – device 2



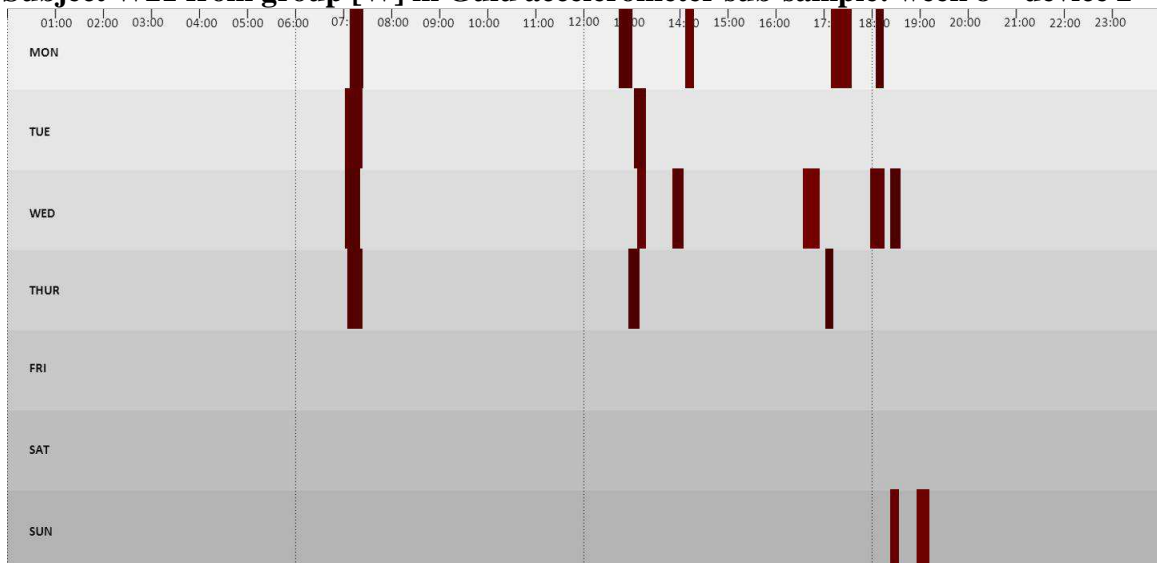
Subject W19 from group [W] in Gulu accelerometer sub-sample: week 7 – device 6



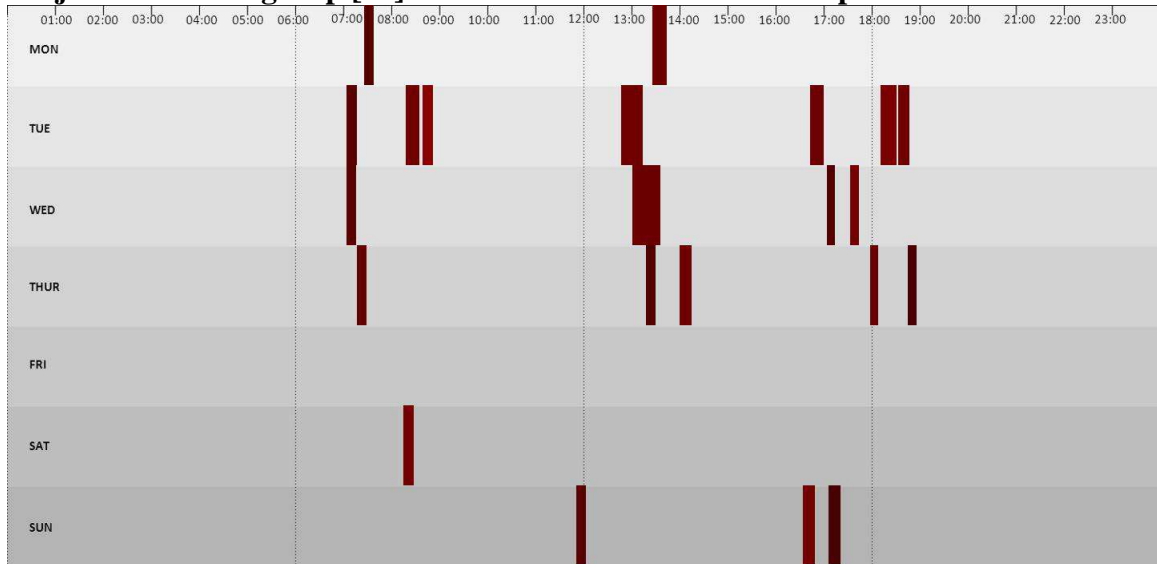
Subject W20 from group [W] in Gulu accelerometer sub-sample: week 8 – device 1



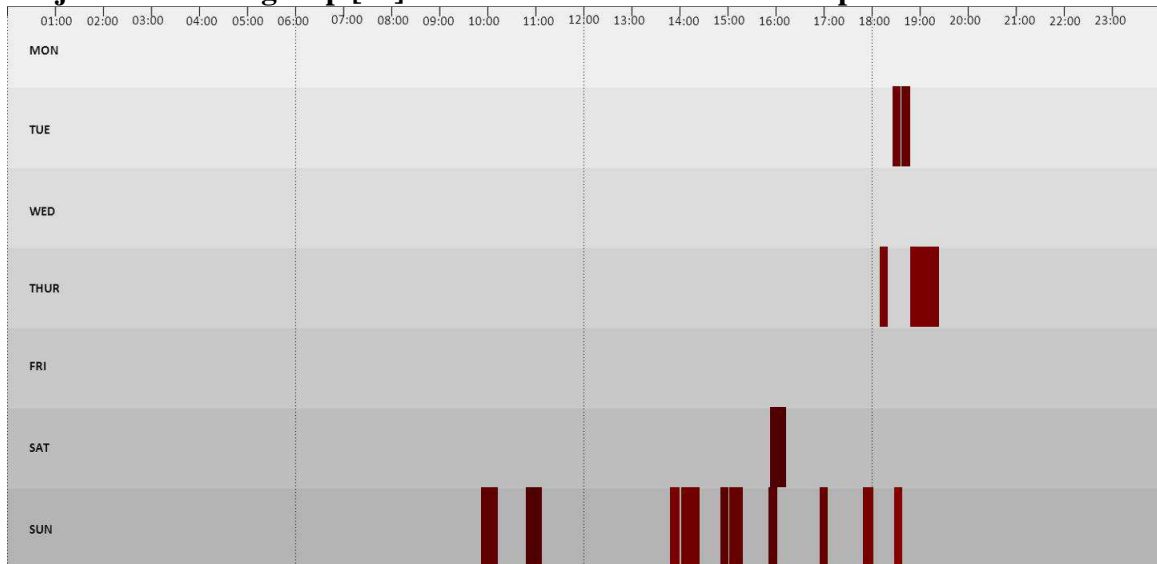
Subject W21 from group [W] in Gulu accelerometer sub-sample: week 8 – device 2



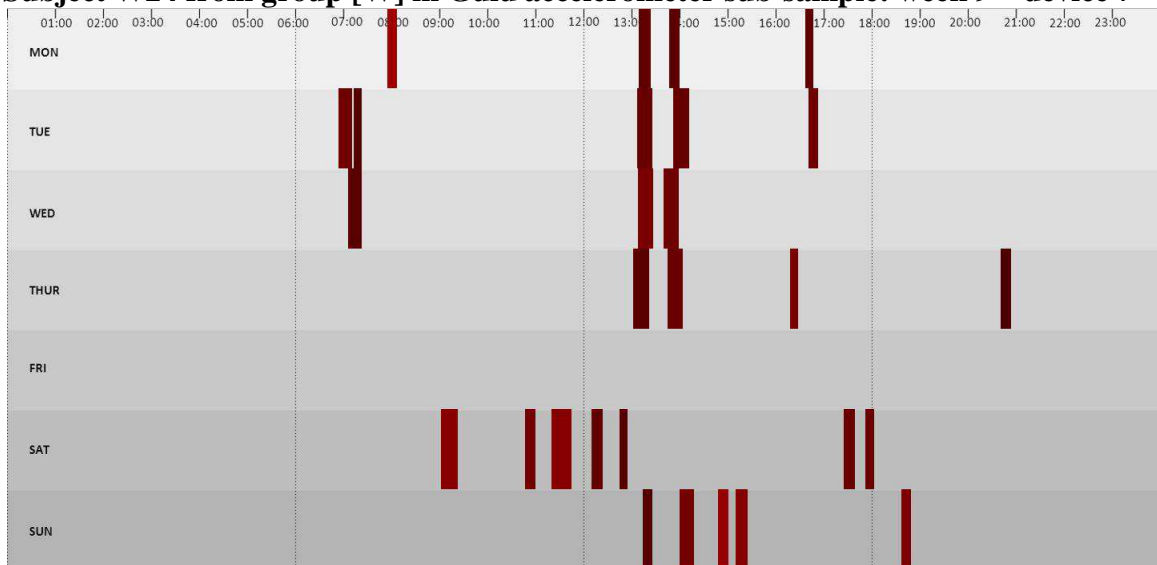
Subject W22 from group [W] in Gulu accelerometer sub-sample: week 8 – device 4



Subject W23 from group [W] in Gulu accelerometer sub-sample: week 9 – device 4

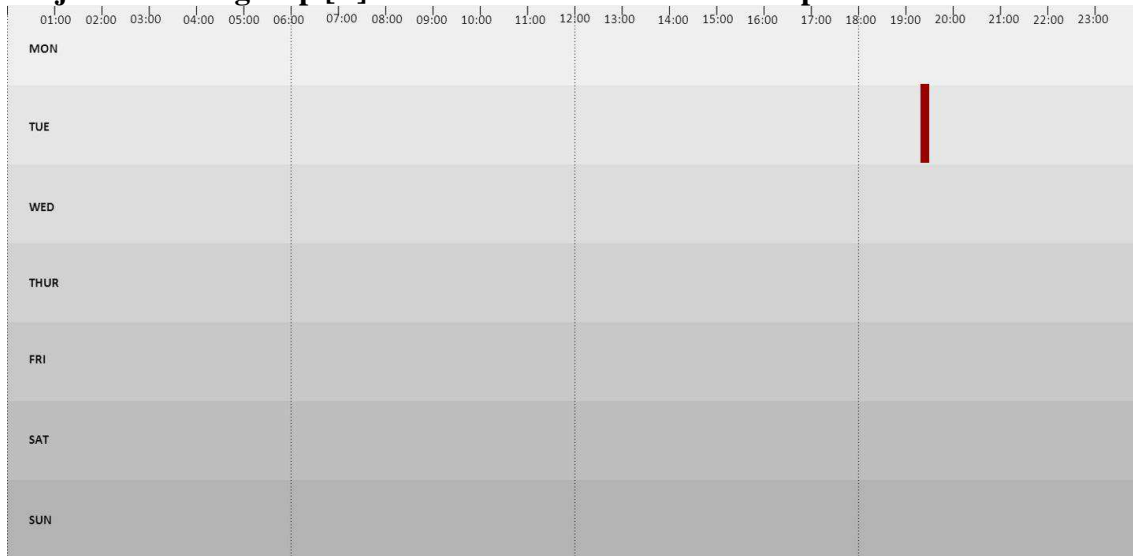


Subject W24 from group [W] in Gulu accelerometer sub-sample: week 9 – device 7



APPENDIX A23: “Heat Maps” for 10-Minute Bouts of VPA (Results)

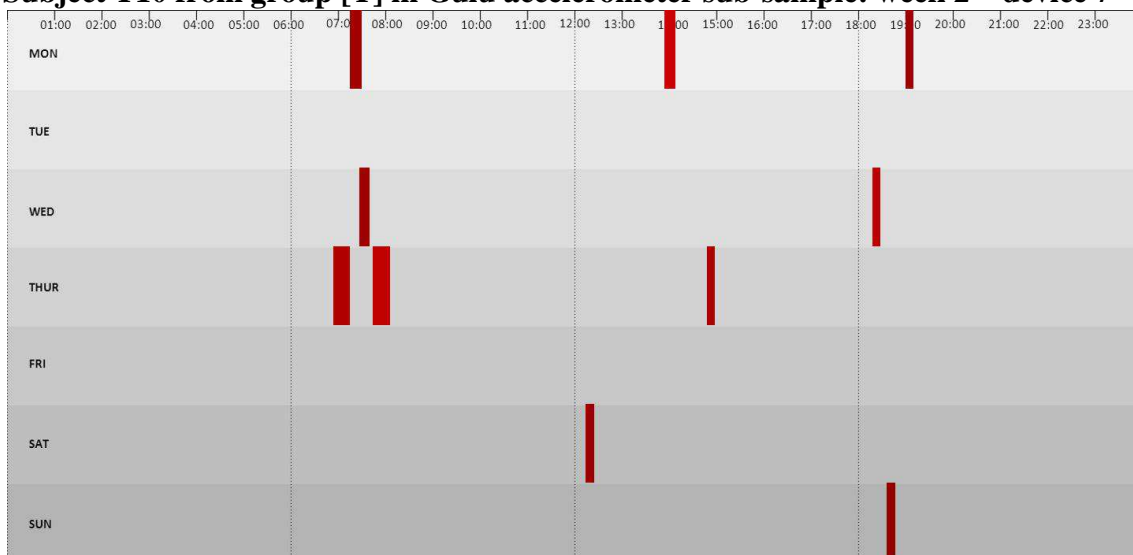
Subject T3 from group [T] in Gulu accelerometer sub-sample: week 1 – device 3



Subject T6 from group [T] in Gulu accelerometer sub-sample: week 2 – device 2



Subject T10 from group [T] in Gulu accelerometer sub-sample: week 2 – device 7



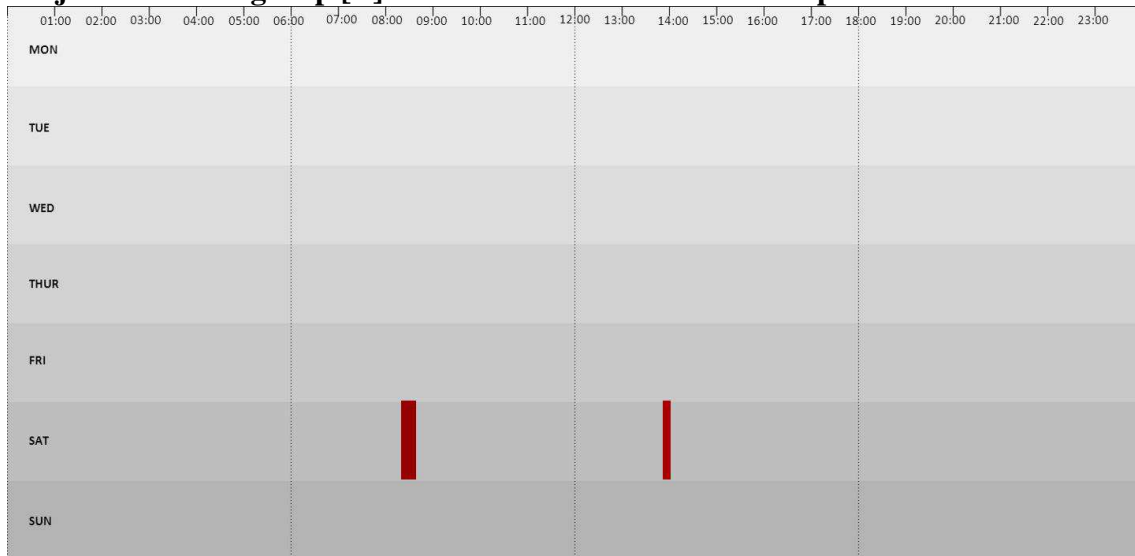
Subject T14 from group [T] in Gulu accelerometer sub-sample: week 4 – device 3



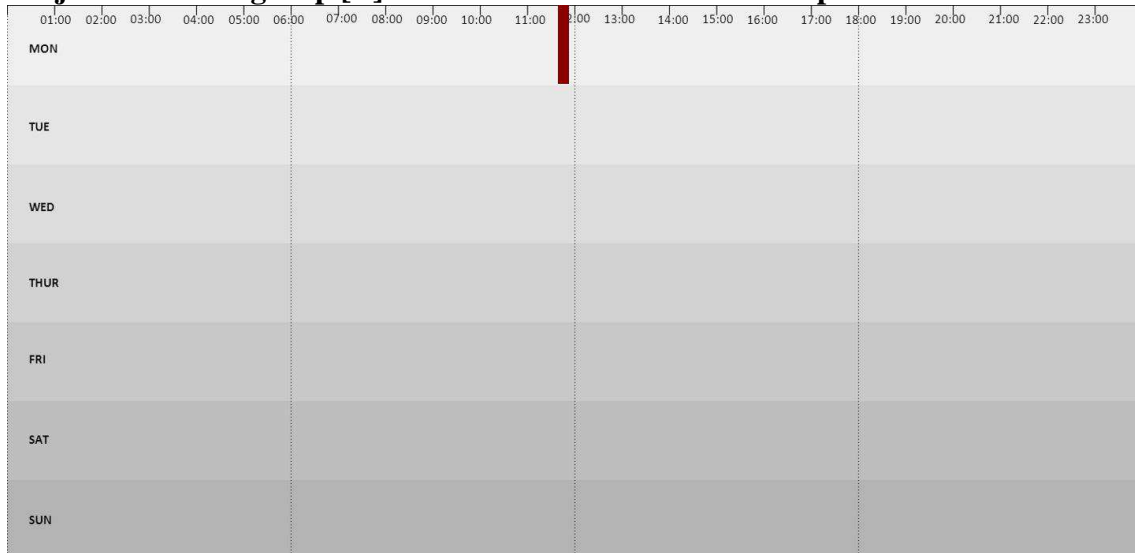
Subject T18 from group [T] in Gulu accelerometer sub-sample: week 5 – device 3



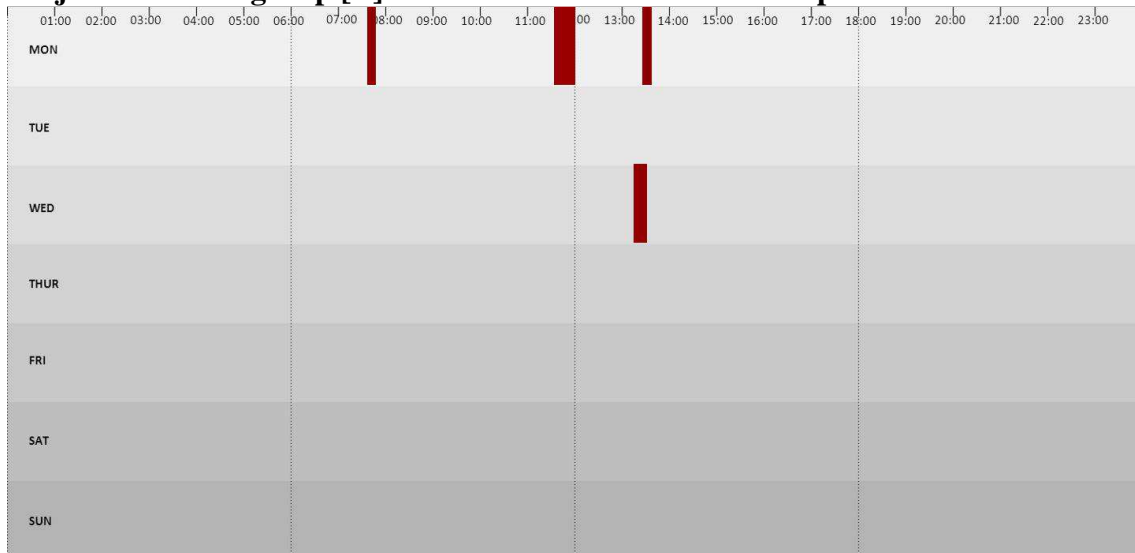
Subject T22 from group [T] in Gulu accelerometer sub-sample: week 7 – device 3



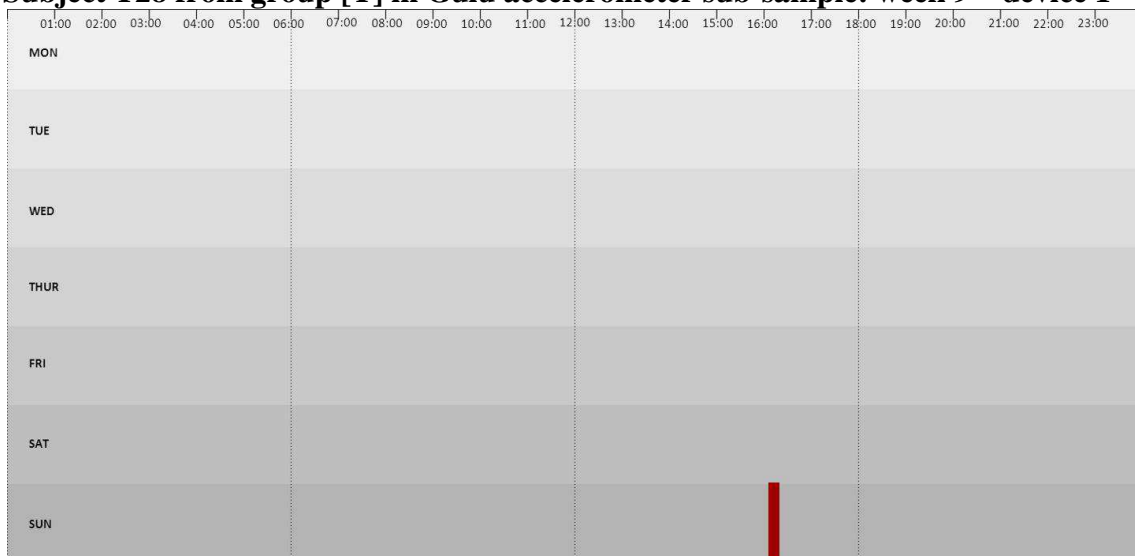
Subject T26 from group [T] in Gulu accelerometer sub-sample: week 8 – device 6



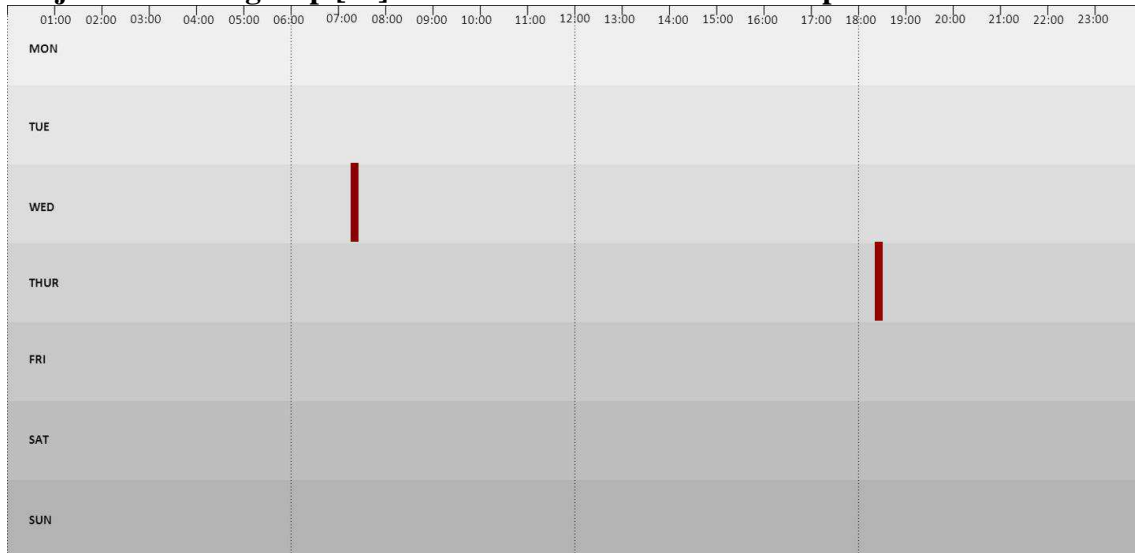
Subject T27 from group [T] in Gulu accelerometer sub-sample: week 8 – device 7



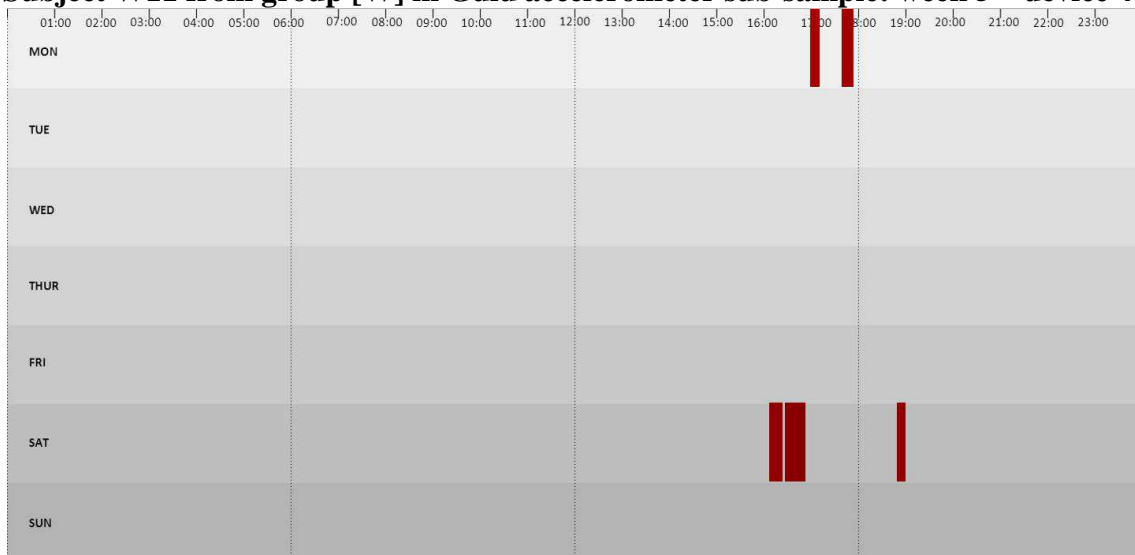
Subject T28 from group [T] in Gulu accelerometer sub-sample: week 9 – device 1



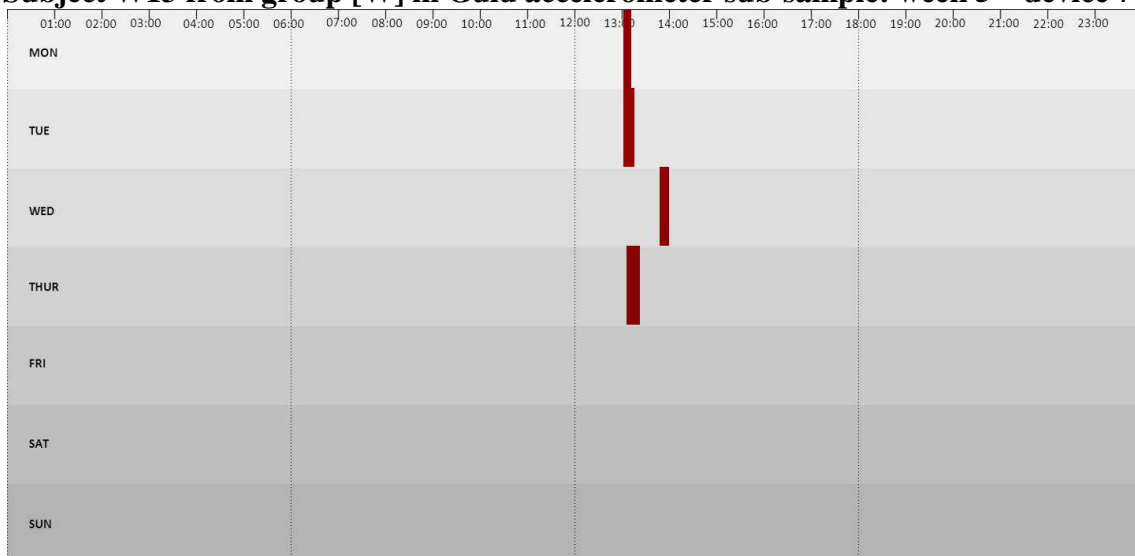
Subject W7 from group [W] in Gulu accelerometer sub-sample: week 3 – device 7



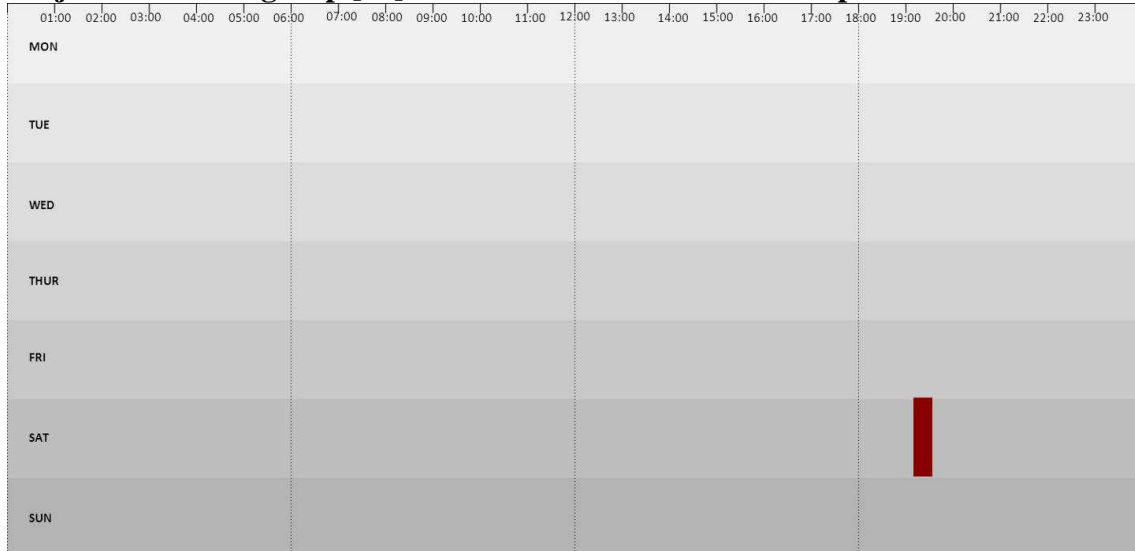
Subject W11 from group [W] in Gulu accelerometer sub-sample: week 5 – device 4



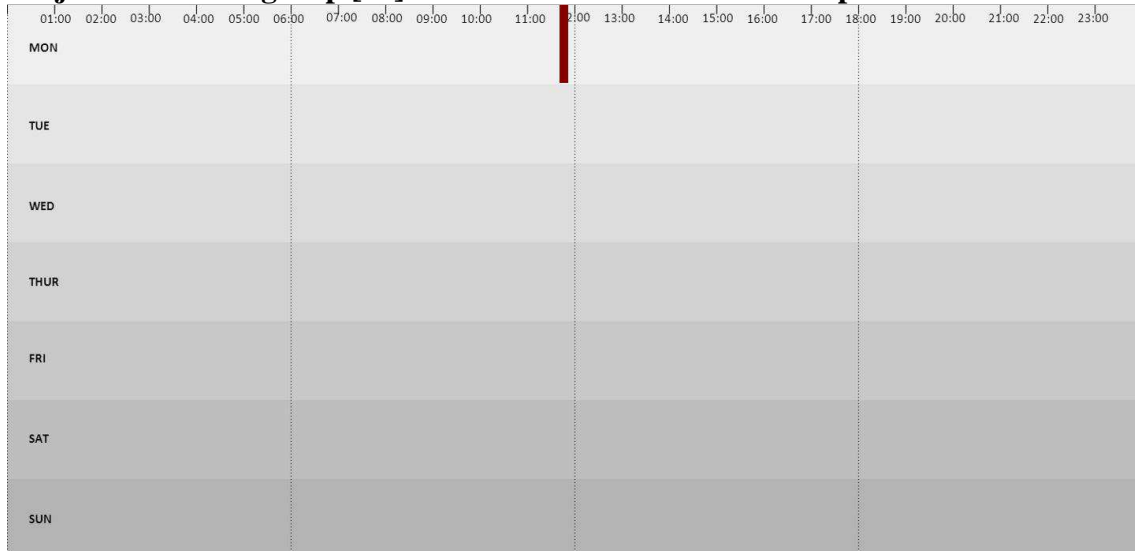
Subject W13 from group [W] in Gulu accelerometer sub-sample: week 5 – device 7



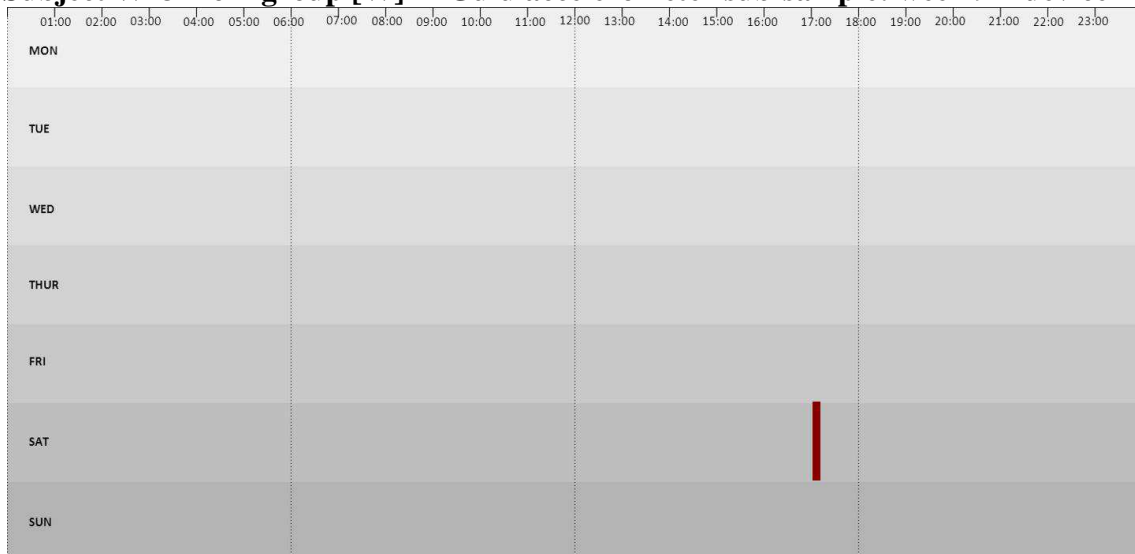
Subject W14 from group [W] in Gulu accelerometer sub-sample: week 6 – device 4



Subject W21 from group [W] in Gulu accelerometer sub-sample: week 8 – device 2



Subject W23 from group [W] in Gulu accelerometer sub-sample: week 9 – device 4



APPENDIX A24: Publication and Conference Plan

The papers below are listed in the same order as in the thesis for journals, books and conferences.

Publication type: Academic journal – original research

Authors: J.A.Richards, C. Foster

Title: Feasibility and reliability of using protocols adapted to the local context to measure the physical fitness and physical activity of young adolescents in Gulu, northern Uganda

Status: Planning stages (target journal: Journal of Sport Science)

Publication type: Academic journal – original research

Authors: J.A.Richards, J.D. McMullen, P.S. O’Callaghan, C. Foster

Title: Feasibility and reliability of using the Acholi Psychosocial Assessment Instrument in questionnaire format to assess the mental health of young adolescents in Gulu, northern Uganda

Status: Planning stages (target journal: Social Psychiatry and Psychiatric Epidemiology)

Publication type: Academic journal – original research

Authors: J.A.Richards, C. Foster

Title: Physical fitness of young adolescents in Gulu, northern Uganda – a cross-sectional comparison of the standing broad jump and multi-stage fitness test to international norms.

Status: Planning stages (target journal: Journal of Sport Science)

Publication type: Academic journal - review

Authors: M.E.G Armstrong, G. Tomkinson, J.A.Richards, M.I. Lambert, E.V. Lambert

Title: Who are the Afrofittest?

Status: Manuscript in preparation (target journal: Sports Medicine)

Publication type: Academic journal – original research

Authors: J.A.Richards, C. Foster

Title: Persistent mental health implications of war-time abduction for young adolescents in a post-conflict setting – a cross-sectional analysis in Gulu, northern Uganda

Status: Planning stages (target journal: Social Psychiatry and Psychiatric Epidemiology)

Publication type: Academic journal – original research

Authors: J.A.Richards, C. Foster

Title: Relationship between physical fitness and mental health in a post-conflict setting in a low-income country – a cross-sectional analysis in Gulu, northern Uganda

Status: Planning stages (target journal: Pediatrics)

Publication type: Academic journal – original research

Authors: J.A.Richards, C. Foster

Title: Who is reached by sport-for-development programmes delivered as voluntary and community-based football leagues – a case-study in Gulu, northern Uganda

Status: Planning stages (target journal: Journal of Sport for Development)

Publication type: Academic journal – original research

Authors: J.A.Richards, C. Foster

Title: Impact of a sport-for-development intervention on the physical fitness and mental health of young adolescents in a post-conflict setting – a cohort study in Gulu, northern Uganda

Status: Planning stages (target journal: Journal of Epidemiology and Health)

Publication type: Academic journal – original research

Authors: J.A.Richards, C. Foster

Title: Impact of a sport-for-development intervention on the physical fitness, physical activity and mental health of young adolescent boys in a post-conflict setting – a randomised control trial in Gulu, northern Uganda

Status: Planning stages (target journal: The Lancet)

Publication type: Academic journal - editorial

Authors: Z. Kaufmann, J.A.Richards, E. Wolff, N. Schulenkorf, K. Gannett

Title: New horizons for sport-for-development: from a movement to a field of study and practice

Status: Manuscript in preparation (target journal: Journal of Sport for Development)

Publication type: Book section

Authors: J.A.Richards, C. Foster

Title: Case-study from Gulu, northern Uganda in *Sport to the Rescue? A critical appraisal of sport-for-development*

Status: Manuscript in preparation

Publication type: Summary of thesis results for dissemination in Gulu and in the grey literature

Authors: J.A.Richards, C. Foster

Title: Physical fitness, physical activity and mental health of young adolescents in Gulu:
Impact of the Gum Marom Kids League in September - November, 2010

Status: Manuscript in preparation

Conference: International Congress on Physical Activity and Public Health – Sydney, 2012

Presenters: J.A.Richards, C. Foster

Title: Impact of a sport-for-development programme on the physical fitness and physical activity of young adolescents in a post-conflict setting – a case-study in Gulu, northern Uganda

Status: Abstract due March, 2012

Conference: International Congress on Physical Activity and Public Health – Sydney, 2012

Presenters: J.A.Richards, C. Foster

Title: Impact of a sport-for-development programme on the mental health of young adolescents in a post-conflict setting – a case-study in Gulu, northern Uganda

Status: Abstract due March, 2012

APPENDIX A25: Published Articles

Soc Psychiatry Psychiatr Epidemiol
DOI 10.1007/s00127-011-0454-9

ORIGINAL PAPER

Screening for traumatic exposure and psychological distress among war-affected adolescents in post-conflict northern Uganda

John D. McMullen · Paul S. O'Callaghan ·
Justin A. Richards · John G. Eakin ·
Harry Rafferty

Received: 20 April 2011 / Accepted: 4 November 2011
© Springer-Verlag 2011

Abstract

Background The war in northern Uganda has had a debilitating effect on the mental health of children and adolescents in the population. This study measures the prevalence and considers the aetiology of psychological distress in war-affected adolescents 4 years after the end of the conflict.

Methods This is a cross-sectional study of 205 adolescents, aged 12–19, from a boarding primary school in Gulu, northern Uganda. A war experiences checklist was developed with the assistance of local professionals. The Impact of Event Scale-Revised (IES-R) measured post-traumatic stress symptoms. Finally, the Acholi Psychosocial Assessment Instrument (APAI) was used to measure locally described mental health constructs similar to the Western concepts of depression and anxiety.

Results Four years after the end of the war, 57% of the students were still found to have clinically significant levels of post-traumatic stress symptoms using a similar cut-off score to previous studies among the same population. Both components of traumatic exposure: (i) the number of types of traumatic event experienced; and (ii) whether the adolescent was abducted were significantly associated with psychological distress. There was a strong correlation between post-traumatic stress symptoms and internalising symptoms.

Conclusion War-affected adolescents may continue to suffer from significant psychological stress in the years following the cessation of conflict. Multiple exposure to a number of different types of traumatic event may directly increase the likelihood of psychological distress especially for those exposed to the most extreme violence. The feasibility of employing a locally developed and validated screening instrument is demonstrated. Implications for future research and intervention in post-conflict areas are considered.

Keywords Trauma · Depression · Adolescent · War · Exposure

Introduction

Globally, just over 1 billion children under the age of 18 live in countries affected by war. In places such as northern Uganda, the consequences of war for children extend beyond killing and injury and include the denial of their material and emotional needs and the complete fragmentation of their society. There is a growing need to increase public awareness of the effects of conflict on children; to understand the impact on their mental health and development; and to develop interventions to treat children traumatised by war [1].

Post-traumatic stress disorder (PTSD) is the most common psychological disorder found in war-affected populations [2]. There is some debate over whether or not the concept of PTSD, with its distinct number of symptoms, can be applied to children and adolescents. Dyregrov and Yule [3] suggest that above the age of 8 years, children's reactions to trauma are almost the same as those manifested by adults. Both adults and children with

J. D. McMullen (✉) · P. S. O'Callaghan ·
J. G. Eakin · H. Rafferty
School of Psychology, Queens University Belfast,
David Kier Building, Belfast BT9 5BP, UK
e-mail: jmcullen01@qub.ac.uk

J. A. Richards
Department of Public Health, University of Oxford,
Rosemary Rue Building, Old Road Campus, Headington,
Oxford OX3 7LF, UK

PTSD suffer greatly due to involuntary sensory, visual, or other recall of the most horrific moments of their lives. Their memory is fragmented and their ability to focus their mind is reduced. A great deal of energy is spent in avoidance, because remembering the traumatic experience brings with it painful emotions, panic, a sense of threat and distress [4].

International efforts have failed to fully protect children from being recruited and abducted by armed groups. The coalition to stop the use of child soldiers (CSUCS) estimate that 300,000 children, sometimes as young as 6 years old, remain in the ranks of armed groups in at least 24 different countries or territories [5]. The CSUCS highlights the fact that girls continue to be involved in military forces in combat and non-combat roles, often being subjected to rape and other forms of sexual violence. As well as being used as combatants and for sexual services, child soldiers are used as messengers, scouts, porters and cooks [6]. Child soldiers are seen as more expendable than adults. Child soldiers are often beaten and continuously exposed to atrocities and violence, so that they do not question the authority of the adults in the group. Sometimes they are forced to kill their own family members or friends to discourage them from trying to escape and return to their communities [7]. Despite this reality, the impact of armed conflict on child and adolescent mental health has often not been considered or even recognized [8].

Northern Uganda

The war in northern Uganda was one of the most persistently brutal the world has seen [9]. During 24 years of conflict between the government of Uganda and the Lord's resistance army (LRA), 1.8 million people were forced to live in squalid camps for internally displaced people (IDP) where the mortality rate reached nearly 1,000 deaths per week. Forty per cent of these deaths were children under 5 [10]. The LRA has committed countless human rights abuses against Ugandan civilians from the Acholi ethnic group, including the abduction of up to 60,000 children to act as soldiers [11].

During the war, researchers in northern Uganda found extremely high levels of trauma among children who had been abducted by the LRA. One study of post-traumatic stress symptoms (PTSS) found that 97% of the former child soldiers had PTS symptoms [12]. Amone-P'Olak [13] measured the psychological impact of war and sexual abuse on adolescent girls in northern Uganda and found 98% of the girls had clinically significant PTS symptoms with 78% in the moderate range and 21% in the severe range. In another study in Gulu, Amone-P'Olak et al. [9] considered the impact of war experiences and physical

abuse on formerly abducted boys (child soldiers). They found PTS symptom rates of 97% with 90% in the moderate range and 7% in the severe range. All of these studies used the Impact of Event Scale-Revised (IES-R) as the screening measure for PTSS.

Exposure effects

A 'dose-effect relationship' between the *number* of traumatic events witnessed and symptoms of emotional distress has been found in children and adolescents exposed to war [17, 18]. This may lead to a 'building-block effect' where experiencing an increasing number of traumatic events directly increases the likelihood of mental-health disorders in a linear fashion [4, 14]. In contrast, other studies with war-affected children have found no significant relationship between total number of traumatic experiences and posttraumatic stress reactions [12, 15]. Repeated traumatic life-threat during war or natural disaster may cause protective factors, such as social support or personal resilience, to be weakened. The *type* of traumatic event has also been found to be important. More severe types of trauma, such as the violent death of a family member or witnessing someone being injured, tortured or killed, are related to higher levels of distress [17]. The individual's perception of a direct threat and their degree of involvement during the traumatic event are also associated with increased psychological distress [19].

Other effects of trauma

In addition to PTSD, other non-trauma-specific psychopathologies have been identified in war-affected individuals [1]. The traumatic effects of war are 'multiple, chronic, interrelated, recurrent, and therefore difficult to disentangle' [20]. Depression is strongly associated with PTSD in war-affected children. Co-morbid depression has been found in up to 80% of the children who were exposed to severe trauma [21]. This pattern was also found among refugee children in Gaza [20] and among war-affected adults in the Democratic Republic of Congo [22]. The nature of the relationship between these two disorders is unclear. It is unknown as yet whether the disorders emerge as distinct responses to trauma, or whether depression is a response to PTSD through mediating life events [20]. A key distinction, where PTSD is closely linked to earlier war trauma but depression is associated with recent life difficulties, has been found [19, 23, 24]. Although the aetiology of PTSD and depression may differ, this does not mean that PTSD and depression are by necessity qualitatively distinct disorders. In fact, PTSD and depression may share similar neural pathways since PTSD interventions have also been associated with reductions in depression [25]. This

suggests that similar cognitions and behaviours may be linked to both conditions.

There are many other consequences of war-related trauma on children. In fact, the entire spectrum of contextual, interpersonal, and intrapersonal realms is affected [16]. Other specific issues identified in the literature are abnormal grief reactions [19], academic impairment [26], fear and dependent behaviour [27], physical symptoms e.g. loss of appetite and multiple physical complaints [28], suicidal ideation [29] and moral development/loss of childhood [7].

Assessment of trauma

Comprehensive epidemiological, population-based screening for traumatic symptoms and psychosocial impairment is essential when planning therapeutic interventions. This is a vital stage in separating those in urgent need of help from those who do not require immediate intervention. Since PTSD rarely occurs in isolation, a cross-cultural psychological assessment battery for children and adolescents should ideally consist of a broad screening of symptoms associated with multiple disorders [18]. Laor and Woolmer [27] stress the importance of assessing risk factors and gathering information concerning the details of the traumatic events. They advocate the integration of this assessment into the everyday activities of schools. Holding assessment and intervention in school can also help to reduce stigma and reach out to individuals who would not otherwise access mental health services [16, 30].

The present study

This school-based study was undertaken primarily to measure the prevalence of psychological distress in war-affected young people 4 years after the war had ended in northern Uganda. We anticipated that a comparison could then be made with the levels of post-traumatic stress found during the war, in order to provide an indication of the continuing presence or remission of these symptoms. Internalising mental health difficulties were assessed using a locally developed, culturally relevant instrument. Association between local understandings of depression/anxiety and post-traumatic stress could then be considered. The association between exposure to traumatic events during the war and psychological distress was also explored. Specifically, we examined whether abduction and witnessing a cumulative number of different types of traumatic events were associated with increased levels of symptomatology. Our results were used to identify those adolescents most in need of therapeutic intervention and to help provide access to this.

Methods

Participants

A total of 205 adolescents from a boarding primary school in Gulu, northern Uganda participated in the study. They were aged 12–19 (mean = 14.4, SD = 1.7), much older than would be expected in a 'western' primary school. The school was established by the Ugandan and Belgian governments to educate former child soldiers and other young people who were directly affected by the war. It was agreed with the school management that all of the adolescents in primary 5, 6, 7 and vocational classes who agreed to participate would be screened. As the study took place in a boarding school for children from all over Uganda, it was not possible to gain parental consent. Therefore, consent was gained from the school staff. Written consent was sought from all adolescents involved in the study. When this was not possible informed oral consent was gained through local interpreters. From the time of first contact, the young people were made aware of their right to withdraw at any time and have their data destroyed. None chose to do so.

An initial sociodemographic questionnaire revealed that the sample comprised 112 boys (55%) and 93 girls (45%). 97 (47%) of the adolescents screened had been abducted by the LRA. For those who were abducted, the length of time spent with the rebels ('in the bush') ranged from 1 week to 10 years with an average time of 21 months. All of the other adolescents screened had been affected by the war either through the loss of immediate family or by direct experience of violence. The adolescents spoke Acholi Luo and had English as a second language.

Procedure

The screening took place during the month of June 2010. In total fifteen local professionals, including a lecturer at Gulu University, a local psychiatrist, a psychologist and various other mental health workers, assisted with the development of the study. These helped to increase the cultural applicability of the screening instruments and to increase face validity by making adjustments to the phrasing of some items. The initial translation of the WEC and IES-R from English to Luo was completed by a local psychotherapist and social worker. The translation was checked and amended by a community director/pastor and other local Acholi partners to inform an appropriate translation that reflected the concepts under study. Back-translation of all three measures was then completed by a local community counsellor. The final translations were checked by the two teachers at the school who lead the counselling/pastoral care team. Prior to assessment, a brief psychoeducational

seminar was held with all of the students and staff to explain the purpose of the study, what it would entail, and how it might affect them.

A pilot study with a representative sample of 20 adolescents took place before commencement of the screening. Adolescents representing each class and a range of ages (from 12 to 19) were randomly selected for this. This enabled practical issues to be resolved such as completion time, sufficient provision of resources and best method of explanation and delivery. Discussion with this group also confirmed that the adolescents understood the questions and format and did not experience negative emotions during completion of the questionnaire. The pilot study scores also allowed test–retest reliability to be measured using Pearson correlation.

The questionnaires were delivered item by item by the school teachers/counselors, assisted by the researchers, in class groups of 40–50. Each questionnaire item was written in Luo and English. The data gathered on the number and types of war-related incidents the adolescents had experienced and the extent of trauma and internalising symptoms among adolescents in the school were shared with the staff in the school. The most traumatised adolescents, according to their score on the IES-R and APAI, received between four and six sessions of Narrative Exposure Therapy [33]. In addition, a list of adolescents with the highest scores (1 SD above the mean) on the IES-R and APAI was provided for the school counselling team. Finally, a further psychoeducational session, employing aspects of the Teaching Recovery Manual [34] was carried out in groups after the screening. This aimed to educate the young people about the effects of trauma; normalise the symptoms that some of them had been experiencing; and offer strategies for dealing with these symptoms.

Measures

War Experiences Checklist (WEC)

A war experiences checklist was developed by the current researchers with assistance and advice from local professionals. The checklists employed by Amone-P'Olak [13] and Mels et al. [24] were modified by the team to comprehensively determine the experiences of the current sample in as few items as possible. This assistance by professionals working with the adolescents enabled cross-cultural issues of the scale to be addressed. 15 events were included (see Table 1) in this self-report checklist and children were required to indicate their response to the instruction. 'Please tick "Yes" or "No" if you experience or witnessed these events during the war'. It was made explicit, verbally and in writing, that we were simply assessing the war events that adolescents had *witnessed*.

They did not have to differentiate between atrocities that had been directly inflicted on them, those that they had been forced to inflict or those that they had simply seen. It was hoped that this would result in greater honesty in the adolescent's responses and limit any distress in completing the checklist. Test–retest reliability was demonstrated in the pilot study to be strong ($r = .818, p < .001$).

Impact Of Event Scale-Revised (IES-R)

Posttraumatic stress symptoms were measured by the Impact of Event Scale-Revised (IES-R) [31]. This scale is based on the DSM-IV criteria for PTSD and consists of three subscales: intrusion (seven items), avoidance (eight) and hyperarousal (seven). Respondents were asked to rate difficulties experienced in the past 7 days on a five-point Likert Scale scale of 0 (not at all), 1 (a little bit), 2 (moderately), 3 (quite a bit), or 4 (extremely). Possible answers were accompanied by a visual probe (a dot growing to ever-larger circles from 0 to 4). The IES-R yields a total score and subscale scores for intrusion, hyperarousal, and avoidance. The IES-R was developed with Western samples but has frequently been used in non-Western contexts. Of relevance to the current study, it has been used to screen for PTSD among children in northern Uganda [9, 12, 13]. Mels et al. [18] adapted the instrument to provide a community-based, cross-cultural instrument to screen for posttraumatic stress symptoms in eastern Democratic Republic of Congo. This adapted version had internal reliabilities for the subscales ranging from .79 to .84 and Cronbach α for the total scale at .92. This version was chosen for the current study due to the similarities in the traumatic experiences of children in the DRC and northern Uganda. Based on previous studies [12, 18], and for purposes of comparison, a score of 22 was selected as a measure of a clinically significant level of PTSS. The current pilot study showed test–test reliability to be strong ($r = .757, p < .001$). Instrument reliability was assessed using Cronbach α coefficients and internal consistency was found to be high ($\alpha = .92$).

The IES-R is not used to assess PTSD directly, but rather to screen for general reactions to traumatic distress. There is growing evidence questioning the validity of the DSM-IV-based construct of PTSD, and it is suggested that it may be more meaningful to address 'traumatic distress' as a broad range of symptoms, rather than a clearly defined syndrome [24]. The current study therefore refers to the PTSS displayed by the adolescents.

Acholi Psychosocial Assessment Instrument (APAI)

To measure internalising disorders, an instrument which was developed and validated in Gulu among the Acholi

Table 1 War experiences checklist (WEC)—the number and percentage of adolescents who witnessed each traumatic event type ($n = 205$)

Alegi ni igwet gin ma itamo ni kakare onyo kuu ka itye ki ngec onyo ineno gin ma-otime ikare me lweny. Please tick 'Yes' or 'No' if you experienced or witnessed these events during the war	<i>n</i>	%	% Boys	% Girls
Ineno kom dano mutoo <i>Dead bodies or body parts</i>	156	76	80	72
Dano acel me dog gangwa kineko <i>Friend or family member killed</i>	148	72	71	74
Mwoc pa gurunet onyo bom <i>Explosions</i>	147	72	73	70
Kec amoro wang dano <i>Severe hunger</i>	136	66	70	62
Goyo dano <i>Flogging or Beating</i>	129	63	65	60
Wango ot <i>Burning homes/houses</i>	122	60	56	63
Dangee wek ki mini kony <i>Shouting or screaming for help</i>	114	57	60	51
Yako onyo kwo <i>Looting or Stealing</i>	113	55	55	55
Celo muduku <i>Gunfire</i>	95	46	50	42
Ki ngolo wiwa iyoo <i>Ambush</i>	95	46	49	43
Neko dano <i>Killing</i>	93	45	55	34
Mako dano <i>Abduction</i>	90	44	47	40
Bwuru dano me aneka <i>Threats to kill</i>	85	41	51	30
Ngolo it, um, cak onyo dog dano <i>Mutilation</i>	53	26	24	28
Mako butu ki dano tektek <i>Rape</i>	36	18	16	19

Luo people was employed. The Acholi Psychosocial Assessment Instrument was developed by Betancourt et al. [32] using data from a qualitative study with young people and adults living in IDP camps. They were asked to identify and describe the important problems of youth. Five common local mental health syndromes affecting Acholi youths were determined: *two tam*, *par*, *kumu*, *ma lwor*, and *kwo maraco*. The symptoms of these local syndromes share similarities with Western definitions for mood disorders or depression (*two tam*, *par* and *kumu*), general anxiety disorder (*ma lwor*) and conduct disorder (*kwo maraco*).

Similar to the IES-R, respondents were asked about the frequency of each symptom during the previous week, with responses coded on a 4-point Likert scale ranging from '0' (never) to '3' (constantly). The APAI yields a total score and subscale scores for each of the local symptoms. The *two tam*, *par* and *kumu* scores can be grouped to gain a score for 'depression-like' syndromes. The *ma lwor* score was used as a measurement of an 'anxiety-like' syndrome. Items that occur in more than one subscale are listed only once in the APAI total problems score.

Betancourt et al. [32] examined reliability using standard test-retest and inter-rater methods. The total APAI score had strong internal consistency ($\alpha = .93$). Inter-rater reliability (.84) and test-retest reliability (.74) were also high. Score for the depression-like syndromes exhibited similar strong internal reliability (ranging from $\alpha = .84$ to .87). As the current study focused on internalising disorders, the *kwo maraco*/conduct disorder items were left out resulting in a 44-item adaptation of the original 60-item questionnaire. Test-test reliability was also found to be strong in the current pilot study ($r = .835$, $p < .001$) and internal consistency was high ($\alpha = .93$).

As the APAI assesses mental health according to the local cultural understanding, we did not feel it was meaningful to determine a cut-off point for clinically significant 'depression/anxiety-like' symptoms. The data were used to determine if there was an association between culturally defined internalising mental health difficulties, trauma exposure and PTSS.

Data analysis

Tests for data normality by examining skew, kurtosis and a histogram were inconclusive. A normal distribution was assumed based on the previous findings of Betancourt et al. [32] and parametric tests (Pearson correlation and *t* test) were applied.

Results

War experiences

The young people witnessed a wide range of horrific events during the war as presented in Table 1. These 15 traumatic events were identified by local professionals as a comprehensive list of war-related events that the adolescents may have been exposed to. Of a possible 15, the mean number of different traumatic events experienced was 8 (SD = 4.04). Table 1 shows the number and percentage of adolescents who witnessed each type of traumatic event. A large majority of both boys and girls had seen dead bodies and had a friend or family member killed. Boys were more

Table 2 Overview of table showing mean scores on the IES-R and APAI main scales and subscales ($n = 205$)

	Mean	Standard deviation	Min/max	25th percentile	75th percentile
IES-R total	30.53	20.528	0/78	15	44
Intrusion	9.94	8.401	0/32	4	15
Avoidance	11.53	8.096	0/30	4.50	17
Hyperarousal	9.06	6.115	0/24	4	13.50
APAI total	36.37	22.231	0/111	19	51
TwoTam	14.65	9.449	0/42	7	21
Kumu	10.65	7.189	0/35	5	15
Par	13.13	8.986	0/50	7	17.50
Depression sub-scale	29.63	18.801	0/93	15	41
MaLwor	9.28	6.128	0/31	4	14

likely than girls to report witnessing killing or threats to kill. The low number of adolescents who report witnessing rape may be in part due to the stigma associated with being a victim of rape. It is well documented that the LRA have used rape as a 'tool of war' in northern Uganda [13].

Post-traumatic stress symptoms (IES-R)

The mean score on the IES-R was 30.53 (SD = 20.52). Table 2 displays the mean scores on the IES-R subscales and the proportion of high and low scores. There was no significant gender difference in post-traumatic stress symptomatology. 57% ($n = 117$) of students were found to have scores higher than 24 on the IES-R. This was the cut-off point used by Derluyn et al. [12] to determine clinically significant levels of post-traumatic stress symptoms.

PTSS and exposure

The number of different types of traumatic event witnessed There was a strong positive correlation between the number of different types of traumatic events witnessed and levels of PTSS ($r = .621, p < .001$). The types of traumatic events most strongly correlated with high levels of PTSS were 'Ambush' ($r = .524, p < .001$), 'Threats to kill' ($r = .455, p < .001$) and 'Friend or family member killed' ($r = .437, p < .001$).

Abducted versus non-abducted Adolescents who were abducted had significantly higher levels of PTSS than those who were not abducted ($t = 7.319, p < .001$).

Depression-like and anxiety-like symptoms (APAI)

The overall mean score on the APAI was 36.37 (SD = 22.2). Table 2 displays the mean scores on the APAI subscales and the proportion of high and low scores. The mean for the males in the sample was 33.15, while for

females it was 40.24. In contrast to their scores on the IES-R, girls in the sample had significantly higher levels of depression and anxiety-like symptoms than the boys ($t = 2.295, p < .05$).

Depression/anxiety and exposure

The number of different types of traumatic event witnessed There was a significant correlation between the number of different types of traumatic events witnessed and levels of depression/anxiety-like symptoms ($r = .466, p < .001$). The types of traumatic events most strongly correlated with depression and anxiety-like symptoms were 'Ambush' ($r = .450, p < .001$), 'Threats to kill' ($r = .383, p < .001$) and 'Friend or family member killed' ($r = .377, p < .001$). These are the same three types most strongly associated with PTSS.

Abducted versus non-abducted Adolescents who were abducted were significantly more likely to have high levels of depression/anxiety-like symptoms than those who were not abducted ($t = 5.745, p < .001$).

Association between post-traumatic stress and internalising symptoms

There was a strong positive correlation between PTSS as measured on the (IES-R) and depression/anxiety-like symptoms as measured on the APAI ($r = .743, p < .001$).

Discussion

This cross-sectional study provided an indication of the prevailing psychological distress in adolescents 4 years after the end of the conflict in northern Uganda. While levels of post-traumatic stress were lower than those found in similar studies during the war, many young people were

still suffering from nightmares, flashbacks, hyperarousal, depression and anxiety. There was a strong correlation between the number of traumatic event types experienced and the likelihood of developing psychological distress. Adolescents who were abducted were more likely to have developed PTSS and depression and anxiety-like symptoms than those who had been affected by the war but did not serve as 'child soldiers'. These findings suggest that there is something intrinsic to the experience of being a child soldier that is a specific risk factor for the emergence of post-traumatic distress. The study also provided further evidence of the close link between post-traumatic stress and internalising difficulties. Further investigation, including longitudinal research, is required to address some of the omissions of the current study such as the impact of daily stressors, trajectory of traumatic stress over time, and consideration of resilience factors.

Trauma levels

As previously mentioned, other research was carried out in northern Uganda when the conflict was at its peak [9, 13, 36]. These studies took place in the same district (Gulu) with a similar population. Assessment of a similar sample, 4 years after the war has ended, enables us to make some suggestions regarding the effect of war on psychological wellbeing over time. Symptoms of post-traumatic stress were lower than those found during the war using the same screening measure with the same cut-off point for clinical significance [12]. Alongside the fact that many children did not display psychological distress despite witnessing many horrific atrocities, this indicates the presence of individual and community resilience. It may have been that attendance at this boarding school had also brought healing to some of the adolescents.

Traumatic exposure

The strong relationship between level of traumatic exposure and psychological distress may have implications for future research and intervention. As found in previous research [4, 14], experiencing an increasing number of different types of traumatic event directly increased the likelihood of psychological distress. The current findings also concurred with previous studies that related more severe types of trauma, such as the violent death of a family member, witnessing someone being killed, or a perception of a direct threat, to higher levels of distress [17, 19].

The effect of cumulative exposure makes former child soldiers a highly vulnerable group, because the 'building-block effect' may cause protective factors such as social support or personal resilience to be weakened. Indeed it is

perhaps not surprising that former child soldiers, who spent years 'in the bush' witnessing and taking part in the most horrific atrocities, are likely to be more traumatised than those who faced less extreme situations. Betancourt et al. [38] singled out being raped, and wounding or killing others as three events that result in qualitatively different psychological outcomes. This has implications for intervention as former child soldiers may require specific treatment that is not necessary for other war-affected youth. Care should be taken, however, as there is evidence that singling out child soldiers for intervention may serve to increase stigma and community resentment [37]. This decision of whether to separate former child soldiers for intervention should be made in conjunction with local representatives through consideration of sociodemographic factors and the trauma levels derived from screening.

Depression/anxiety levels

This study provided further evidence of the close link between post-traumatic stress and depression. According to Betancourt and colleagues [32], the locally described mental health difficulties of *two tam, par* and *kumu* resemble Western definitions of mood disorders and depression. In agreement with previous research that used a Western classification of depression, the levels of these symptoms were generally higher when the same child had high levels of PTSS. Because this is a cross-sectional study, the direction of this relationship cannot be clearly established. However, in line with the findings of Smith et al. [19], the correlations in the current study showed that depression was much more strongly associated with PTSS than with exposure. This was the case whatever the measure of 'exposure'—the number of different types of traumatic event witnessed; whether or not they were abducted; or the length of time they were abducted for. This seems to suggest that depressive feelings among these adolescents are a secondary consequence of PTSS and mediating life events, rather than of exposure directly. Longitudinal research is required to conclusively determine the direction of this relationship. Since most trauma interventions focus narrowly on PTSD, more research is required to determine the implications of the co-morbidity with depression in choosing assessment tools and intervention programmes.

The fact that girls were significantly more likely than boys to suffer from depression and anxiety was interesting, especially since this gender difference was not reflected in the levels of PTSS. Two recent studies on gender differences among war-affected adolescents from developing countries found that girls are more vulnerable to internalising problems, while boys are more vulnerable to externalising problems [24, 37]. This pattern of boys'

externalising and girls internalising their responses to maltreatment has also been observed in Western culture [35]. The gender differences in the current sample may be partially explained by the stage of pubertal developmental or by culture-bound gender-expectations of acceptable and unacceptable behaviour. Mels and colleagues [24] added a further insight when they noted that boys tended to develop externalising behaviour problems as a result of higher traumatic exposure, while girls tended to develop internalising problems when faced with cumulative daily stressors. Limited access to food, clothing and healthcare facilities are examples of daily stressors that impact more on females than males, since girls and women are responsible for feeding, clothing and tending to the sick in traditional cultures. Thus, on-going daily stressors linked to traditional gender work roles may militate against girls' recovery from PTSD. For the girls in this study, the fact that they are in a boarding school setting and hence detached from these traditional roles may be exacerbating these daily stressors.

Assessing psychological distress in post-conflict areas

The results of this mental health screening further emphasise the importance of early identification and treatment of children and adolescents at risk of psychopathology after exposure to war-related trauma. However, mental health, especially in children, does not seem to be a high priority for the governmental and non-governmental organisations that operate in post-conflict zones. For example, northern Uganda has a psychiatrist-to-population ratio of 1:1.3 million [39]. Assessment in low-income, war-affected areas can be extremely challenging. The context and culture in which assessment is occurring must be considered. Given the extreme circumstances of traumatised individuals, data gathering for research purposes may be seen as unsympathetic, foreign and exploitive; and may be met with resistance [27].

Completing the screening in a school that was specifically designated for war-affected youth provided a group that was relatively homogeneous. Some of the teachers in the school were also trained counsellors. These teachers played a vital role in screening, acting as interpreters and also as co-facilitators of the psycho-education seminar. The teachers had already developed strong positive relationships with the adolescents and played an important role in advising the authors on the cultural applicability of each measure and in adapting the wording of each item to aid understanding. There are many areas of trauma research that could take place in a school such as this one. Questions remain about how trauma affects education: what is the nature of any impairment in attention, reasoning, executive capacity and/or general intellectual functioning?

Limitations and future direction

Many of the adolescents in the school were suffering from an array of daily stressors such as poverty, grief, responsibility for younger children and abuse. These are very likely to have had some effect on the pre and post measures [24]. The written and oral guidance on the screening measures stated the adolescents should tick a particular item in accordance with how frequently the statement was true for them over the previous 7 days. Therefore, for some adolescents, the depression and anxiety-like symptoms rated highly on the checklist may have been due in part to their daily stressors as well as a traumatic event related to the war.

It is also important to note that, while the current study found evidence for symptoms that are comparable to Western conceptualisations of trauma, depression and anxiety, this does not essentially imply Western classifications are relevant in their entirety. Every conflict or disaster occurs in its own unique cultural context. We must be mindful of Summerfield's statements that 'every culture has its own frameworks for mental health, and norms for help-seeking at times of crisis' and 'social healing and the remaking of worlds cannot be managed by outsiders' [40]. However, there does seem to be a consensus that post-traumatic stress reactions in children and adolescents are more similar across cultures than they are different [41]. Hence, it seems that there is an obligation to balance concerns about cultural applicability with the principle that scientific methods of assessment and evidence-based treatment should be available for all, especially those in war-affected communities who need it most. As posited by Barenbaum et al. [1], the best approach to assessing and treating war-affected children should be to draw both on principles of cultural specificity, and cross-cultural universality.

The current study did not address the question of why a number of those who had been exposed to the most horrific situations during the war had not developed symptoms of post-traumatic stress, depression or anxiety. It would be of great interest, especially when considering intervention, to understand better how individuals cope and rebuild their lives despite living through war-related trauma. This concept of resilience has proved difficult to conceptualise and measure. Even when conceptualised as an outcome there is no consensus as to whether it should be defined as the absence of negative outcomes, presence of positive outcome, or a combination of both [36]. However, the fact that 43% of the adolescents in this study did not display clinically significant symptoms of post-traumatic stress suggests that resilience factors should be considered in future studies especially when planning subsequent intervention. The operational definition employed by Klasen et al. [36]

may provide the best modus operandi. They defined 'posttraumatic resilience' as the 'absence of clinically significant psychopathology in the aftermath of severe trauma exposure' and considered factors (such as dispositions, positive future orientation, guilt, motivations and perceived external support) which may be associated with a lack of psychopathology.

It is vital that greater effort is made to develop assessment tools and intervention therapies within home countries and cultures of those being assessed and treated. At present, there are few assessment tools (the APAI notwithstanding) and evidence-based trauma treatment approaches that have been developed within the cultures of the war-affected populations and based on a local understanding of trauma and psychological distress [42].

Conclusion

This study has demonstrated that war-affected adolescents may continue to suffer from considerable psychological stress in the years following the cessation of conflict. A building-block effect of multiple trauma can exacerbate this distress especially for those exposed to the most extreme violence such as child soldiers. More research is required to expand the current knowledge base of the effects of war-related trauma on children and adolescents. Crucially, this knowledge needs to be built into interventions to treat traumatised young people. To this end, the results of the current study would suggest tailoring intervention according to presentation with consideration of: the severity of post-traumatic symptoms; co-existing mental health difficulties; gender; the number and type of traumatic events exposed to; and personal or community resilience factors. An evidence-based group intervention that could be delivered in a school would be of particular help to a population such as that in the current study, where the sheer numbers of war-affected people make individual trauma therapy almost impossible. It is hoped that, as our understanding improves, so too will the efficacy of our attempts to treat the debilitating effects of trauma on young people.

Acknowledgments With sincere appreciation and thanks to all the young people who participated in this study, to their teachers and caregivers and to our friends in Gulu without whom this study would not have been possible. The authors declare that they have no financial, personal, or professional interests that could be construed to have influenced their article.

References

- Barenbaum J, Ruchkin V, Schwab-Stone M (2004) The psychosocial aspects of children exposed to war: practice and policy initiatives. *J Child Psychol Psychiatry* 45(1):41–62
- Neuner F, Catani C, Ruf M, Schauer E, Schauer M, Elbert T (2008) Narrative exposure therapy for the treatment of traumatized children and adolescents (KidNET): from neuro-cognitive theory to field intervention. *Child Adolesc Psychiatry Clinics N Am* 17(3):641–664
- Dyregrov A, Yule W (2006) A review of PTSD in children. *Child Adolesc Mental Health* 11(4):176
- Schauer M, Schauer E (2010) Trauma-focused public mental health interventions: a paradigm shift in humanitarian assistance and aid work. In: Marz E (ed) *Trauma rehabilitation after war and conflict: community and individual perspectives*. Springer, New York
- Coalition to Stop the Use of Child Soldiers (CSUCS) (2008) *Child Soldiers Global Report 2008*. Coalition to Stop the Use of Child Soldiers (CSUCS), London
- Machel G (2001) *The impact of war on children*. Hurst & Company, London
- Boothby NG, Knudsen CM (2000) Waging a new kind of war: children of the gun. *Sci Am* 282:60–65
- Machel G (1996) *Impact of armed conflict on children*. Report of the Expert of the Secretary General of the United Nations. New York
- Amone-P'Olak K, Garnefski N, Kraaij V (2007) The impact of war experiences and physical abuse on formerly abducted boys in northern Uganda. *S Afr Psychiatry Rev* 10(2):76–82
- Machel G (2009). Machel study 10-year strategic review: children and conflict in a changing world. Office of the special representative of the Secretary-General for children and armed conflict in collaboration with UNICEF UN children's fund. www.unicef.org/publications/index_49985.html. Accessed 2 July 2010
- Annan J, Brier M, Aryemo F (2009) From 'rebel' to 'returnee': the reintegration of young soldiers in northern Uganda. *J Adolesc Res* 24:639–667
- Derluyn I, Broekaert E, Schuyten G, De Temmerman E (2004) Post-traumatic stress in former Ugandan child soldiers. *Lancet* 363(9412):861–863
- Amone-P'Olak K (2005) Psychological impact of war and sexual abuse on adolescent girls in northern Uganda. *Intervention* 3(1):33–45
- Karunakara UK, Neuner F, Schauer M, Singh K, Hill K, Elbert T et al (2004) Traumatic events and symptoms of post-traumatic stress disorder amongst Sudanese nationals, refugees and Ugandans in the West Nile. *Afr Health Sci* 4(2):83–93
- Bayer C, Klasen F, Adam H (2007) Association of trauma and PTSD symptoms with openness to reconciliation and feelings of revenge among former Ugandan and Congolese Child Soldiers. *J Am Med Assoc* 298(5):555–559
- Schauer E (2008) *Trauma treatment for children in war. Build-up of an evidence-based large-scale mental health intervention in North Eastern Sri Lanka*. Dissertation (PhD) thesis, University of Konstanz, Konstanz, Germany, Department of Clinical Psychology and Neuropsychology. <http://www.ub.unikonstanz.de/kops/volltexte/2008/5424/>
- Ehnholt KA, Yule W (2006) Practitioner review: assessment and treatment of refugee children and adolescents who have experienced war-related trauma. *J Child Psychol Psychiatry* 47(12):1197–1210
- Mels C, Derluyn I, Broekaert E, Rosseel Y (2009) Screening for traumatic exposure and posttraumatic stress symptoms in adolescents in the war-affected eastern democratic republic of Congo. *Arch Pediatr Adolesc Med* 163(6):525
- Smith P, Perrin S, Yule W, Hacam B, Stuvland R (2002) War exposure among children from Bosnia–Herzegovina: psychological adjustment in a community sample. *J Trauma Stress* 15:147–156

20. Thabet AA, Abed Y, Vostanis P (2002) Emotional problems in Palestinian children living in a war zone: a cross-sectional study. *Lancet* 359(9320):1801–1804
21. Goenjian AK, Molina L, Steinberg AM, Fairbanks LA, Alvarez ML (2001) Posttraumatic stress and depressive reactions among Nicaraguan adolescents after hurricane Mitch. *Am J Psychiatry* 158(5):788–794
22. Pham PN, Vinck P, Kinkodi DK, Weinstein H (2010) Sense of cohesion and its association with exposure to traumatic events, post-traumatic stress disorder, and depression in eastern Democratic Republic of the Congo. *J Trauma Stress* 23(3):313–321
23. Hepinstall E, Sethna V, Taylor E (2004) PTSD and depression in refugee children: association with premigration trauma and postmigration stress. *Eur Child Adolesc Psychiatry* 13:373–380
24. Mels C, Derluyn I, Broekaert E, Rosseel Y (2010) Community-based cross-cultural adaptation of mental health measures in emergency settings: validating the IES-R and HSCL-37A in eastern Democratic Republic of Congo. *Soc Psychiatry Psychiatr Epidemiol* 45(9):899–910
25. Taylor S, Thordarson D, Maxfield L, Fedoroff I, Lovell K, Ogradniczuk J (2003) Comparative efficacy, speed, and adverse effects of three PTSD treatments. *J Consult Clin Psychol* 71(2):330–338
26. Saigh PA, Mroueh M, Bremner JD (1997) Scholastic impairments among traumatized adolescents. *Behav Res Ther* 35(5):429–436
27. Laor N, Wolmer L (2002) Children exposed to disaster: the role of the mental health professional. In: Lewis M (ed) *Child and adolescent psychiatry: a comprehensive textbook*. Lippincott Williams & Wilkins, Philadelphia, pp 925–937
28. Joshi PT, O'Donnell DA (2003) Consequences of child exposure to war and terrorism. *Clin Child Fam Psychol Rev* 2003 6(4):275–292
29. Famularo R, Fenton T, Kinscherff R, Augustyn M (1996) Psychiatric comorbidity in childhood post traumatic stress disorder. *Child Abuse Negl* 20:953–961
30. Saltzman WR, Layne CM, Steinberg AM, Arslanagic B, Pynoos RS (2003) Developing a culturally-ecologically sound intervention program for youth exposed to war and terrorism. *Child Adolesc Psychiatr Clin N. Am* 12:319–342
31. Weiss DS, Marmar CR (1997) The impact of events scale-revised. In: Wilson JP, Keane TM (eds) *Assessing Psychological Trauma and PTSD*. Guilford Press, New York
32. Betancourt T, Bass J, Borisova I, Neugebauer R, Speelman L, Onyango G, Bolton P (2009) Assessing local instrument reliability and validity: a field-based example from northern Uganda. *Soc Psychiatry Psychiatr Epidemiol* 44(8):685–692
33. Schauer M, Neuner F, Elbert T (2005) Narrative exposure therapy: a short-term intervention for traumatic stress disorders after war, terror, or torture. Hogrefe & Huber, Göttingen
34. Smith P, Dyregrov A, Yule W, Perrin S, Gjestad R, Gupta L (2002) *Children and war: teaching recovery techniques*. Foundation for Children and War, Bergen. Retrieved 15 May 2010 <http://www.childrenandwar.org>
35. Broidy L, Agnew R (1997) A general strain theory perspective. *J Res Crime Delinq* 34:275–306
36. Klasen F, Oettingen G, Daniels J, Post M, Hoyer C, Adam H (2010) Posttraumatic resilience in former Ugandan child soldiers. *Child Dev* 81(4):1096–1113
37. Derluyn I, Broekaert E, Schuyten G (2008) Emotional and behavioural problems in migrant adolescents in Belgium. *Eur Child Adolesc Psychiatry* 17(1):54–62
38. Betancourt T, Agnew-Blais J, Gilman S, Williams D, Ellis B (2010) Past horrors, present struggles: the role of stigma in the association between war experiences and psychosocial adjustment among former child soldiers in Sierra Leone. *Soc Sci Med* (1982) 70(1):17–26
39. Ovuga E, Boardman J, Wasserman D (2007) Integrating mental health into primary health care: local initiatives from Uganda. *World Psychiatry* 6:60–61
40. Summerfield D (1999) A critique of seven assumptions behind psychological trauma programmes in war-affected areas. *Soc Sci Med* 48(10):1449–1462
41. Smith P, Perrin S, Dyregrov A, Yule W (2003) Principal components analysis of the impact of event scale with children in war. *Pers Individ Differ* 34:315–322
42. Crumlish N, O'Rourke K (2010) A systematic review of treatments for post-traumatic stress disorder among refugees and asylum-seekers. *J Nerv Ment Dis* 198(4):237

APPENDIX A26: Accepted Conference Abstracts

Conference:

Annual meeting for the International Society for Behavioural Nutrition – Melbourne, 2011

Symposia title:

Physical Activity Research in Developing Countries – What are the opportunities and challenges?

Presenters:

J. Richards, K. Siefken, A. McManus, C. Foster.

Presentation focus:

Young Adolescents in a Post-Conflict Setting

Purpose:

It is thought that decreasing physical activity (PA) levels are contributing to the emergence of the double burden of disease in low- and middle-income countries (LMICs). Behavioural, environmental and socio-economic characteristics of post-conflict settings may expedite this epidemiological transition. The objective of this study was to assess the feasibility of obtaining reliable and valid measures of adolescent PA, physical fitness (PF) and associated NCD risk factors in northern Uganda.

Methods:

1655 children from 10 schools in Gulu were assessed using the standardised “beep” test, standing broad jump, BMI-for-age and Acholi Psychosocial Assessment Instrument. 250 of these children voluntarily registered for a 9-week “sport-for-development” football league in Gulu Municipality. 50% of them were randomly selected as league participants (I) and the others were wait-listed (W). A sub-sample of 28xI and 28xW were randomly selected for accelerometry measurement during the intervention period.

Results:

Post-conflict settings within LMICs present numerous challenges when utilising common measurement tools and applying rigorous scientific methodology. Several infrastructure and resource limitations accompanied social, language and cultural differences that complicated data collection. This was compounded by less predictable factors that included teaching practices, timetabling and attendance patterns in schools as well as the capacity of the local population to recall personal information and plan ahead.

Conclusion:

There are several challenges to measuring adolescent PA, PF and NCD risk factors in post-conflict zones of LMICs. How this limits the establishment of baseline levels in this setting and the impact evaluation of a “sport-for-development” intervention are yet to be determined.

Conference:

Annual meeting for the International Society for Behavioural Nutrition – Melbourne, 2011

Poster title:

The reliability of physical fitness testing for children in Gulu, Northern Uganda – a post-conflict setting in a low-income country

Authors:

J. Richards, C. Foster.

Purpose:

Transferring standardised measurement methods for physical fitness from developed countries is often limited by the local environment and resources available in low-income countries (LIC). Our study examined the feasibility and reliability of assessing a range of indicators for physical fitness in the post-conflict setting of Gulu.

Methods:

We adapted standardised protocols for assessing height (H), weight (W), standing broad jump (SBJ) and the multi-stage “beep” fitness test (MFT) according to the local context and available resources. Repeat measures were taken by trained local assessors on a convenient sample of one class (n=70, 12-16 years) from a boarding school for war-affected children. Single measure intra-class correlation coefficients (ICC – two-way random, absolute agreement) were calculated to assess the inter- and intra-tester reliability.

Results:

The modified measurement methods had to: 1) be rapidly set-up and disassembled; 2) be capable of efficiently measuring many subjects in variable environments; 3) utilise materials that are locally available and sustainable; 4) follow protocols that are culturally appropriate; 5) enable school access by the assessors transporting the equipment by motorbike.

Initial measures were taken for H and W (n=70), SBJ (n=54) and MFT (n=57). Repeat measures were taken for H and W (n=70), SBJ (n=49) and MFT (n=55). Assessment of the inter-tester and intra-tester reliability found strong correlations that were statistically significant (ICC: H>0.9 [p<0.000]; W>0.9 [p<0.000]; SBJ>0.9 [p<0.000]; MFT>0.8 [p<0.000]).

Conclusion:

Measuring physical fitness (H, W, SBJ, MFT) using local resources in the post-conflict setting of Gulu required creative adaptation of standardised assessment procedures. These modified techniques have good inter- and intra-tester reliability.

REFERENCES (Appendices)

1. D. Pedersen. *Political Violence, Ethnic Conflict, and Contemporary Wars: Broad Implications for Health and Social Well-Being*. Soc Sci Med, 2002. **55**(2): p. 175-90.
2. H.A. Ghobarah, P. Huth, B. Russett. *The Post-War Public Health Effects of Civil Conflict*. Soc Sci Med, 2004. **59**(4): p. 869-84.
3. B. Roberts, P. Patel, M. McKee. *Noncommunicable Diseases and Post-Conflict Countries*. Bull World Health Organ, 2012. **90**: p. 2-2A.
4. P. B. Spiegel, F. Checchi, S. Colombo, E. Paik. *Health-Care Needs of People Affected by Conflict: Future Trends and Changing Frameworks*. Lancet, 2010. **375**(9711): p. 341-5.
5. Tele Atlas, Geocentre Consulting. *Map of Africa (Modified)*. 2011, Google Maps.
6. United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA). *Map of Uganda (Modified)*. 2009, UNOCHA.
7. United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA). *Map of Gulu District (Modified)*. 2009, UNOCHA.
8. Tele Atlas, Geocentre Consulting. *Map of Gulu Municipality (Modified)*. 2011, Google Maps.
9. M. de Onis, A. Onyango, E. Borghi, A. Siyam, C. Nishida, J. Siekmann. *Development of a Who Growth Reference for School-Aged Children and Adolescents*. Bulletin of the World Health Organization, 2007. **85**: p. 661-668.
10. I. Price. *Analysing the Data: Determining If Skewness and Kurtosis Are Significantly Non-Normal*. 2000, University of New England: Armidale.
11. H.M. Park. *Univariate Analysis and Normality Test Using Sas, Stata, and Spss*. 2008, University Information Technology Services, Center for Statistical and Mathematical Computing, Indiana University: Bloomington.
12. G. Keppel. *Design and Analysis: A Researcher's Handbook*. 4th ed. 2004, Englewood Cliffs, N.J: Prentice-Hall. xii, 611 p.
13. G. Bravo, P. Louise. *Estimating the Reliability of Continuous Measures with Cronbach's Alpha or the Intraclass Correlation Coefficient: Toward the Integration of Two Traditions*. J Clin Epidemiol, 1991. **44**(4/5): p. 381-390.
14. D.C. Montgomery. *Design and Analysis of Experiments*. 7 ed. 2009, Hoboken: John Wiley and Sons.
15. J.M. Bland, D.G. Altman. *Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement*. Lancet, 1986. **1**(8476): p. 307-10.
16. H.A. David. *Upper 5 and 1% Points of the Maximum F-Ratio*. Biometrika, 1952. **39**(3/4).