

РЕПУБЛИКА СЕВЕРНА МАКЕДОНИЈА  
УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“ – СКОПЈЕ  
ФАКУЛТЕТ ЗА ФИЗИЧКО ОБРАЗОВАНИЕ, СПОРТ И ЗДРАВЈЕ



ТРЕТ ЦИКЛУС СТУДИИ ПО КИНЕЗИОЛОГИЈА

**ПРИМЕНЛИВОСТ И РЕЛИЈАБИЛНОСТ НА ТЕСТОВИТЕ ЗА  
ПРОЦЕНУВАЊЕ НА ФИЗИЧКИОТ И МОТОРИЧКИОТ РАЗВОЈ КАЈ  
МАКЕДОНСКИТЕ УЧЕНИЦИ: МАКФИТ СТУДИЈА**

(ДОКТОРСКА ДИСЕРТАЦИЈА)

Кандидат:

**М-р Миодраг Тодоровиќ**

Ментор:

**Проф. д-р Серјожа Гонтарев**

Скопје, 2026 година

**Комисија за одбрана на докторската дисертација:**

**Проф. д-р Серјожа Гонтарев (ментор)**

*Факултет за физичко образование, спорт и здравје - Скопје*

**Проф. д-р Жарко Костовски (претседател)**

*Факултет за физичко образование, спорт и здравје - Скопје*

**Проф. д-р Ленче Алексовска Величковска (член)**

*Факултет за физичко образование, спорт и здравје - Скопје*

**Проф. д-р Борче Даскаловски (член)**

*Факултет за физичко образование, спорт и здравје - Скопје*

**Проф. д-р Влатко Неделковски (член)**

*Факултет за физичко образование, спорт и здравје - Скопје*

**Наука во која се стекнува звањето:**

**ОПШТЕСТВЕНИ НАУКИ – КИНЕЗИОЛОГИЈА**

**Научно звање :**

**ДОКТОР НА КИНЕЗИОЛОГИЈА**

## БЛАГОДАРНОСТ

Искрено им благодарам на сите кои ме поддржуваа, охрабруваа и веруваа во мене во текот на изработката на оваа докторска дисертација. Нивната помош, доверба и поттик беа моја сила и инспирација на овој пат. Особена благодарност до мојот ментор, проф. д-р Серјожа Гонтарев, чие разбирање, поддршка и мотивирачки зборови во текот на целиот истражувачки процес беа од непроценливо значење и придонесоа за конечна изработка на овој труд.

## АПСТРАКТ

Физичкиот и моторичкиот развој на децата и адолесцентите претставува клучен индикатор за тековното здравје и идниот кардиометаболен ризик, но неговото валидно следење бара употреба на тестови кои се релијабилни, објективни, безбедни и реално применливи во училишни услови. Во тој контекст, целта на ова истражување е да се процени применливоста и релијабилноста на батеријата МАКФИТ за проценување на физичкиот и моторичкиот развој кај македонските ученици, како и да се утврди дали резултатите што ги обезбедува оваа батерија се доволно стабилни и чувствителни за употреба во педагошка и јавно здравствена практика. Истражувањето е спроведено на примерок од 102 адолесценти од средни училишта во Скопскиот регион, од кои 62 момчиња и 40 девојчиња, психофизички здрави и редовни на наставата по физичко и здравствено образование. Применети беа антропометриски мерки (телесна висина, телесна тежина, индекс на телесна маса, обем на половина) и теренски тестови за физички фитнес (флексибилност, скок во далечина од место, сила на стисок на дланка, силова издржливост на трупот и агилност, брзина и координација). Сите тестови беа администрирани двапати во интервал од една седмица. Статистичката обработка опфати дескриптивни показатели, проверка на нормалност, интракласни корелациски коефициенти за процена на релативната релијабилност, параметри на апсолутна релијабилност (стандардна грешка на мерење и минимално детектибилна промена), тестови на разлики меѓу првото и второто мерење и Bland–Altman анализа на согласност. Резултатите покажаа дека антропометриските мерки и поголемиот број моторички тестови имаат добра до одлична тест ретест релијабилност и кај момчињата и кај девојчињата, со високи интракласни корелациски коефициенти и релативно ниски вредности на стандардна грешка и варијабилност во процент. Кај дел од моторичките тестови беше евидентиран мал, но систематски ефект на запознавање со тестот, додека кај основните антропометриски показатели не се регистрираа значајни систематски разлики. Анализата на согласност укажа дека резултатите од ретестот можат практично да го заменат првото мерење при следење на истото дете, без промена на интерпретацијата за неговиот вистински статус. Истовремено, беше потврден очекуван полов диморфизам во корист на момчињата во силовите и брзинско експлозивните способности. Батеријата се покажа и како високо применлива: тестирањето е временски економично, лесно спроведливо од страна на наставниците и добро прифатено од страна

на учениците, без сериозни несакани настани. Заклучно, МАКФИТ батеријата претставува релијабилен, безбеден и практично изводлив инструмент за процена на физичкиот и моторичкиот развој кај македонските ученици и може да се користи за индивидуално следење, изработка на норми по пол и возраст и планирање на таргетирани интервенции во рамките на образовниот и јавно здравствениот систем.

**Клучни зборови:** физички фитнес, релијабилност, тест ретест, адолесценти, МАКФИТ батерија

## ABSTRACT

The physical and motor development of children and adolescents is a key indicator of current health and future cardiometabolic risk, yet its valid monitoring requires the use of tests that are reliable, objective, safe, and genuinely feasible in school settings. In this context, the aim of the present study was to evaluate the feasibility and reliability of the MAKFIT test battery for assessing physical and motor development in Macedonian schoolchildren, and to determine whether the outcomes provided by this battery are sufficiently stable and sensitive for use in pedagogical and public health practice. The study was conducted on a sample of 102 adolescents from secondary schools in the Skopje region, including 62 boys and 40 girls, all psychophysically healthy and regularly attending physical and health education classes. The applied measures comprised anthropometric indicators (body height, body weight, body mass index, waist circumference) and field-based physical fitness tests (flexibility, standing long jump, handgrip strength, trunk muscular endurance, and agility, speed and coordination). All tests were administered twice, with a one-week interval between measurements. Statistical analysis included descriptive indicators, tests of normality, intraclass correlation coefficients to estimate relative reliability, parameters of absolute reliability (standard error of measurement and minimal detectable change), tests of mean differences between the first and second measurement, and Bland–Altman agreement analysis. The results showed that the anthropometric measures and the majority of motor tests demonstrated good to excellent test–retest reliability in both boys and girls, with high intraclass correlation coefficients and relatively low values of standard error and percentage variability. A small but systematic familiarization effect was observed for some motor tests, whereas no statistically significant systematic differences were registered for the core anthropometric indicators. The agreement analysis indicated that retest scores can practically replace baseline values when monitoring the same child over time, without altering the interpretation of the child’s true status. At the same time, the expected sex dimorphism in favour of boys was confirmed in strength-related and speed–power abilities. The battery also proved to be highly feasible: testing was time-efficient, easy to implement by physical education teachers, and well accepted by students, with no serious adverse events reported. In conclusion, the MAKFIT battery represents a reliable, safe, and practically implementable instrument for assessing physical and motor development in Macedonian school-aged youth, and can be used for individual monitoring, development of sex- and age-specific norms, and for planning targeted interventions within the educational and public health systems.

**Keywords:** physical fitness, reliability, test–retest, adolescents, MAKFIT battery

# СОДРЖИНА

<b>1. ВОВЕД .....</b>	<b>1</b>
<b>2. ДОСЕГАШНИ ИСТРАЖУВАЊА.....</b>	<b>5</b>
<b>3. ПРОБЛЕМ, ПРЕДМЕТ, ЦЕЛИ И ХИПОТЕЗИ НА ИСТРАЖУВАЊЕТО .....</b>	<b>11</b>
3.1. ПРОБЛЕМ, ПРЕДМЕТ И ЦЕЛИ НА ИСТРАЖУВАЊЕТО .....	11
3.2. ХИПОТЕЗИ НА ИСТРАЖУВАЊЕТО.....	13
<b>4. МЕТОДИ НА ИСТРАЖУВАЊЕТО.....</b>	<b>15</b>
4.1. ПРИМЕРОК НА ИСПИТАНИЦИ.....	15
4.2. ПРИМЕРОК НА ВАРИЈАБЛИ.....	16
4.2.1. Варијабли за проценка на антропометриските карактеристики .....	16
4.2.2. Варијабли за проценка на моторичките способности .....	17
4.3. ОПИС НА АНТРОПОМЕТРИСКИТЕ МЕРЕЊА.....	17
4.4. УСЛОВИ И ТЕХНИКА НА МЕРЕЊЕ НА АНТРОПОМЕТРИСКИТЕ КАРАКТЕРИСТИКИ .....	18
4.5. ОПИС НА ВАРИЈАБЛИ ЗА ПРОЦЕНКА НА МОТОРИЧКИТЕ СПОСОБНОСТИ.....	19
4.6. УСЛОВИ И ТЕХНИКА НА МЕРЕЊЕ НА МОТОРИЧКИТЕ СПОСОБНОСТИ .....	23
4.7. УТВРДУВАЊЕ НА ПРИМЕЛИВОСТ НА МЕРКИТЕ И ТЕСТОВИТЕ ВО УЧИЛИШНО ОКРУЖУВАЊЕ .....	23
4.8. УТВРДУВАЊЕ НА БЕЗБЕДНОСТ НА МЕРКИТЕ И ТЕСТОВИТЕ .....	24
4.9. УТВРДУВАЊЕ НА ТЕСТ-РЕТЕСТ РЕЛИЈАБИЛНОСТА НА МЕРКИТЕ И ТЕСТОВИТЕ .....	24
4.10. МЕТОД ЗА ОБРАБОТКА НА ПОДАТОЦИТЕ .....	25
<b>5. РЕЗУЛТАТИ .....</b>	<b>28</b>
5.1. ОСНОВНИ ДЕСКРИПТИВНИ СТАТИСТИЧКИ ПАРАМЕТРИ И НОРМАЛНОСТА НА ДИСТРИБУЦИЈАТА НА АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ОД ПРВОТО И ВТОРОТО МЕРЕЊЕ (ТЕСТ-РЕТЕСТ МЕРЕЊА) КАЈ МОМЧИЊАТА.....	29
5.2. ОСНОВНИ ДЕСКРИПТИВНИ СТАТИСТИЧКИ ПАРАМЕТРИ И НОРМАЛНОСТА НА ДИСТРИБУЦИЈАТА НА АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ОД ПРВОТО И ВТОРОТО МЕРЕЊЕ (ТЕСТ-РЕТЕСТ МЕРЕЊА) КАЈ ДЕВОЈЧИЊАТА.....	38
5.3. ТЕСТ-РЕТЕСТ РЕЛИЈАБИЛНОСТА И СОГЛАСНОСТА НА МАКФИТ БАТЕРИЈАТА, ОДДЕЛНО ЗА МОМЧИЊА И ДЕВОЈЧИЊА .....	47
5.4. КОМПАРАТИВЕН УВИД МЕЃУ МОМЧИЊАТА И ДЕВОЈЧИЊАТА НИЗ ТРИТЕ АНАЛИТИЧКИ РАМКИ.....	82
5.5. ПРАГ НА „РЕАЛНА ПРОМЕНА“ (MDC95/SDD) КАЈ МОМЧИЊА И ДЕВОЈЧИЊА ...	85
5.6. ICC SUMMARY (ICC(2,1), 95% CI) – МОМЧИЊА ВО ОДНОС НА ДЕВОЈЧИЊА (РЕЛАТИВНА РЕЛИЈАБИЛНОСТ) .....	88

5.7. РАЗЛИКИ ВО АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ПОМЕЃУ МОМЧИЊАТА И ДЕВОЈЧИЊАТА ВО ПРВОТО МЕРЕЊЕ.....	91
5.8. РАЗЛИКИ ВО АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ПОМЕЃУ МОМЧИЊАТА И ДЕВОЈЧИЊАТА ВО ВТРОТО (РЕТЕСТ) МЕРЕЊЕ .....	100
5.9. ПРИМЕНЛИВОСТА НА ПРЕДЛОЖЕНАТА БАТЕРИЈА МАКФИТ .....	108
5.10. БЕЗБЕДНОСТА НА ПРИМЕНЕТИТЕ АНТРОПОМЕТРИСКИ МЕРЕЊА И МОТОРИЧКИ ТЕСТОВИ .....	110
<b>6. ДИСКУСИЈА .....</b>	<b>112</b>
<b>7. ЗАКЛУЧОЦИ.....</b>	<b>129</b>
<b>8. ТЕОРЕТСКО И ПРАКТИЧНО ЗНАЧЕЊЕ НА ИСТРАЖУВАЊЕТО.....</b>	<b>134</b>
<b>ЛИТЕРАТУРА .....</b>	<b>138</b>
<b>ПРИЛОГ .....</b>	<b>147</b>

## ПОПИС НА ТАБЕЛИ

Табела 1. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од првото мерење кај момчињата .....	33
Табела 2. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од второто мерење (ретест мерења) кај момчињата .....	37
Табела 3. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од првото мерење кај девојчињата .....	41
Табела 4. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од второто мерење (ретест мерења) кај девојчињата .....	46
Табела 5. Тест-ретест релијабилност (релативна релијабилност (ICC) и систематска разлика кај момчињата).....	51
Табела 6. Апсолутна релијабилност (SEM, SEM%, MDC/SDD, nSDD, CV% кај момчињата)...	54
Табела 7. Bland–Altman согласност на варијаблите кај момчињата .....	59
Табела 8. Тест-ретест релијабилност (релативна релијабилност (ICC) и систематска разлика кај девојчињата).....	68
Табела 9. Апсолутна релијабилност (SEM, SEM%, MDC/SDD, nSDD, CV% кај девојчињата) .	72
Табела 10. Bland–Altman согласност на варијаблите кај девојчињата.....	76
Табела 11. Праг на „реална промена“ (MDC95/SDD) кај момчиња и девојчињаж .....	87
Табела 12. ICC summary (ICC(2,1), 95% CI) – момчиња vs девојчиња (релативна релијабилност).....	90
Табела 13. Разлики во антропометриските мерки и моторичките тестови помеѓу момчињата и девојчињата во првото мерење .....	93
Табела 14. Разлики во антропометриските мерки и моторичките тестови помеѓу момчињата и девојчињата во второто (ретест) мерење .....	102

## ПОПИС НА ГРАФИКОНИ

Графикон 1. Bland-Altman дијаграм на телесната висина кај момчињата.....	60
Графикон 2. Bland-Altman дијаграм на телесна тежина кај момчињата.....	60
Графикон 3. Bland-Altman дијаграм на индексот на телесна маса (BMI) кај момчињата.....	61
Графикон 4. Bland-Altman дијаграм на обем на половината (WC) кај момчињата.....	61
Графикон 5. Bland-Altman дијаграм на тестот претклон во сед кај момчињата.....	62
Графикон 6. Bland-Altman дијаграм на тестот скок во далечина од место кај момчињата.....	62
Графикон 7. Bland-Altman дијаграм на тестот динамометрија на дланка на десна рака кај момчињата.....	63
Графикон 8. Bland-Altman дијаграм на тестот динамометрија на дланка на лева рака кај момчињата.....	63
Графикон 9. Bland-Altman дијаграм на тестот динамометрија на дланка просечна вредност од десна и лева рака кај момчињата.....	64
Графикон 10. Bland-Altman дијаграм на тестот подигнување труп 30 секунди кај момчињата.....	64
Графикон 11. Bland-Altman дијаграм на тестот Чунесто трчање 4×10 m кај момчињата.....	65
Графикон 12. Bland-Altman дијаграм на телесната висина кај девојчињата.....	77
Графикон 13. Bland-Altman дијаграм на телесна тежина кај девојчињата.....	77
Графикон 14. Bland-Altman дијаграм на индексот на телесна маса (BMI) кај девојчињата.....	78
Графикон 15. Bland-Altman дијаграм на обем на половината (WC) кај девојчињата.....	78
Графикон 16. Bland-Altman дијаграм на тестот претклон во сед кај девојчињата.....	79
Графикон 17. Bland-Altman дијаграм на тестот скок во далечина од место кај девојчињата.....	79
Графикон 18. Bland-Altman дијаграм на тестот динамометрија на дланка десна рака кај девојчињата.....	80
Графикон 19. Bland-Altman дијаграм на тестот динамометрија на дланка на лева рака кај девојчињата.....	80
Графикон 20. Bland-Altman дијаграм на тестот динамометрија на дланка просечна вредност од десна и лева рака кај девојчињата.....	81
Графикон 21. Bland-Altman дијаграм на тестот подигнување труп 30 секунди кај девојчињата.....	81
Графикон 22. Bland-Altman дијаграм на тестот Чунесто трчање 4×10 m кај девојчињата.....	82
Графикон 23. Преглед на аритметичките средини на антропометриската мерка телесна висина помеѓу момчињата и девојчињата во првото мерење.....	94

Графикон 24. Преглед на аритметичките средини на антропометриската мерка телесна тежина помеѓу момчињата и девојчињата во првото мерење .....	94
Графикон 25. Преглед на аритметичките средини на антропометриската мерка индекс на телесна маса помеѓу момчињата и девојчињата во првото мерење.....	95
Графикон 26. Преглед на аритметичките средини на антропометриската мерка обем на половината помеѓу момчињата и девојчињата во првото мерење.....	95
Графикон 27. Преглед на аритметичките средини на моторичкиот тест претклон во сед помеѓу момчињата и девојчињата во првото мерење .....	96
Графикон 28. Преглед на аритметичките средини на моторичкиот тест скок во далечина помеѓу момчињата и девојчињата во првото мерење .....	96
Графикон 29. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на десна рака помеѓу момчињата и девојчињата во првото мерење .....	97
Графикон 30. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на лева рака помеѓу момчињата и девојчињата во првото мерење .....	97
Графикон 31. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка просечна вредност од десна и лева рака помеѓу момчињата и девојчињата во првото мерење .....	98
Графикон 32. Преглед на аритметичките средини на моторичкиот тест подигнување труп 30 s помеѓу момчињата и девојчињата во првото мерење .....	98
Графикон 33. Преглед на аритметичките средини на моторичкиот тест чунесто 4×10 m помеѓу момчињата и девојчињата во првото мерење .....	99
Графикон 34. Преглед на аритметичките средини на антропометриската мерка телесна висина помеѓу момчињата и девојчињата во второто мерење.....	102
Графикон 35. Преглед на аритметичките средини на антропометриската мерка телесна тежина помеѓу момчињата и девојчињата во второто мерење.....	103
Графикон 36. Преглед на аритметичките средини на антропометриската мерка индекс на телесна маса помеѓу момчињата и девојчињата во второто мерење .....	103
Графикон 37. Преглед на аритметичките средини на антропометриската мерка обем на половината помеѓу момчињата и девојчињата во второто мерење .....	104
Графикон 38. Преглед на аритметичките средини на моторичкиот тест претклон во сед помеѓу момчињата и девојчињата во второто мерење .....	104
Графикон 39. Преглед на аритметичките средини на моторичкиот тест скок во далечина помеѓу момчињата и девојчињата во второто мерење .....	105
Графикон 40. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на десна рака помеѓу момчињата и девојчињата во второто мерење.....	105
Графикон 41. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на лева рака помеѓу момчињата и девојчињата во второто мерење .....	106

Графикон 42. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка просечна вредност од десна и лева рака помеѓу момчињата и девојчињата во второто мерење .....	106
Графикон 43. Преглед на аритметичките средини на моторичкиот тест подигнување труп 30 s помеѓу момчињата и девојчињата во второто мерење .....	107
Графикон 44. Преглед на аритметичките средини на моторичкиот тест чунесто 4×10 m помеѓу момчињата и девојчињата во второто мерење .....	107

## 1. ВОВЕД

Физичкиот фитнес на децата и адолесцентите претставува суштински индикатор за нивното сегашно здравје и за ризикот од долгорочни хронични болести. Таа е мултидимензионална категорија, која вклучува компоненти како кардиореспираторна издржливост, мускулна сила и издржливост, флексибилност и состав на телото. Во теоријата и праксата се нагласува дека и моторичките способности (брзина, агилност, координација, рамнотежа) се составен дел од развојот, бидејќи овие способности ја овозможуваат секојдневната активност и го поддржуваат учењето и растот кај децата. Научното знаење јасно покажува дека ниските вредности на физичката кондиција кај младите корелираат со неповолни здравствени исходи кои предвидуваат повисок крвен притисок, нарушена липидна слика, инсулинска резистенција и други метаболички и кардиоваскуларни ризици. Всушност, физичката активност и кондицијата во детството се сметаат за многу значаен здравствен маркер, што укажува дека редовното, објективно и системско тестирање на овие параметри може да послужи како превентивен чекор за рано откривање на проблеми. Со други зборови, во идеална рамка, набљудувањето на физичкиот и моторичкиот развој ќе се овозможи навремено коригирање на животниот стил и училишните програми за поддршка на здравиот раст на младите.

Во идеална образовна и здравствена рамка, секоја држава би воспоставила систем за редовно следење на физичкиот и моторичкиот развој на учениците, со јасни и униформни референтни вредности. Учениците би поминувале стандардизирани фитнес-тестови на определени возрасни интервали, а резултатите би се споредувале со национални и меѓународни стандарди. Таквата пракса овозможува формирање на референтни криви и перцентили по возраст и пол, што им помага на наставниците и здравствените служби да го следат напредокот и потенцијалните ризици за здравјето.

Меѓутоа, во Северна Македонија сè уште не постои таков систематизиран мониторинг. Во тековната реалност се спроведуваат само поединечни истражувања или скромни проекти во рамки на одделни училишта, што нема национален обем. Испитувањата покажуваат дека македонските наставници обично не располагаат со дефинирани национални нормативи за интерпретација на резултатите, а самите тестирања често се изведуваат различно од училиште до училиште. Овие околности укажуваат дека сегашната состојба далеку се разликува од идеалот: без официјални референци и стандарди, процената на физичкиот развој останува субјективна и фрагментирана, и тешко се оценува вистинскиот тренд во кондицијата на учениците.

Имајќи предвид меѓународната практика, постојат бројни признаени фитнес батерии и научни напори кои овозможуваат методолошка подлога за проценка на развојот. Американската FITNESSGRAM® батерија и европската EUROFIT батерија се најшироко користени алатки за проценка на физичката кондиција кај децата. Овие комплексни протоколи вклучуваат множество тестови: на пример, FITNESSGRAM® користи PACER (20-метарски трчачки тест со прогресивно зголемување на брзината) и тестови за сила и флексибилност, додека EUROFIT содржи класични теренски тестови (на пр. скок во далечина, клекнувања, 20-метарско трчање, 'sit-and-reach' тест за флексибилност). Во рамките на европските иницијативи, HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) разви ALPHA батеријата како унифициран сет за проценка во Европа. ALPHA се состои од тестови што одговараат на здравствените компоненти (20-m шатл, динамометрија на дланка, скок во далечина од место, тест за флексибилност и чунесто трчање 4×10 m, меѓу другите) и е дизајнирана за споредливо следење во различни земји. Покрај ова, неколку национални проекти во светот исто така прилагодија слични комплексни батерии. На пример во Франција, Vanhelst et al. (2016) во рамките на програмата „MOVE” (со BOUGE тест-батерија) пријавија висока релијабилност на теренските фитнес-тестови што ги администрираа наставниците (ICC > 0.90 за сите компоненти). Слично, Ramírez-Vélez et al. (2015) ги тестираа компонентите на фитнес (морфолошки, мускулен, моторен, флексибилност, кардио) во училишна средина во Колумбија и заклучија дека батеријата на FUPRECOL е релијабилна: разликите меѓу повторувањата беа минимални, а систематскиот ефект речиси нула (освен кај тестот за флексибилност каде имаше мала негативна пристапна грешка). Овие резултати укажуваат дека, под адекватна обука на администраторите и стандардизиран протокол, теренските фитнес-тестови може да дадат стабилни и репродуцибилни мерења.

Сепак, јасно е дека и покрај успехот на овие интернационални напори, постојат значајни јазови во знаењето кога се работи за македонската средина. Во Северна Македонија до сега нема систематски анализирано релијабилноста на овие мерки. Единствениот обид за создавање национални референци доаѓа од МАКФИТ студијата (Гонтарев и сор., 2018), која опфати 9.241 ученик и прикажа возрастна и пол-зависни перцентили за фитнес до 14 години. Таа студија обезбеди важни статистички моменти, но не ја испитуваше релијабилноста на самите тестови во македонски услови.

Исто така, таа ги опфати само учениците од претпубертетна возраст до почеток на адолесценција (до 14 г.), но оставајќи ги тинејџерските години непокриени.

Дополнително, некои поранешни локални студии (на пр. норми за скок во далечина) беа ограничени на поединечни тестови и помали популации. Заради тоа, постои празнина во целосната слика: сегашните податоци во Северна Македонија се делумни и не ја решаваат дилемата за применливоста и точноста на мерните инструменти. Последица од овие недостатоци е тоа што во образовните и здравствените институции недостасуваат квалитетни податоци за разработка на политики. Без научно поткрепена релијабилност, резултатите од тестирањата можат да бидат неверливи: не е јасно дали забележаните промени се вистински или само случајни варијации во мерењето.

Од големо значење е и да се разгледаат веќе постоечките меѓународни батерии за фитнес и нивната улога. FITNESSGRAM® батеријата (САД) годишно се применува на милиони ученици; таа вклучува тестови како PACER (20-метар трчање со прогресивно зголемување на брзината) и стандарден сет за сила и флексибилност. EUROFIT батеријата (Европски Совет) опфаќа единаесет теренски тестови на пример, скок во далечина, тест за лежечки допир (sit-and-reach), склекови, и 20-метар трчање и ги користи за оценка на кардио-издржливост, мускулна сила и флексибилност. ALPHA батеријата, како дел од HELENA, ги обединува најрелевантните мерки со научни основи; таа го содржи 20-метарскиот шатл, ручдинамометрија на дланка, чунесто трчање 4×10 м, скок во далечина од место и тест за флексибилност. Сличен мулти компонентен пристап има и FUPRECOL батеријата (Колумбија), која комбинира морфолошки мерења со мускулни, моторни и кардио тестови, чија примена (под надзор на обучени наставници) покажа висока релијабилност. Секој од овие системи има за цел да ја отслика „health-related fitness“ на поединецот, така што училиштата и здравствените служби во земјите го следат целокупниот здравствен профил преку стандардизиран сет тестови и референтни скали.

МАКФИТ студијата е лоцирана како логичен и неопходен одговор на идентификуваните недостатоци. Оваа национална иницијатива ги интегрира меѓународно проверените методи со специфичноста на македонската популација. Таа ја користи EUROFIT методологијата, прилагодена на локалните возрасни групи и услови. Претходно, Гонтарев и сор. (2018) ги објавија националните перцентили за физичката кондиција кај македонските деца (6–14 г.), што претставува значителен придонес. Сепак, тие студии не беа посветени на тоа дали самите тестови даваат стабилни резултати. Затоа, ова истражување се фокусира на испитување на применливоста (реализација во училишни услови) и релијабилноста (точност при повторено тестирање) на МАКФИТ батеријата. За таа цел, физичко образованите наставници се обучуваат и оспособуваат

да ги изведат сите избрани тестови во две последователни изведби. Анализата ќе ја тестира доследноста на резултатите користејќи ги современите методи од психометријата: интра-класа корелации (ICC) за проценка на релативната релијабилност, тестови за систематски разлики во просечните резултати, како и индекси на апсолутна грешка (стандардна грешка на мерење – SEM, минимален детектичен прагот – MDC). Ова е во согласност со препораките на реномирани истражувања за релијабилност (на пример, Artero et al., 2011), кои советуваат комбинирање на различни метрики за целосна проценка на стабилноста. На тој начин, оваа студија ќе утврди дали разликите во мерењата ги надминуваат очекуваната варијација, односно дали МАКФИТ тестовите се доволно чувствителни и прецизни за да го отсликаат вистинскиот развој на македонските ученици.

Концептуалниот модел на ова истражување ги опфаќа сите клучни димензии на физичкиот развој. Таа ги интегрира морфолошките параметри (висина, тежина, БМИ, обем на струк) како маркери на растот и развојот, заедно со мускулно-функционалните перформанси (тестови за сила и издржливост) и моторичките вештини (брзина, координација, агилност). МАКФИТ батеријата вклучува тестови за секоја од овие категории: на пример, чунесто трчање 4×10 м, подигнување на трупот за 30 секунди, скок во далечина од место, динамометрија на дланка и длабок претклон во сед. Овие тестови се усогласени со концептот на „health-related fitness”. Во аналитичкиот фокус е процената на релијабилноста на овие мерки: студијата предвидува повторно тестирање на истите ученици под исти услови, со цел да се споредат резултатите и да се утврди нивната повторливост. Ова технички се мери преку ICC за релативна конзистентност и преку анализи на апсолутната грешка (SEM, MDC). Како што објаснуваат Artero et al., (2011), релијабилноста на тестот се дефинира како повторливост на вредностите при двојно тестирање на истиот испитаник.

Следствено, оценката ќе покаже дали измерените разлики се надвор од очекуваната случајна варијација. Овој мулти-индикаторски пристап (релативни и апсолутни метрики) овозможува сеопфатна проценка: ако МАКФИТ тестовите ја исполнат задачата, тие ќе претстават доверлив инструмент за следење на физичкиот и моторичкиот развој кај македонските ученици и за унапредување на националните програми за физичко образование.

## 2. ДОСЕГАШНИ ИСТРАЖУВАЊА

Rikli et al. (1992) ја истражувале тест-ретест релијабилноста на тестовите за проценување на кардиореспираторниот фитнес, трчање на 1 една милја и трчање на пола милја кај деца на возраст од 5 до 9 години. Истражувањето било реализирано на примерок од 177 момчиња и 185 девојчиња. Резултатите од истражувањето покажале дека тестот трчање на 1 милја имал прифатлива релијабилноста за деца на возраст од 8-9 години ( $0,84 < R < 0,90$ ), но немал задоволителна релијабилност за децата од 5-6 години ( $0,39 < R < 0,56$ ). Тестот трчање на пола милја ги исполнувал стандардите за минимална релијабилност за момчиња и девојчиња на возраст од 5-6 години ( $0,77 < R < 0,77$ ).

Pitetti et al. (2002) ја истражувале релијабилноста на тестот трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT) кај деца и адолесценти на возраст од 8–15 години. Резултатите од истражувањето покажале висока тест-ретест релијабилност на тестот трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT) ( $R = 0.89$ ).

Liu et al. (1992) ја истражувале релијабилноста на тестот трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT) кај адолесценти на возраст од 12–15 години. Врз основа на добиените резултати авторите констатираат дека тестот трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT) е релијабилен тест ( $ICC = 0.93$ ).

Beets & Pitetti (2006) ја истражувале тест-ретест релијабилност на тестовите трчање-одење на една милја и трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT) кај адолесценти на возраст од 13–18. Истражувањето било реализирано на примерок од 165 момчиња и 76 девојчиња. Резултатите покажале дека и двата теста имаат слична релијабилност. Според тоа, ниту еден тест не може да се смета дека обезбедува постабилна дијагностичка повратна информација за учениците во однос на другиот.

Tong et al. (2001) ја истражувале релијабилноста на тестот за проценување на кардиореспираторниот фитнес, трчање 5 минути кај адолесценти со просечна возраст од  $17.0 \pm 0.2$  години. Истражувањето било реализирано на примерок од 14 момчиња и 31 девојче. Испитаниците го повториле тестот 4 пати во временски интервал од 3 дена. Резултатите од истражувањето покажала дека релијабилноста меѓу 3 и 4 мерење била повисока во однос на 1 и 2 мерење. Авторите препорачуваат, за постигнување на

најдобри перформанси во овој тест, испитаниците двапати да го пробаат тестот со цел да се запознаат со истиот.

Ruiz et al. (2006) ја истражувале релијабилноста на тестот динамометрија на дланка кај адолесценти на возраст од 13–18 години. Истражувањето било реализирано на примерок од 13 момчиња и 4 девојчиња. Резултатите од истражувањето покажале високи коефициенти на релијабилност ( $ICC=0.98$  и  $0.96$ ) и не била утврдени статистички значајни разлики меѓу првото и второто мерење (тест-ретест мерењето).

España - Romero et al. (2008) ја истражувале релијабилноста на тестот динамометрија на дланка кај деца и адолесценти на возраст 6 до 12 години. Истражувањето било реализирано на 17 момчиња и 5 девојчиња. Тестот бил мерен два пати во временски интервал од 2 часа. Резултатите од тестот покажале високи коефициенти на релијабилност, статистички значајна корелација и не биле утврдени статистички значајни разлики помеѓу резултатите од првото и второто мерење.

Molenaar et al. (2008) ја истражувале релијабилноста на два различни рачни динамометри. Истражувањето било реализирано на примерок од 104 испитаника (45 момчиња и 59 девојчиња) на возраст од 4 до 12 години. Резултатите од истражувањето покажале дека и динамометарот од марката Lode и вигориметарот Martin се релијабилни инструменти со кои може да се проценува силината на стисокот на дланката кај деца под 12-годишна возраст. Сепак динамометарот од марката Lode покажал повисоки коефициенти на релијабилност меѓу првото и второто мерење во однос на вигориметарот Martin. Понатаму, споредбата на најмалите забележливи разлики укажуваат дека динамометарот Lode е попрецизен инструмент.

Patterson et al. (2001) ја истражувале тест-ретест релијабилноста на тестот подигнување на трупот (curl-up) кај деца на возраст од 10 до 12 години. Истражувањето било реализирано на примерок од 36 момчиња и 48 девојчиња. Временскиот интервал меѓу првото и второто мерење се движело од 48–72 часа. Резултатите од истражувањето покажале дека објективност на тестот изнесувале  $.99$  ( $R = 0,99$ ) и немало статистички значајни разлики меѓу оценувачите во првото мерење, додека во второто мерење биле утврдени статистички значајни разлики меѓу оценувачите. Тест-ретест релијабилноста пријавено од наставникот се движела  $R$  (95 % CI) =  $0,89$  ( $0,78-0,94$ ) за момчиња и  $0,86$  ( $0,74-0,92$ ) за девојчиња.  $R$  (95 % CI) проценет за едно испитување =  $0,80$  ( $0,64-0,89$ ) за момчиња и  $0,75$  ( $0,59-0,85$ ) за девојчињата. Не биле утврдени статистички значајни разлики во текот на 2. дена (помеѓу првото и второто мерење). Процентот на усогласеност изнесувал  $0,89$ , а модифицираниот карра коефициент се движел од  $0,78$  за

момчиња и 0,81 и 0,62 за девојчињата. Тест-ретест релијабилноста пријавено од учениците се движела  $R$  (95 % CI) = 0,82 (0,66-0,91) за момчиња и 0,81 (0,67-0,90) за девојчиња.  $R$  (95 % CI) проценет за едно испитување = 0,70 (0,49-0,84) за момчиња и 0,69 (0,50-0,81) за девојчиња. Биле утврдени статистички значајни разлики меѓу првото и второто мерење кај девојчињата. Процентот на согласување изнесувал 0,86, а модифицираниот карра коефициент се движел од 0,72 за момчиња и 0,79 и 0,58 за девојчињата. Корелацијата помеѓу резултатите пријавени од наставникот и ученикот изнесувал  $r$  (95 % CI) = 0,54 (0,46-0,61). Анализата на АНОВА покажа дека резултатите пријавени од децата биле значително повисоки од резултатите пријавени од наставникот.

Romain & Mahar (2001) ја истражувале релијабилноста на тестовите склекови до  $90^{\circ}$  и модифицирани згибови. Истражувањето било реализирано на примерок од 62 испитаника (30 момчиња и 32 девојчиња) со просечна возраст од  $11.4 \pm 0.9$  години. Врз основа на добиените резултати авторите констатираат дека и двата теста покажале задоволителна релијабилност.

Pritchard & O'Bryant (2001) ја истражувале релијабилноста на тестот подигнување до сед (half sit-up). Истражувањето било реализирано на примерок од 72 испитаника (40 момчиња и 32 девојчиња) со просечна возраст од  $12.7 \pm 0.1$  години. Авторите констатирале дека тестот подигнување до сед (half sit-up) е релијабилен и може да се користи како тест за проценување на силовата издржливост на абдоминалната мускулатура во наставата по физичко образование.

Hannibal et al. (2006) ја истражувале релијабилноста на тестот екстензија на трупот на кутија до  $90^{\circ}$ . Истражувањето било реализирано на примерок од 43 испитаника на возраст од 14–18 години. Резултатите од истражувањето покажале статистички значајни разлики меѓу момчињата и девојчињата, но не биле утврдени статистички значајни разлики во тестот меѓу првото и второто мерење. Кога се проценува мускулната сила и издржливоста со овој тест, истиот треба да се повтори еднаш, но претходно испитаниците треба да бидат запознаени со истиот.

Alricsson et al. (2001) ја истражувале релијабилноста на тестовите за проценување на моторичкиот фитнес, слалом тест и тест со совладување на препреки кај примерок ја деца на возраст од 11 години. Истражувањето било реализирано на примерок од 8 момчиња и 3 девојчиња. Не биле утврдени статистички значајни разлики помеѓу интервалите на тестирање ниту за тестот слалом ( $P = 0,99$ ) или за тестот совладување препреки ( $P = 0,96$ );  $R$  (95 % CI) = 0,96 (0,89-0,99) и 0,90 (0,75-0,97). Просечното CV од

4-те тест интервали за сите учесници во тестовите за слалом и совладување препреки изнесувала 2,3% и 4,9%, соодветно. Тестот за слалом и тестот совладување препреки се релијабилни и функционални тестови за проценување на брзина и агилност во групи млади атлетски поединци.

Ortega et al. (2008) ја истражувале релијабилноста и останатите мерни карактеристики на тестовите трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT), скок во далечина од место, динамометрија на дланка, издржај во згиб, чунесто трчање 4 x 10 метри и модифицираната верзија на тестот длабок претклон во сед кај адолесценти со просечна возраст од  $13.6 \pm 0.8$ . Истражувањето било реализирано на примерок 123 испитаника (69 момчиња и 54 девојчиња). Авторите констатираат дека не биле утврдена систематска пристрасност и полови разлики за ниту еден од испитуваните тестови, што укажува на прифатлива релијабилност на истите.

Gutin et al. (1996) ја истражувале релијабилноста на пресметката на процентот на масно ткиво преку антропометриски мерења кај деца на возраст од 9 до 11 години. Резултатите од истражувањето покажале дека нема значителни разлики помеѓу испитувањата; ICC = 0,99; минимална разлика = - 1,35 %, максимална разлика = 2,82 %; средна разлика = 0,13 %; Границата на прифатливост = - 1,74 % и 1,99 %.

Esraña-Romero et al. (2010) реализирале истражување со цел да ја утврдат релијабилноста, применливоста и безбедноста на фитнес батерија тестови кои треба да ги применуваат наставниците по физичко образование во наставата по физичко образование. Шест наставници по физичко образование од три основни училишта и три средни училишта, двапати (во интервал од 7 дена) ги примениле тестовите: трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT), скок во далечина од место, динамометрија на дланка; и ја измериле тежината, висината, кожните дупли на трицепсот и супскапуларно и обемот на половината кај 58 деца. (возраст: 6 – 11 год.) и 80 адолесценти (возраст: 12 – 18 год.). Применливоста и безбедноста на тестовите била проценета со методата на директно набљудување. Резултатите од истражувањето покажале разлики меѓу првото и второто мерење во тестовите скок во далечина од место ( $3.8 \pm 12.7$  cm,  $P < 0.05$ ) и телесната висина ( $0.73 \pm 0.8$  cm,  $P < 0.001$ ) кај децата, како и обемот на половината и кај децата и адолесцентите ( $-0.82 \pm 1.2$  cm and  $-0.35 \pm 0.8$  cm  $P = 0.001$ ). Безбедноста и изводливоста на тестовите била на задоволително ниво. Врз основа на добиените резултати, авторите заклучиле дека фитнес тестовите поврзани со здравјето применети во ова истражување се релијабилни, применливи и безбедни за примена во наставата по физичко образование.

Lubans et al. (2011) ја истражувале тест-ретест релијабилноста на фитнес батерија тестови за проценување на здравствениот фитнес. Истражувањето било реализирано на примерок од 68 испитаника (42 момчиња и 26 девојчиња) со просечна возраст  $14.8 \pm 0.4$  години. Биле реализирани две мерења во временски интервал од една недела. За реализирање на целите на истражувањето биле применети следниве тестови и мерки: анализа на биоелектрична импеданса (БИА) за пресметување на процентот на маснотии во телото, динамометрија на нога, склекови до  $90^{\circ}$ , 7 степен тест за проценување на силова издржливост на абдоминалната мускулатура, издражај во получучањ на сид. Пресметана е интер-класна корелација (ICC), t-тестови за зависни примероци и коефициент на варијацијата. Коефициентот на интер-класна корелација (ICC), за процентот на телесни маснотии беше сличен и кај момчињата (ICC = 0,95) и девојчињата (ICC = 0,93), но средниот коефициент на варијација беше значително повисок кај момчињата отколку кај девојчињата (22,2% наспроти 12,2%). Коефициентите на варијацијата кај момчињата на тестовите за проценување на мускулниот фитнес се движеа од 9,0% за тестот динамометар на ногата до 26,5% за тестот издражај во получучањ на сид. Коефициентите на варијација кај девојчињата се движеа од 17,1% за тестот за проценување на силовата издржливост на абдоминалната мускулатура до 21,4% за тестот склекови. Иако, биоелектричната импеданса покажа висока релијабилност при проценка на телесните маснотии, тестовите за проценување на мускулниот фитнес покажале висока систематска грешка, што укажува дека истите може да бараат опсежна фаза на запознавање пред резултатите да се сметаат за релијабилни.

Ramírez-Vélez et al. (2015) реализирале истражување со цел да ја утврдат релијабилноста на тестовите за проценување на физичкиот фитнес поврзан со здравјето кај колумбиски ученици. Истражувањето било реализирано на примерок од 229 колумбиски деца и адолесценти (момчиња  $n = 124$  и девојчиња  $n = 105$ ) на возраст од 9 до 17,9 години. За реализирање на целите на истражувањето биле применети пет компоненти на физичкиот фитнес поврзан со здравјето 1) морфолошка компонента, висина, тежина на телесна маса, индекс на телесна маса, обем на половината, кожни дупли на трицепсот и супскапуларно, процент на масно ткиво со методата на биоелектрична импеданса 2) мускулно-скелетна компонента: динамометрија на дланка и скок во далечина од место. 3) моторната компонента: тестот за брзина/агилност (чунесто трчање  $4 \times 10$  m); 4) компонента на флексибилност, длабок претклон во сед 5) кардиореспираторна компонента: трчање со прогресивно зголемување на брзината на 20

метри (20m-SRT), врз основа на кој е проценета и максималната потрошувачка на кислород. Тестовите се повторуваат по два пати, со временски интервал од 1 недела во ист ден во неделата. Врз основа на добиените резултати авторите заклучуваат дека батеријата за физички фитнес поврзана со здравјето „Fuprecol study“, администрирана од наставници по физичко образование, била релијабилна за мерење на компонентите на физичкиот фитнес поврзани со здравјето кај децата и адолесцентите на возраст од 9-17,9 години во училишна средина во Колумбија.

Резултатите од досегашните истражувања укажуваат дека тестот трчање со прогресивно зголемување на брзината на 20 метри (20m-SRT) е релијабилен тест за мерење на кардиореспираторниот фитнес, тестот динамометрија на дланка и скок во далечина од место се релијабилни тестови за проценување на мускулно-скелетниот фитнес. Тестот чунесто трчање  $4 \times 10$  m е релијабилен тест за проценување на моторниот фитнес, а висината, тежината, БМИ, кожните набори, обемот на половината и процентот на телесни масти проценети од дебелината на кожните наборите се релијабилни тестови за мерење на составот на телото. Тестот длабок претклон во сед се покажал како релијабилен за проценување на фитнес компонента на флексибилност. Неопходно е да се истакне дека сите овие тестови се покажаа како релијабилни, но доказите за некои од нив се ограничени поради малиот број студии.

### **3. ПРОБЛЕМ, ПРЕДМЕТ, ЦЕЛИ И ХИПОТЕЗИ НА ИСТРАЖУВАЊЕТО**

#### **3.1. ПРОБЛЕМ, ПРЕДМЕТ И ЦЕЛИ НА ИСТРАЖУВАЊЕТО**

Проблемот на истражувањето произлегува од суштинската потреба систематски, објективно и доверливо да се процени физичкиот и моторичкиот развој кај децата од училишна возраст, во реални училишни услови. Во контекст на современиот образовен систем и алармантните трендови на намалена физичка активност, наставниците по физичко и здравствено образование се соочуваат со барања не само да реализираат наставна програма, туку и да ја следат, документираат и интерпретираат физички фитнес и развој на секое дете. За таа цел им е потребен валидиран, стандардизиран и практично применлив сет мерни инструменти, кој во исто време е научно заснован и изводлив во рамките на ограничувањата на училишната средина. Во македонскиот контекст, ваква батерија на тестови специјално адаптирана за деца и адолесценти и систематски проверена од аспект на релијабилност, објективност, безбедност и применливост – досега не е доволно разработена и емпириски потврдена.

Оттука, централниот проблем на истражувањето е да се утврди до кој степен мерките и тестовите вклучени во батеријата МАКФИТ обезбедуваат стабилни, репродуктивни и објективни резултати кога се применуваат во училишни услови од страна на наставници, а не само во контролирана лабораториска средина. Ова подразбира оценување на нивната тест ретест релијабилност, односно дали резултатите остануваат конзистентни при повторени мерења во краток временски интервал, како и процена на степенот на објективност, односно независност од субјективниот фактор на мерителот. Дополнително, проблемот се надградува со прашањето за безбедност: дали тестовите можат да се спроведуваат без ризик по здравјето и благосостојбата на децата, во рамки на нивниот развоен капацитет и типичните услови во спортските сали и училишните дворови.

Особено значаен аспект на проблемот е и практичната применливост на МАКФИТ во секојдневната педагошка практика. Не е доволно тестовите да бидат статистички релијабилни доколку нивната реализација е логистички сложена, преголемо временско оптоварување за наставникот, бара специфична опрема која не е достапна во училиштата, или пак, резултатите се тешко интерпретабилни од страна на практичарите. Затоа, истражувањето се стреми да даде одговор дали оваа батерија тестови, во реални

услови на настава, може да се користи како ефективна алатка за рутински скрининг, мониторинг и педагошка дијагностика на физичкиот и моторичкиот развој кај адолесцентите. На тој начин, проблемот на истражувањето не е само техничко прашање на релијабилност, туку и пошироко прашање за тоа дали МАКФИТ може да стане функционална врска или „мост“ помеѓу научните знаења и секојдневната наставна пракса, со директни импликации за планирање, индивидуализација и евалуација на физичкото образование.

Предмет на истражување се антропометриските мерки и моторичките (фитнес) тестови вклучени во батеријата МАКФИТ, администрирани кај адолесценти од средношколска возраст. Посебен фокус е ставен на нивото и распределбата на овие показатели, како и на нивните метрички својства, пред сè релијабилност, објективност и практична применливост во училишни услови. Преку систематска анализа на резултатите од овие мерни инструменти, истражувањето го опфаќа и начинот на кој антропометричкиот статус и моторичките способности на адолесцентите можат да се следат, споредуваат и интерпретираат во функција на педагошка дијагностика, планирање и евалуација на програмите по физичко и здравствено образование.

Врз основа на проблемот и предметот на истражувањето поставени се и повеќе конкретни цели и тоа:

1. Да се утврди тест-ретест релијабилноста (релативна релијабилност) на антропометриските мерки и моторичките (фитнес) тестови кои се составен дел на батеријата МАКФИТ кај учениците од училишна возраст, одделно кај момчињата и кај девојчињата.
2. Да се определи апсолутната релијабилност на поединечните антропометриски мерки и моторички тестови (SEM, SEM%, MDC, CV%) кај момчињата и девојчињата, со цел да се утврдат праговите на минимална детектибилна промена.
3. Да се провери постоењето на систематска разлика меѓу првото и второто мерење (T1–T2) за секоја антропометриска мерка и моторички тест, преку тестови на разлики и големина на ефект, со цел да се идентификува евентуален ретест-ефект (ефект на запознавање со тестот).
4. Да се процени согласноста меѓу првото и второто мерење на индивидуално ниво преку Bland–Altman анализа (bias, граници на согласност и хетероскедастичност) за поединечните тестови, со цел да се утврди дали резултатот од ретестот може практично да го замени резултатот од првото мерење при следење на истото дете,

без притоа да се промени интерпретацијата за неговиот вистински статус и евентуалната промена.

5. Да се утврдат половите разлики во нивото на антропометриските карактеристики и моторичките способности во првото мерење, со контрола на возраста и да се процени големината на ефектот на полот врз поединечните показатели.

### 3.2. ХИПОТЕЗИ НА ИСТРАЖУВАЊЕТО

Имајќи ги предвид резултатите од досегашните истражувања, проблемот, предметот и целите на истражувањето, поставени се следниве хипотези:

X<sub>1</sub>. Антропометриските мерки и моторичките (фитнес) тестови кои се составен дел на батеријата МАКФИТ ќе покажат добра до одлична тест–ретест релијабилност, така што интракласните корелациски коефициенти ICC(2,1) за повеќето варијабли ќе бидат во домен на висока релијабилност, одделно кај момчињата и кај девојчињата.

X<sub>2</sub>. Апсолутната релијабилност на поединечните антропометриски мерки и моторички тестови ќе биде во прифатливи граници и ќе овозможи детекција на реални индивидуални промени, така што кај повеќето тестови ќе се добијат релативно ниски вредности на SEM% и CV%, а минимално детектибилната промена (MDC) ќе биде доволно мала во однос на варијабилноста на тестот за да има практично значење при следење на поединци.

X<sub>3</sub>. Ќе бидат утврдени статистички значајни систематски разлики меѓу првото и второто мерење (T1–T2) кај дел од моторичките тестови, кои ќе се манифестираат преку значајни вредности на t-тестот и мала до умерена/голема големина на ефект на влијание (dz), додека кај основните антропометриски мерки не се очекуваат статистички значајни систематски разлики.

X<sub>4</sub>. Согласноста меѓу првото и второто мерење на индивидуално ниво, проценета со Bland–Altman анализа, ќе биде задоволителна кај повеќето антропометриски мерки и моторички тестови, така што просечниот bias нема да се разликува значајно од нула, најголем дел од индивидуалните разлики T1–T2 ќе се наоѓаат во прифатливи 95% граници на согласност, а нема да се евидентира изразена хетероскедастичност. Ова ќе укаже дека резултатот од ретестот може практично да го замени резултатот од првото мерење при следење на истото дете, без да се промени интерпретацијата за неговиот вистински статус.

H<sub>5</sub>. Ќе бидат утврдени статистички значаен мултиваријантен ефект на полот врз збирната варијација на антропометриските карактеристики и моторичките способности во првото и второто (ретест) мерење , при што кај дел од поединечните варијабли ќе се евидентираат значајни униваријатни полови разлики со умерена до голема големина на ефект на влијание.

## 4. МЕТОДИ НА ИСТРАЖУВАЊЕТО

### 4.1. ПРИМЕРОК НА ИСПИТАНИЦИ

Истражувањето е спроведено на примерок од вкупно 102 адолесценти, ученици од средношколска возраст, избрани по случаен избор од неколку средни училишта на територија на Скопскиот регион. Примерокот е структуриран според полот и опфаќа 62 испитаници од машки и 40 испитаници од женски пол, со што е обезбедена реална основа за споредбена анализа на антропометриските карактеристики и моторичките способности меѓу половите. Ваквата структура на примерокот овозможува да се добие релевантен пресек на физичкиот и моторичкиот статус на средношколската популација во урбана средина, кој е методолошки соодветен за поставените цели на истражувањето.

Во истражувањето беа вклучени само ученици за кои родителите или законските старатели претходно дадоа писмена информирана согласност, а самите испитаници доброволно се согласија да учествуваат во мерењата и тестирањата. Дополнителен критериум за вклучување беше учениците да бидат психофизички здрави, без дијагностицирани хронични заболувања, акутни повреди или функционални ограничувања кои би можеле да влијаат на изведбата на моторичките тестови или да го загорзат нивното здравје за време на процената. Исто така, како предуслов се земаше редовното посетување на наставата по физичко и здравствено образование, со цел да се намали веројатноста во примерокот да влезат ученици со исклучително ниско ниво на физичка активност кои би можеле непропорционално да ја искриват сликата за популацијата.

Од понатамошната статистичка анализа беа исклучени сите ученици кои, од било која причина, не ги комплетираа сите предвидени антропометриски мерења и моторички тестови, како и оние кај кои беше евидентирана очигледно нецелосна или несоодветна изведба на тестовите што можеше да ја наруши веродостојноста на податоците. На тој начин, финалниот аналитички примерок опфаќа адолесценти кај кои е обезбедена целосна, конзистентна и методолошки квалитетна база на податоци, што претставува неопходен предуслов за валидна процена на релијабилноста и за прецизна интерпретација на добиените резултати.

Истражувањето е спроведено во согласност со основните етички начела пропишани во Хелсиншката декларација за етички принципи во медицинските истражувања со луѓе, како и со важечките национални и институционални прописи за заштита на учесниците во истражувања. Протоколот на студијата беше разгледан и одобрен од надлежна етичка комисија на универзитетската установа во која се реализираше истражувањето, при што беа особено почитувани принципите на автономија, добротворност и правичност.

При спроведување на тестирањата беше обезбедена целосна доверливост и анонимност на податоците: сите резултати беа кодирани, без можност поединечниот ученик да биде идентификуван во базата на податоци или во објавените резултати. Учесниците беа јасно информирани дека учеството е доброволно и дека во секој момент можат да се повлечат од истражувањето без никакви негативни последици по нивниот наставно образовен статус. Мерењата и тестирањата беа изведувани од стручно лице во контролирани услови, со внимателна процена на ризикот и придобивките, при што беше водено сметка протоколите да бидат безбедни, соодветни за возраста и физичкиот развој на учениците и да не предизвикуваат прекумерен напор или непријатност. На овој начин, примерокот на испитаници не само што е методолошки соодветен, туку и етички оправдан, што ја засилува валидноста и применливоста на добиените наоди.

## 4.2. ПРИМЕРОК НА ВАРИЈАБЛИ

Во истражувањето беа применети вкупно 11 варијабли, од кои 4 варијабли за проценка на антропометриските карактеристики и 7 варијабли за проценка на моторичките способности.

### 4.2.1. Варијабли за проценка на антропометриските карактеристики

1. Висина на телото.
2. Тежина на телото.
3. Обем на половината.
4. Индекс на телесна маса (Weight & Height - BMI).

#### 4.2.2. Варијабли за проценка на моторичките способности

1. Претклон во сед.
2. Динамометрија на дланка лева рака.
3. Динамометрија на дланка десна рака.
4. Динамометрија на дланка просек од десна и лева рака.
5. Скок во далечина од место.
6. Лежење - сед за 30 секунди.
7. Чуњесто трчање 4 x 10 метри (4 x 10 m shuttle run test).

#### 4.3. ОПИС НА АНТРОПОМЕТРИСКИТЕ МЕРЕЊА

##### **Висина на телото**

**Инструменти:** Антропометар по Мартин.

**Задача:** Висината на телото е измерена со висинометар (антропометар по Мартин). При мерењето испитаникот задолжително е бос во исправен став, при што стои на тврда хоризонтална подлога под висинометарот. Главата на испитаникот се наоѓа во таква положба при што франкфуртската рамнина е во хоризонтална положба. Испитаникот го исправа грбот и ги составува стапалата. Испитувачот стои лево од испитаникот при што го контролира висинометарот дали е поставен вертикално спуштајќи го лизгачот на темето на испитаникот.

**Оценување:** Резултатот се чита со точност од 0,1 cm.

**Напомена:** Потребен е еден мерач и еден записничар.

##### **Тежина на телото**

**Инструмент:** Медицинска децимална вага.

**Задача:** Тежината на телото е мерена со медицинска вага поставена на тврда хоризонтална подлога. Испитаникот, кој е бос и во спортска опрема, застанува на средина на вагата и мирно стои во исправена положба.

**Оценување:** Во моментот кога стрелката на вагата целосно ќе биде смирена, испитувачот го чита и внесува резултатот со точност од 0.5 kg.

**Напомена:** Потребен е еден мерач кој истовремено е и записничар.

### **Обем на половината**

**Инструмент:** Нееластична лента.

Детето е минимално облечено за да може лентата правилно да се позиционира. Детето е во исправен став, со опуштен мев, рацете се од страна, а стапалата се споени. Тестерот е свртен кон детето и ја става нееластичната лента во хоризонтална рамнина, на ниво на најтесниот дел на струкот, кога се гледа трупот од предната страна. Кај некои подебели деца тешко е да се идентификува најтесниот дел на струкот. Во таков случај најмалиот хоризонтален обем треба да се мери во подрачјето меѓу спина илијака супериор и работ на долното ребро (midaxillary line).

**Број на повторувања:** Две мерења се изведуваат последователно и средната вредност се користи во анализите. Мерењето започнува кога детето ќе ја заземе правилната почетна позиција. Мерењето не треба да се прави над облеката, треба да се земе вредноста на крајот од нормалниот истегнување без лентата да прави компресија на кожата и кога рацете на детето ќе бидат поставени од страна на трупот.

**Оценување:** Се евидентира најблискиот резултат до 0,1 cm. Пример: резултат од 60,7 cm се запишува како 60,7.

**Напомена:** Потребен е еден мерач кој истовремено е и записничар.

### **Индекс на телесна маса – БМИ**

Индекс на телесна маса (анг. *body mass index* – BMI) се пресметува преку математичка формула која го пресметува односот на висината и масата на поединецот или попрецизно претставува масата на телото (во кг) поделена со висината (изразена во метри на квадрат). Односно,  $BMI = kg/m^2$

#### **4.4. УСЛОВИ И ТЕХНИКА НА МЕРЕЊЕ НА АНТРОПОМЕТРИСКИТЕ КАРАКТЕРИСТИКИ**

Мерењето е реализирано во стандардни училишни услови на редовните часови по физичко и здравствено образование. Мерењето го реализираат стручни лица од областа на кинезиологијата, кои претходно ќе бидат оспособени за мерење на определена антропометриска мерка. Мерењата е спроведени секој работен ден во време од 7:30 до 14:30 часот (за време на редовната претпладневна настава).

1. Инструментите беа стандардни и соодветно баждарени пред нивната употреба;
2. Просториите во кои се вршеа мерењата се чисти, доволно топли и соодветно осветлени;
3. При мерењето испитаниците беа боси и минимално облечени во спортска опрема;
4. При реализација на мерењата обележени се и одредени антропометриски точки и нивоа.

Сите антропометриски мерки се мерени по методот на Интернационалната биолошка програма (Lohman, Roche & Martorell, 1988). Инструментите за мерење се со стандардна изработка. За мерењата се обезбедени следните инструменти:

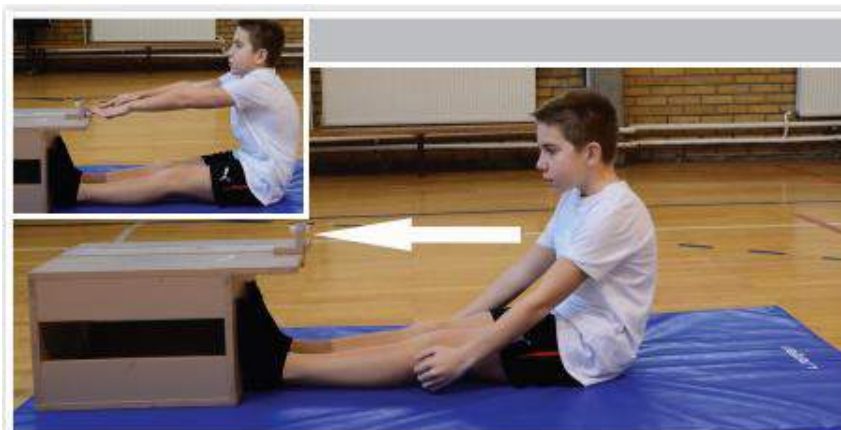
- Висинометар со точност на мерење од 0.1 cm.
- Електронска вага (транспортабилна).
- Други технички и мерни инструменти потребни за реализација на мерењата.

#### 4.5. ОПИС НА ВАРИЈАБЛИ ЗА ПРОЦЕНКА НА МОТОРИЧКИТЕ СПОСОБНОСТИ

##### **Претклон во сед**

**Опрема и реквизити потребни за изведување на тестот:** дрвен сандак (кутија) за тестирање со должина 45 cm, ширина 35 cm и висина 32 cm. Мерките на горната плоча се: 60 cm должина и 35 cm ширина, оваа горна плоча преоѓа 15 cm во страната на сандакот на кој испитаникот ги потпира нозете, скалата е во распон од 0 до 60 cm и е означена на средината на горната плоча, со лизгачи линијар кој се наоѓа на скалата на која испитаникот ја турка со рацете.

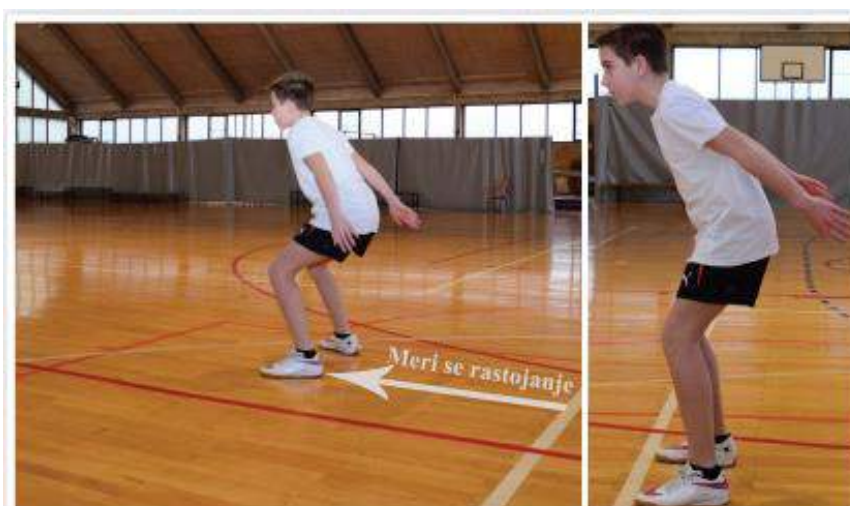
**Опис на изведување на тестот:** Ученикот седи бос пред сандакот, со испружени нозе и стапала поставени на предната страна од сандакот. Наставникот треба да клекне пред ученикот и со притисок на колената да му оневозможи да ги свитка нозете. Со испружени раце (една преку друга) пред себе, ученикот, без да ги свитка нозете, се наведнува напред колку што може повеќе и рамномерно со рацете го турка лизгачкиот линијар полека и без замав. Резултатот е одреден како најоддалечена позиција што ученикот ќе ја достигне со врвовите на своите прсти. Тестот се повторува двапати и се запишува подобриот резултат (изразен во сантиметри).



### **Скок во далечина од место**

**Опрема и реквизити потребни за изведување на тестот:** мерна лента со точност од 1 cm, рамна нелизгачка површина со обележено скокалиште на кое местото на одразот е на исто ниво како и местото на доскокот.

**Опис на изведување на тестот:** Отскокот и доскокот се задолжително суножни. Ученикот скока во опрема за настава по физичко и здравствено образование. Задачата на ученикот е со суножен отскок да скокне што подалеку. Како точна мерка се земе точката на допир на петицата со површината, која е најблиску до линијата на отскокот. Се изведуваат два скока, неправилно изведениот скок се повторува. Се забележува најдалеку изведениот скок. Резултатот се прикажува во сантиметри.



### **Подигнување на трупот за 30 сек.**

**Опрема и реквизити потребни за изведување на тестот:** душеци и стоперица.

**Опис на изведување на тестот:** Ученикот лежи на грб со рацете поставени на тилот и нозете свиткани во колената под агол од 90 степени, фиксирани од страна на друг ученик (помошник во реализирање на тестот). Од лежечка положба, ученикот треба да направи максимален број свиткување напред, на тој начин што секојпат со лактите ќе ги допре колената.

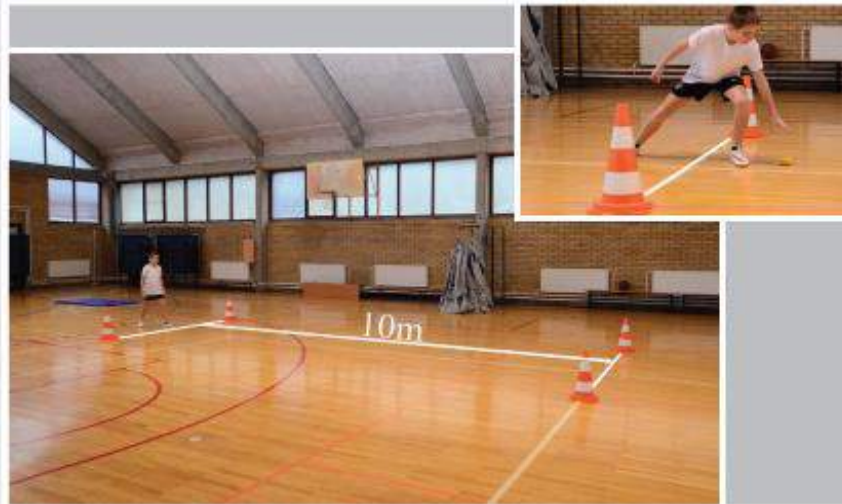


### **Чунено трчање 4 x 10 метри**

**Опрема и реквизити потребни за изведување на тестот:** обележена патека за трчање на чист и нелизгав под, два сунѓери и стоперица. Патека се бележи така што на растојание од 10 м се цртаат или бележат на подот две паралелни линии со должина од 1,2 м (со креда или леплива лента во боја). На другата обележана паралелна линија, на крајот од патеката се поставуваат двата сунѓери.

**Опис на изведување на тестот:** Ученикот од стартната линија на даден знак трча што побргу може по означената патека до линијата на другиот крај од патеката која мора да ја пројде со стапалото на едната нога и да го земе првиот сунѓер, потоа право вртење и се враќа назад во истиот правец, потоа застанува со стапалото на едната нога преку стартната линија, оставајќи го сунѓерот и вртејќи се трча назад до другата линија каде што го зема и другиот сунѓер, прави вртење и го завршува тестот со спринт преку стартната линија (за тоа време претрчува 40 метри). Мерењето на времето започнува на дадениот знак (кога испитаникот почнува да трча), а завршува кога ученикот ќе помине преку стартната линија со цело стапало, после четири пати претрчана делница од 10 м.

Делниците треба да се трчаат со максимална брзина. За грешка се смета ако ученикот не прејде со стапалото преку линијата или ако не трча по патеката и во тој случај мерењето се повторува.



### **Динамометрија на дланка**

**Опрема и реквизити потребни за изведување на тестот:** калибриран рачен динамометар со нагодлива рачка.

**Опис на изведување на тестот:** Максимално силен еднократен стисок со дланката, односно динамометарот и регистрирање на статичката сила во килограми. Ученикот постепено и без прекин ја стиска рачката на динамометарот најмалку две секунди. Тестот се повторува двапати со десната и двапати со левата рака. Оценка е просечната вредност од двата подобри резултати од стисокот на левата и десната рака.

**Напомена:** во текот на тестирањето раката и дланката со која го држите динамометарот не смее да го допира телото, инструментот се држи во линија со подлактицата од страна на телото; нагодете ја рачката така што двете шипки ќе одговараат на големината на првата фаланга на средниот прст. Резултатот се прикажува во килограми (со точност од 1 кг).



#### 4.6. УСЛОВИ И ТЕХНИКА НА МЕРЕЊЕ НА МОТОРИЧКИТЕ СПОСОБНОСТИ

Мерењето е реализирано во стандардни училишни услови на редовните часови по физичко и здравствено образование. Мерењето го реализираа стручни лица од областа на кинезиологија кои претходно беа оспособени за мерење на определен моторички тест.

Моторните тестови беа мерени во сала и на отворен терен. Просторот каде се вршеше мерењето беше опремен со сите потребни реквизити и инструменти за предложените тестови. Температурата во салата и на отворен терен се движи од 17 - 22°C. Сите испитаници беа во спортска опрема и поделени во групи.

Предвидените тестови беа мерени според методологијата препорачана од страна на Советот на Европа (батерија на ЕУРОФИТ тестови), а дел од истите се модифицирани и прилагодени во меѓународните научни проекти „Feeding and Assessment of Nutritional Status of Spanish Adolescents“ (AVENA study), „The Healthy Lifestyle in Europe by Nutrition in Adolescence“ - (HELENA study) (Ruiz et al., 2006), „Assessing fitness in children and adolescents“ - (ALPHA Project) и „Identification and prevention of dietary and lifestyle induced health effects in children and infants“ - (IDEFICS study).

#### 4.7. УТВРДУВАЊЕ НА ПРИМЕЛИВОСТ НА МЕРКИТЕ И ТЕСТОВИТЕ ВО УЧИЛИШНО ОКРУЖУВАЊЕ

Применливоста на мерките и тестовите ќе биде утврдено со помош на структуриран анкетен прашалник. Прашалникот ќе биде од затворен тип и наставниците ќе одговараат со да/не на следните прашања. 1. Дали децата носеа соодветна спортска облека при изведување на фитнес тестовите поврзани со здравјето. 2. Информации за тоа дали инструкциите на тестовите биле правилно разбрани од учениците. 3. Дали некој ученик одбива да ги изврши мерењата, и причината/е (1 = срамежливост, 2 = недостаток на мотивација). 4. Исто така, на наставниците по физичко образование ќе им бидат поставени прашања за (i) соодветноста на капацитети (објектите) во училиштето за спроведување на тестовите, (ii) дали сметаат дека тестовите се лесни за спроведување и (iii) нивното претходно искуство во спроведувањето на овие тестови. 5. Понатаму, ќе биде регистрирано и времето потребно од наставникот по физичко образование за подготовка и спроведување на тестовите. Прифатливо ниво на применливост ќе биде

земено ако наставникот одговори „позитивни“ на 95% од одговорите (Waninge et al., 2009).

#### 4.8. УТВРДУВАЊЕ НА БЕЗБЕДНОСТ НА МЕРКИТЕ И ТЕСТОВИТЕ

За да се утврди безбедноста на мерките и тестовите ќе бидат поставени следните прашања на кои наставниците ќе треба да одговорат со да/не. 1. Алергија на инструментите, во случај на антропометриските проценки и моторичките тестирања. 2. Дали ученикот има чувство на болка и мачнина при изведување на тестовите. 3. Дали чувствуваат болка на дланката и подлактицата при изведување на тестот динамометрија на дланка. 4. Ќе бидат евидентирани мускулно-скелетните повреди за време или по процената на фитнес тестовите. 5. Болка во мускулите со одложен почеток (DOMS) ќе биде оценета со прашалник кој ќе биде пополнет 2 до 3 дена по тестирањето. Овој прашалник ќе вклучува прашања за искуството и сериозноста на DOMS, локацијата на болката или болките, од конкретен тест што може да предизвика DOMS и функционалните последици на DOMS во вообичаените активности. Прашалникот поврзан со DOMS ќе биде пополнет со индивидуални интервјуа. Прифатливо ниво на безбедност ќе се смета кога на прашањата е позитивно одговорено во 99% од случаите.

Нивото на безбедност треба да биде повисока од нивото на применливост, бидејќи сметавме дека безбедноста е важен аспект на тестовите за проценка физичкиот фитнес. Применливоста и безбедноста на тестовите ќе бидат проценета од двајца обучени истражувачи со претходно искуство. Истите ќе ја користат методата на директно набљудување при изведување на тестовите од страна на наставниците по физичко образование.

#### 4.9. УТВРДУВАЊЕ НА ТЕСТ-РЕТЕСТ РЕЛИЈАБИЛНОСТА НА МЕРКИТЕ И ТЕСТОВИТЕ

За да го испитаме степенот до кој резултатите од мерките и тестовите ќе останат стабилни, планиравме истите да бидат измерени во два наврати, при што второто мерење ќе биде во период помеѓу седум и десет дена по првото мерење. Второто мерење ќе биде реализирано под исти услови и во исто време како и првото мерење за да се минимизира варијабилноста на деноноќниот ритам.

#### 4.10. МЕТОД ЗА ОБРАБОТКА НА ПОДАТОЦИТЕ

Во рамките на ова истражување беше применет сет на статистички процедури кои овозможија систематична кондензација, трансформација и интерпретација на примарните податоци, на начин што директно се поврзуваше со поставените цели и работни хипотези. При изборот на методите се водеше грижа тие да бидат методолошки конзистентни со современите препораки за студии на релијабилност и за анализа на полови разлики во физичкиот и моторичкиот развој кај деца и адолесценти, како и резултатите да бидат лесно репродуцибилни и употребливи во научната и педагошката практика.

За сите антропометриски мерки и моторички тестови, кои беа на интервална и рачно мерна скала, најнапред беа пресметани основните дескриптивни статистички показатели: аритметичка средина, стандардна девијација, коефициент на варијабилност изразен во проценти, како и минимална и максимална вредност. Паралелно, за секоја променлива беше определена асиметричноста на дистрибуцијата (скјунис) и издолженоста односно плоскавоста (куртозис), со цел да се добие појасна слика за формата на распределбата и можните отстапувања од теоретски нормалната крива. Нормалноста на дистрибуцијата на резултатите беше тестирана со Колмогоров Смирнов тестот, а добиените вредности на статистиката и нивото на статистичка значајност послужија како основа за избор на последователните параметарски или непараметарски процедури, каде што тоа беше релевантно.

Релативната релијабилност на антропометриските и моторичките показатели беше проценета преку интракласни корелациски коефициенти од типот ICC(2,1), базирани на двонасочен модел со случајни ефекти и критериум на апсолутна согласност. Овој пристап овозможи да се процени степенот до кој рангирањето на испитаниците остана стабилно помеѓу првото и второто мерење, одделно кај момчињата и кај девојчињата. За секој тест беа пресметани точковни процени на ICC со соодветни интервали на доверба, а релијабилноста беше интерпретирана според вообичаените критериуми, при што вредности над 0,80 беа толкувани како одлична, вредности меѓу 0,61 и 0,80 како добра, а пониски вредности како умерена или недоволна релијабилност. Интракласните корелациски коефициенти беа пресметани со помош на специјализиран статистички софтвер MedCalc за Windows, верзија 15.2.2.

Пред да се процени апсолутната релијабилност, за секоја променлива беше испитана можната појава на хетероскедастичност, односно зависност на варијансата од големината на измерените вредности. Таа проценка беше изведена преку пресметување на Пирсоновите коефициенти на корелација помеѓу апсолутната разлика меѓу тестот и ретестот, од една страна; и средната вредност од тестот и ретестот, од друга страна. Кога коефициентот на корелација беше близок до нула и не надминуваше вредност од околу 0,10 податоците се сметаа за хомоскедастички и апсолутната релијабилност беше проценувана со стандардната грешка на мерењето. Во случаи кога корелацијата беше поизразена, податоците беа интерпретирани како хетероскедастички и како поадекватен показател за апсолутната релијабилност се користеше коефициентот на варијација.

Стандардната грешка на мерењето беше пресметана според стандардната равенка  $SEM = SD \times \sqrt{(1 - ICC)}$ , при што  $SD$  претставуваше здружено стандардно отстапување на резултатите од првото и второто мерење, а  $ICC$  интракласниот корелациски коефициент за соодветниот тест. За да се добие споредлив показател во релативни единици,  $SEM$  беше нормализирана преку делење со средната вредност од двете мерења и множење со 100, со што се доби  $SEM\%$  како индикатор за типична грешка изразена во проценти од просечниот резултат. Паралелно, за хетероскедастичните променливи беше пресметан и коефициентот на варијација, дефиниран како однос меѓу стандардната девијација на разликите и просечните вредности, исто така изразен во проценти.

Од  $SEM$  беше изведена најмалата детектибилна промена, дефинирана како  $SDD = 1,96 \times \sqrt{2} \times SEM$ , која претставува граница над која со голема веројатност може да се тврди дека разликата меѓу две мерења ја надминува случајната мерна грешка. Во одделни анализи оваа вредност беше нормализирана во однос на просечниот резултат, со цел да се добие и процентуален показател за минимално значајната промена. На тој начин апсолутната релијабилност беше квантитативно опишана и беше можно прецизно да се одреди колкава промена во резултатот е потребна за да се интерпретира како реална, а не како последица на мерен шум.

За да се утврди постоењето на систематска разлика помеѓу првото и второто мерење, односно можен ретест ефект, за секоја променлива беше спроведен парен  $t$  тест за зависни примероци, одделно кај момчињата и кај девојчињата. Паралелно со статистичката значајност, беше пресметана и големината на ефектот преку коефициентот  $d_z$ , со што се оценуваше практичната значајност на утврдените разлики. Вака беше можно да се разграничи ситуацијата во која статистички значајната разлика

е всушност мала и без поголемо практично влијание, од ситуацијата во која промената е и статистички и клинички или педагошки релевантна.

Согласноста меѓу првото и второто мерење на индивидуално ниво дополнително беше анализирана со Bland и Altman методологијата. За секоја променлива беа конструирани графички прикази во кои на хоризонталната оска се прикажуваше средната вредност од тестот и ретестот за секој испитаник, а на вертикалната оска нивната разлика. Врз основа на овие податоци беа пресметани средната разлика, која го претставуваше систематскиот bias, како и девијацијата на разликите, од која произлегуваа 95 проценти граници на согласност. Со анализа на распоредот на точките околу линијата на bias и ширината на границите на согласност беше проценето дали при повторното мерење резултатите остануваат во прифатлив опсег на варијација и дали ретест вредностите можат практично да го заменат почетното мерење без промена на интерпретацијата на статусот на поединецот. Истовремено, преку визуелна и корелациска анализа се проценуваше и можната хетероскедастичност, односно тенденцијата поголемите вредности да имаат поголеми апсолутни отстапувања.

За проценка на половите разлики во антропометриските карактеристики и моторичките способности беше применета мултиваријантна анализа на коваријанса со полот како фиксен фактор и хронолошката возраст како коваријата. Преку Wilks lambda, придружните F статистики и нивото на значајност се оценуваше мултиваријантниот ефект на полот врз збирната варијација на сите вклучени зависни променливи. Во случаи кога мултиваријантниот модел беше статистички значаен, следеше униваријатна анализа на коваријанса за поединечните променливи, при што се пресметуваа F вредности, ниво на значајност и парцијални  $\eta^2$  коефициенти како мерка на големината на ефектот. На овој начин беше можно не само да се утврди дали постојат полови разлики, туку и да се определи нивната практична важност и да се идентификуваат оние тестови кај кои полот има најсилно влијание.

Сите анализи беа спроведени со однапред утврдено ниво на статистичка значајност од  $p < 0,05$ , при што граничните вредности до  $p < 0,10$  беа интерпретирани внимателно и во контекст на големината на ефектот и параметрите на релијабилност. Комбинацијата од дескриптивни показатели, индикатори за релативна и апсолутна релијабилност, тестови на разлики и мултиваријантни модели овозможи сеопфатна статистичка обработка на податоците и формулирање на заклучоци со високо ниво на методолошка сигурност и практична применливост.

## 5. РЕЗУЛТАТИ

При процесот на собирање на податоците беа целосно почитувани сите основни методолошки принципи што се однесуваат на валидно, репрезентативно и етички изведено емпириско истражување. Истражувањето беше спроведено во строго контролирани услови, со јасно дефинирани процедури за мерење и протоколи за стандардизација на инструментите, со цел обезбедување на висока прецизност и веродостојност на добиените резултати.

По завршувањето на теренската фаза, следеше систематска обработка на податоците, која започна со нивна проверка и внесување во аналитичката матрица. Во оваа почетна етапа беше извршена детална логичка контрола на сите мерни листи, со цел идентификација и елиминација на можни грешки при мерењето и внесувањето. Притоа, беа отстранети или коригирани евидентно нелогичните и статистички неконзистентни вредности, како и податоците за кои се утврди дека произлегуваат од технички пропусти или од отстапувања од стандардната процедура на мерење.

Понатаму, беше спроведена иницијална анализа на дистрибуцијата на податоците, со интерпретација на нивната форма и со утврдување на можните причини за статистички значајни отстапувања од теоретски очекуваната Гаусова (нормална) дистрибуција. Анализата на нормалноста беше од суштинско значење за избор на соодветните статистички процедури во понатамошните фази на обработката.

За таа цел, кај сите антропометриски мерки и моторички тестови беа пресметани основните дескриптивни статистички показатели, вклучувајќи ја аритметичката средина (Mean) како мерка на централна тенденција и стандардната девијација (SD) како индикатор за степенот на распрнатост на резултатите околу средината. Дополнително, беа определени минималната и максималната вредност (Min–Max), со што се определи распонот на варијациите, како и коефициентот на варијабилност (CV%) кој овозможува споредба на релативната хомогеност помеѓу различни сетови на податоци.

Со цел подетално да се оцени обликот на дистрибуцијата, беа пресметани и индикаторите за асиметрија (Skewness) и издолженост или сплоштеност (Kurtosis), преку кои се утврдува степенот и насоката на девијацијата од нормалниот модел. За проверка на статистичката значајност на отстапувањата од нормалната дистрибуција беше применет Колмогоров–Смирновиот тест (KS), како стандардна процедура за тестирање на нормалност кај примероци.

Резултатите од спроведените дескриптивни анализи се прикажани во табелите 1 до 4, (поединечно за момчињата и девојчињата), кои обезбедуваат основа за понатамошната интерпретација, споредбена анализа и тестирање на хипотезите поставени во рамките на истражувањето.

### 5.1. ОСНОВНИ ДЕСКРИПТИВНИ СТАТИСТИЧКИ ПАРАМЕТРИ И НОРМАЛНОСТА НА ДИСТРИБУЦИЈАТА НА АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ОД ПРВОТО И ВТОРОТО МЕРЕЊЕ (ТЕСТ-РЕТЕСТ МЕРЕЊА) КАЈ МОМЧИЊАТА

Во Табела 1 се прикажани основните дескриптивни статистички параметри и резултатите од тестот на нормалност (Колмогоров–Смирнов) за антропометриските мерки и моторичките тестови од првото мерење кај момчињата. Најнапред, дистрибутивните индикатори сугерираат дека варијаблите во целина имаат уреден статистички профил, бидејќи кај сите мерки е добиено  $p > .20$ , што значи дека не се регистрира статистички значајно отстапување од нормална распределба. Овој наод е важен методолошки, затоа што укажува дека параметарските анализи можат да се применуваат без потреба од трансформации на податоците, а истовремено овозможува појасно толкување на средните вредности и варијабилноста како репрезентативни показатели за групата. Дополнително, вредностите на скјунис и куртоза во најголем дел се умерени, што укажува дека распределбите се приближно симетрични и без изразена концентрација на екстремни резултати.

Календарската возраст на момчињата изнесува во просек 16,69 години, со релативно тесен распон од 15,85 до 17,99 години, што потврдува дека примерокот е хомоген во однос на возраста и дека се работи за релативно компактна адолесцентна кохорта. Стандардната девијација е 0,57, а  $CV = 3,45\%$  дополнително ја нагласува ниската релативна варијабилност и стабилноста на возраста во примерокот. Стандардната грешка на средната вредност е многу мала ( $S.E = 0,07$ ), што укажува на висока прецизност на проценката на просекот. Скјунисот (0,39) покажува благо позитивно навалување, односно мал број испитаници со нешто повисока возраст ја продолжуваат распределбата кон десно, додека негативната куртоза ( $-0,86$ ) сугерира порамнета форма, со пошироко распоредени вредности околу средината, без нагласено групирање. Колмогоров–Смирновиот тест со  $p > .20$  потврдува дека овие отстапувања се мали и не ја нарушуваат нормалноста.

Телесната висина има просечна вредност од 178,13 cm, со распон од 162,00 до 195,00 cm, што е типичен биолошки распон за машка адолесцентна популација и укажува на присуство на очекувана интериндивидуална варијабилност во линеарниот раст. Стандардната девијација изнесува 6,90 cm, а  $CV = 3,87\%$  посочува ниска релативна варијабилност, што значи дека висината е релативно стабилна и хомогена во групата. Стандардната грешка ( $S.E = 0,88$ ) укажува на солидна прецизност на средната вредност. Скјунисот е низок (0,20), што значи дека распределбата е речиси симетрична, а куртозата е блиска до нула ( $-0,05$ ), што укажува на форма блиска до нормалната. Со  $p > .20$  се потврдува дека телесната висина е статистички уредно распределена и погодна за параметарска обработка.

Телесната тежина во првото мерење има просек од 69,20 kg, со распон од 50,70 до 102,00 kg, што укажува на значајна хетерогеност во телесната маса, како очекуван резултат од комбинацијата на различни степени на биолошка зрелост, телесна композиција и можни разлики во физичката активност. Стандардната девијација е 10,19 kg, а  $CV = 14,72\%$  означува умерена релативна варијабилност, односно дека тежината варира повеќе од висината во однос на просекот. Стандардната грешка ( $S.E = 1,29$ ) е умерена и укажува дека проценката на средната тежина е стабилна, но сепак посензитивна на дисперзијата на индивидуалните вредности. Скјунисот (0,63) укажува на умерена позитивна асиметрија, што значи дека мал број момчиња со повисока телесна маса ја продолжуваат распределбата кон десно. Куртозата (1,17) сугерира зголемена зашиленост, односно поголема концентрација на резултатите околу средината со присуство на поизразени крајни вредности. Сепак,  $p > .20$  потврдува дека и покрај оваа структура, распределбата не отстапува значајно од нормалната.

BMI има средна вредност од 21,75 kg/m<sup>2</sup>, со минимална и максимална вредност од 17,34 до 27,57 kg/m<sup>2</sup>. Овој распон укажува дека во примерокот се присутни различни телесни профили, од послаби кон поизразено покрупни структури, што е во согласност со варијабилноста забележана кај телесната тежина. Стандардната девијација е 2,51, а  $CV = 11,54\%$  укажува на умерена релативна варијабилност, нешто помала од тежината поради стандардизацијата преку висината. Стандардната грешка е ниска ( $S.E = 0,32$ ), што укажува на прецизна проценка на групниот просек. Скјунисот (0,39) покажува благо позитивно навалување, додека куртозата ( $-0,33$ ) упатува на порамнета форма на распределбата. Со  $p > .20$ , BMI е распределен без значајни отстапувања од нормалноста, што овозможува негово користење како континуирана варијабла во понатамошните модели и споредби.

Обемот на половината има средна вредност од 77,61 cm, со распон од 67,00 до 93,00 cm, што укажува на умерена варијабилност во централната телесна димензија. Стандардната девијација изнесува 5,99 cm, додека  $CV = 7,72\%$  означува ниска до умерена релативна варијабилност, што е повисока од висината, но пониска од тежината и силовите тестови. Стандардната грешка ( $S.E = 0,76$ ) е ниска и укажува на стабилна проценка на просечната вредност. Скјунисот (0,42) сугерира блага позитивна асиметрија, односно мал број испитаници со поголем обем ја продолжуваат распределбата кон десно, додека куртозата (-0,42) упатува на порамнета распределба без нагласено групирање. Колмогоров–Смирновиот тест со  $p > .20$  потврдува дека распределбата останува во рамките на нормалноста.

Претклонот во сед има просек од 18,40 cm и екстремно широк распон од 0,00 до 37,00 cm, што укажува на изразена хетерогеност во флексибилноста кај момчињата. Ова е дополнително потврдено со високата стандардна девијација (9,50) и особено со највисокиот CV во табелата (51,62%), што значи дека релативната варијабилност е многу голема и дека индивидуалните разлики во овој домен доминираат над групната средина. Стандардната грешка ( $S.E = 1,21$ ) покажува дека просекот е проценет со умерена прецизност, но природата на варијаблоста носи поголема дисперзија. Скјунисот е речиси нула (0,06), што укажува на речиси симетрична распределба и покрај големата варијабилност, додека куртозата (-0,72) упатува на порамнета форма со пошироко распоредување на резултатите. Со  $p > .20$  се потврдува дека и оваа варијабла не отстапува значајно од нормалноста, што е важен предуслов за валидна параметарска анализа иако хетерогеноста останува висока.

Скокот од место има просечен резултат од 199,43 cm, со распон од 150,00 до 250,00 cm, што укажува на широк спектар на експлозивна сила на долните екстремитети во примерокот. Стандардната девијација е 23,48 cm, а  $CV = 11,77\%$  укажува на умерена релативна варијабилност, што значи дека иако постојат значајни индивидуални разлики, резултатите се доволно компактни за да се добие стабилен групен профил. Стандардната грешка ( $S.E = 3,03$ ) е повисока во апсолутни единици поради природата на тестот и дисперзијата, но останува прифатлива за проценка на средната вредност. Скјунисот е негативен (-0,23), што укажува на благо лево навалување, односно нешто поголема концентрација на повисоки резултати со мал број пониски вредности. Куртозата (-0,43) сугерира порамнета форма.  $p > .20$  потврдува дека распределбата е нормална во статистичка смисла.

Динамометријата на дланка покажува просечни вредности од 46,77 kg за десната и 44,79 kg за левата рака, со распони од 28,20 до 66,20 kg и од 27,00 до 63,00 kg, што ја отсликува очекуваната варијабилност на максималната изометриска сила кај адолесценти. Стандардните девијации (7,91 и 7,73) се слични, а CV вредностите (16,91% и 17,26%) укажуваат на релативно висока варијабилност, поголема од антропометриските мерки и ВМІ, но пониска од флексибилноста. Ова значи дека силата е домен со значајни индивидуални разлики, но сепак со доволна конзистентност за да се користи како дискриминативен тест во популацијата. Стандардните грешки (1,00 и 0,98) се умерени. Скјунисот кај десната рака е близок до нула и негативен (-0,08), што означува речиси симетрична распределба, додека кај левата рака е благо позитивен (0,25). Куртозата е умерена и позитивна кај двете (0,49 и 0,26), што сугерира благо повисока концентрација околу средината. Кога силата се сумира преку просечната динамометрија, се добива просек од 45,78 kg, SD = 7,52 и CV = 16,42%, со речиси нулта асиметрија (Skewn = 0,02) и умерена куртоза (0,53), што укажува дека агрегирањето преку просек создава најсиметричен и статистички уреден профил. Кај сите три показатели,  $p > .20$  потврдува дека распределбите се во рамки на нормалност.

Подигнувањето на трупот за 30 секунди има просек од 25,97 повторувања, со распон од 18,00 до 36,00, што покажува умерен до широк спектар на силова издржливост на трупот кај момчињата. Стандардната девијација е 4,46, а CV = 17,18% укажува на релативно висока варијабилност, но значително пониска од онаа кај флексибилноста. Стандардната грешка (S.E = 0,58) е ниска, што укажува на добра прецизност на проценката на просекот. Скјунисот (0,10) покажува речиси симетрична распределба, додека куртозата (-0,98) упатува на порамнета форма, односно резултатите се пошироко распоредени без силно групирање околу средината. К-S тестот со  $p > .20$  повторно потврдува нормалност.

Чунестото трчање 4×10 m има просек од 10,93 s со распон од 9,09 до 12,46 s, што укажува на релативно компактни резултати во агилноста кај оваа група. Стандардната девијација е 0,74 s, а CV = 6,78% е меѓу најниските во табелата, што означува висока хомогеност и релативно мала дисперзија во перформансот. Стандардната грешка (S.E = 0,10) е многу мала, што укажува на висока прецизност на средната вредност. Скјунисот е мал и негативен (-0,10), што сугерира речиси симетрична распределба со минимално лево навалување, а куртозата (-0,54) укажува на благо порамнета форма. Со  $p > .20$ , распределбата е усогласена со нормалноста, што го прави овој тест статистички стабилен за понатамошни анализи.

Сумирано, првото мерење кај момчињата покажува дека антропометриските индикатори се карактеризираат со ниска релативна варијабилност и приближно нормални распределби, додека моторичките тестови очекувано имаат поголема дисперзија, особено флексибилноста, која се издвојува со највисок CV и најголема хетерогеност. Сепак, фактот што кај сите варијабли е потврдено  $p > .20$  укажува дека дистрибутивните претпоставки се задоволени и дека описните параметри веродостојно го отсликуваат профилот на примерокот, обезбедувајќи стабилна основа за понатамошни анализи на релијабилност и групни споредби.

Табела 1. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од првото мерење кај момчињата

Варијабли	Mean	Min	Max	SD	CV%	S.E	Skewn	Kurto	K-S
Години	16,69	15,85	17,99	0,57	3,45	0,07	0,39	-0,86	$p > .20$
Телесна висина	178,13	162,00	195,00	6,90	3,87	0,88	0,20	-0,05	$p > .20$
Телесна тежина	69,20	50,70	102,00	10,19	14,72	1,29	0,63	1,17	$p > .20$
ВМІ (kg/m <sup>2</sup> )	21,75	17,34	27,57	2,51	11,54	0,32	0,39	-0,33	$p > .20$
Обем на половина	77,61	67,00	93,00	5,99	7,72	0,76	0,42	-0,42	$p > .20$
Претклон во сед	18,40	0,00	37,00	9,50	51,62	1,21	0,06	-0,72	$p > .20$
Скок од место	199,43	150,00	250,00	23,48	11,77	3,03	-0,23	-0,43	$p > .20$
Динамометрија десна	46,77	28,20	66,20	7,91	16,91	1,00	-0,08	0,49	$p > .20$
Динамометрија лева	44,79	27,00	63,00	7,73	17,26	0,98	0,25	0,26	$p > .20$
Динамометрија	45,78	27,60	62,85	7,52	16,42	0,95	0,02	0,53	$p > .20$
Поди. груп 30 s	25,97	18,00	36,00	4,46	17,18	0,58	0,10	-0,98	$p > .20$
Чунесто 4×10 m	10,93	9,09	12,46	0,74	6,78	0,10	-0,10	-0,54	$p > .20$

Во Табела 2 се прикажани основните дескриптивни статистички параметри и резултатите од тестот на нормалност (Колмогоров Смирнов) за антропометриските мерки и моторичките тестови од второто мерење, односно ретест мерењата кај момчињата. Општиот профил на резултатите укажува на висока стабилност на централните тенденции и на очекувани варијации меѓу испитаниците, при што формата на распределбите во најголем дел останува приближно нормална. За речиси сите варијабли е добиено  $p > .20$  на Колмогоров Смирновиот тест, што значи дека нема статистички значајни докази за отстапување од нормална дистрибуција и дека податоците се соодветни за параметарски пристапи во натамошните анализи. Единствено кај обемот на половината се бележи маргинално отстапување од нормалност ( $p < .10$ ), што упатува дека распределбата може да биде благо нарушена, веројатно поради поголема хетерогеност или присуство на поизразени вредности во десниот дел на распределбата, па оваа варијабла заслужува нешто повнимателно толкување и евентуална проверка и со графички методи.

Кај телесната висина, средната вредност на ретест изнесува 178,24 cm, со минимална и максимална вредност од 162,00 до 195,50 cm, што покажува широк но очекуван распон за адолесцентна машка популација. Стандардната девијација е 6,91 cm, што укажува на умерено расејување околу просекот, а коефициентот на варијација ( $CV = 3,88\%$ ) дополнително потврдува дека релативната варијабилност е ниска и дека резултатите се концентрирани околу средната вредност. Стандардната грешка на средината ( $S.E = 0,88$ ) е мала, што значи дека проценката на популацискиот просек е прецизна и дека средната вредност е релативно стабилна во однос на примерочната варијација. Скјунисот ( $Skewn = 0,19$ ) е многу близок до нула, што упатува на речиси симетрична распределба, а куртозата ( $Kurto = -0,04$ ) е практично нулта, што значи дека распределбата по форма е блиска до нормална без изразена зашиленост или порамнетост. Конзистентно со ова, Колмогоров Смирновиот тест покажува  $p > .20$ , односно нема статистички значајно отстапување од нормалност.

Телесната тежина на ретест има средна вредност 69,18 kg, со распон од 50,00 до 103,00 kg, што укажува на значајни индивидуални разлики во телесната маса, типични за оваа возраст. Стандардната девијација е 10,24 kg и е повисока во апсолутни единици од висината, што ја одразува поголемата хетерогеност на оваа варијабла. Коефициентот на варијација ( $CV = 14,81\%$ ) е умерено висок, што покажува дека релативната варијабилност е значајна и дека телесната тежина се разликува повеќе меѓу поединците отколку телесната висина. Стандардната грешка ( $S.E = 1,30$ ) сугерира дека проценката на средната тежина е нешто помалку прецизна споредено со висината, што е очекувано поради поголемата дисперзија. Скјунисот е умерено позитивен ( $Skewn = 0,75$ ), што значи дека има тенденција распределбата да биде навалена кон десно, односно постои помал број испитаници со повисоки вредности кои ја подигаат „опашката“ на распределбата. Куртозата ( $Kurto = 1,48$ ) укажува на поизразена зашиленост и поголема концентрација на вредности околу центарот со потенцијално поизразени крајни вредности, што е конзистентно со присуство на испитаници со повисока телесна маса. Сепак, и покрај оваа форма, K S резултатот  $p > .20$  сугерира дека отстапувањето не е доволно изразено за статистички да ја наруши нормалноста.

Индексот на телесна маса на ретест (BMI) има просек 21,72 kg/m<sup>2</sup>, со распон од 17,10 до 27,74 kg/m<sup>2</sup>. Стандардната девијација е 2,49, а  $CV = 11,45\%$ , што означува умерена релативна варијабилност, помала од телесната тежина но поголема од висината, што е логично бидејќи BMI интегрира информација за масата во однос на висината. Стандардната грешка ( $S.E = 0,32$ ) е ниска и покажува добра прецизност на проценетиот

просек. Скјунисот ( $Skewn = 0,43$ ) укажува на блага позитивна асиметрија, односно на дискретно поместување кон повисоки ВМІ вредности кај помал број испитаници, додека куртозата ( $Kurto = -0,18$ ) е малку негативна и упатува на благо порамнета распределба, без изразена концентрација на екстрими. Колмогоров Смирновиот тест повторно дава  $p > .20$ , што сугерира дека распределбата е прифатливо блиска до нормалната.

Обемот на половината на ретест има средна вредност 77,56 cm, со распон од 67,00 до 99,00 cm. Стандардната девијација е 6,01 cm, со  $CV = 7,75\%$ , што покажува ниска до умерена релативна варијабилност, поголема од висината но значително пониска од тежината. Стандардната грешка ( $S.E = 0,76$ ) укажува на солидна прецизност на средната вредност. Меѓутоа, скјунисот е изразено позитивен ( $Skewn = 0,92$ ) и куртозата е позитивна ( $Kurto = 1,27$ ), што сугерира дека распределбата е навалена кон десно и дека постои поголема концентрација околу центарот со присуство на повисоки вредности кај одделни испитаници, односно потенцијално постојат поизразени „крајни“ вредности кои ја нарушуваат идеалната симетрија. Ова е во согласност со маргинално значајниот резултат на K S тестот ( $p < .10$ ), што укажува дека за оваа варијабла распределбата може да биде посензитивна на отстапувања од нормалност. Практично, тоа не мора да биде проблем за општа интерпретација, но оправдува повнимателен избор на аналитички пристап во зависност од целта, како и дополнителна визуелна проверка со хистограми или Q-Q графикони, особено ако се планираат чувствителни параметарски модели.

Кај претклонот во сед, ретест просекот е 20,96 cm, со распон од 0,00 до 38,00 cm, што покажува дека дел од испитаниците имаат многу ниска флексибилност, додека други достигнуваат високи вредности. Стандардната девијација е 9,64 cm, а  $CV = 45,99\%$ , што укажува на многу висока релативна варијабилност и силно изразени индивидуални разлики во флексибилноста. Стандардната грешка ( $S.E = 1,22$ ) е умерена и ја одразува оваа дисперзија, односно просекот е нешто помалку прецизно проценет во споредба со антропометриските мерки. Скјунисот ( $Skewn = -0,08$ ) е практично нула и упатува на симетрична распределба, додека куртозата ( $Kurto = -0,88$ ) е негативна и укажува на порамнета распределба, односно резултатите се пошироко распоредени низ опсегот без силна концентрација околу средината. K-S тестот е  $p > .20$ , што значи дека и покрај големата варијабилност, распределбата не е статистички далеку од нормална.

Скокот од место при ретест има средна вредност 206,26 cm, со распон од 150,00 до 270,00 cm. Стандардната девијација е 26,07 cm, а  $CV = 12,64\%$ , што укажува на умерена релативна варијабилност и значајни, но не екстремни разлики во експлозивната сила на долните екстремитети. Стандардната грешка ( $S.E = 3,37$ ) е релативно повисока

затоа што скалата е во сантиметри и варијабилноста е вообичаено поголема кај моторичките тестови. Скјунисот ( $Skewn = 0,06$ ) и куртозата ( $Kurto = 0,00$ ) укажуваат на речиси идеално симетрична распределба со форма практично идентична на нормалната. Колмогоров Смирновиот тест ( $p > .20$ ) ја потврдува оваа оценка, што значи дека резултатите се дистрибуирани на начин кој поддржува параметарска интерпретација.

Динамометријата на дланка покажува многу конзистентен профил меѓу десната и левата рака. За десната рака просекот е  $46,97 \text{ kg}$  ( $Min = 29,50$ ;  $Max = 66,80$ ), со  $SD = 7,84$  и  $CV = 16,68\%$ , што означува умерена релативна варијабилност во максималната сила. Стандардната грешка ( $S.E = 1,00$ ) упатува на стабилна проценка на средината. Скјунисот ( $0,20$ ) и куртозата ( $0,50$ ) укажуваат на блага позитивна асиметрија и умерено зашилена распределба, без сериозни отстапувања. За левата рака просекот изнесува  $44,65 \text{ kg}$  ( $Min = 28,80$ ;  $Max = 60,00$ ), со  $SD = 7,51$  и  $CV = 16,82\%$ , што е речиси идентично ниво на релативна дисперзија. Скјунисот ( $0,22$ ) е исто така благо позитивен, додека куртозата е негативна ( $-0,49$ ), што сугерира нешто порамнета распределба во споредба со десната рака, но без нарушување на нормалноста ( $p > .20$ ). Кога двете раце се разгледуваат преку агрегирана мерка на сила, просекот е  $45,81 \text{ kg}$ , со  $SD = 7,40$  и  $CV = 16,16\%$ , што укажува дека комбинирањето на информацијата доведува до минимално намалување на релативната варијабилност. Скјунисот ( $0,22$ ) и куртозата ( $0,17$ ) повторно укажуваат на стабилна и приближно нормална распределба ( $p > .20$ ), што ја поддржува употребата на оваа мерка како општ индикатор за максимална сила на дланка.

Кај подигнувањето на трупот за 30 секунди, ретест просекот е  $27,68$  повторувања, со распон од 18 до 38 повторувања. Стандардната девијација е  $4,31$ , а  $CV = 15,56\%$ , што укажува на умерена релативна варијабилност, односно испитаниците се разликуваат во силовата издржливост на трупот, но распоредот на резултати е релативно стабилен. Стандардната грешка ( $S.E = 0,56$ ) е мала и упатува на прецизна проценка на средната вредност. Скјунисот ( $0,06$ ) е речиси нулта, а куртозата ( $-0,70$ ) е умерено негативна, што сугерира дека распределбата е симетрична и нешто порамнета, односно резултатите се распоредени пошироко низ опсегот без силна концентрација околу средината. K-S тестот ( $p > .20$ ) повторно потврдува прифатлива нормалност.

Чунестото трчање  $4 \times 10 \text{ m}$  на ретест има средна вредност  $10,90 \text{ s}$ , со распон од  $9,32$  до  $12,77 \text{ s}$ . Стандардната девијација е  $0,78 \text{ s}$ , а  $CV = 7,14\%$ , што упатува на релативно ниска варијабилност и добра хомогеност на агилноста во примерокот. Стандардната грешка ( $S.E = 0,10$ ) е многу мала, што значи дека средната вредност е проценета со висока прецизност. Скјунисот ( $0,19$ ) укажува на мала позитивна асиметрија, додека

куртозата (-0,55) сугерира благо порамнета распределба, односно без изразени екстреми. Со оглед на  $p > .20$  од K-S тестот, распределбата е статистички прифатлива како нормална.

Сумирано, Табела 2 покажува дека во ретест мерењата кај момчињата централните вредности остануваат стабилни и дистрибуциите на повеќето антропометриски и моторички индикатори се приближно нормални. Ниските CV вредности кај возраста, телесната висина и чунестото трчање укажуваат на хомогени карактеристики во примерокот, додека повисоките CV кај телесната тежина, динамометријата и особено флексибилноста го рефлектираат очекуваниот степен на индивидуални разлики во телесниот состав и моторичките способности. Единствениот сигнал за потенцијално отстапување од нормалност се забележува кај обемот на половината, каде асиметријата и куртозата сугерираат дека кај дел од испитаниците постојат повисоки вредности кои ја нарушуваат симетријата, што е важно да се има предвид при натамошна интерпретација и моделирање. Во целина, дистрибутивните карактеристики од ретест мерењата обезбедуваат солидна статистичка основа за последователни параметарски анализи и за споредба со првото мерење во контекст на тест ретест релијабилноста.

Табела 2. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од второто мерење (ретест мерења) кај момчињата

Варијабли	Mean	Min	Max	SD	CV%	S.E	Skewn	Kurto	K-S
Телесна висина	178,24	162,00	195,50	6,91	3,88	0,88	0,19	-0,04	$p > .20$
Телесна тежина	69,18	50,00	103,00	10,24	14,81	1,30	0,75	1,48	$p > .20$
ВМI (kg/m <sup>2</sup> )	21,72	17,10	27,74	2,49	11,45	0,32	0,43	-0,18	$p > .20$
Обем на половина	77,56	67,00	99,00	6,01	7,75	0,76	0,92	1,27	$p < ,10$
Претклон во сед	20,96	0,00	38,00	9,64	45,99	1,22	-0,08	-0,88	$p > .20$
Скок од место	206,26	150,00	270,00	26,07	12,64	3,37	0,06	0,00	$p > .20$
Динамометрија десна	46,97	29,50	66,80	7,84	16,68	1,00	0,20	0,50	$p > .20$
Динамометрија лева	44,65	28,80	60,00	7,51	16,82	0,95	0,22	-0,49	$p > .20$
Динамометрија	45,81	29,15	63,20	7,40	16,16	0,94	0,22	0,17	$p > .20$
Поди. труп 30 s	27,68	18,00	38,00	4,31	15,56	0,56	0,06	-0,70	$p > .20$
Чунесто 4×10 m	10,90	9,32	12,77	0,78	7,14	0,10	0,19	-0,55	$p > .20$

## 5.2. ОСНОВНИ ДЕСКРИПТИВНИ СТАТИСТИЧКИ ПАРАМЕТРИ И НОРМАЛНОСТА НА ДИСТРИБУЦИЈАТА НА АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ОД ПРВОТО И ВТОРОТО МЕРЕЊЕ (ТЕСТ-РЕТЕСТ МЕРЕЊА) КАЈ ДЕВОЈЧИЊАТА

Во Табела 3 се прикажани основните дескриптивни статистички параметри и резултатите од тестот на нормалност (Колмогоров–Смирнов) за антропометриските мерки и моторичките тестови од првото мерење кај девојчињата. Генерално, за сите варијабли е добиено  $p > .20$ , што значи дека не постојат статистички значајни докази за отстапување од нормална распределба. Овој наод е важен од методолошки аспект, бидејќи сугерира дека распределбите се доволно блиски до нормалната форма и дека примената на параметарски процедури во понатамошните анализи е оправдана. Дополнително, вредностите на скјунис и куртоза кај повеќето променливи се умерени и без екстреми, што ја поткрепува оценката за стабилна и статистички уредна дистрибутивна структура на податоците во овој примерок.

Во однос на возраста, девојчињата имаат просечна вредност од 16,65 години, со распон од 15,75 до 17,76 години, што укажува на релативно хомогена група во рамките на средношколска возраст. Стандардната девијација е 0,59 години, а коефициентот на варијација ( $CV = 3,55\%$ ) потврдува ниска релативна варијабилност, односно дека возраста е добро контролирана и не претставува извор на големи индивидуални разлики. Стандардната грешка на средината ( $S.E = 0,13$ ) е мала, што упатува на прецизна проценка на просечната возраст во примерокот. Скјунисот (0,37) е благо позитивен, што значи дека има минимална тенденција кон поголем број испитаници со нешто пониска возраст и мал број со повисока возраст во рамките на опсегот, додека куртозата ( $-0,97$ ) укажува на порамнета распределба, односно поширока распределеност на вредностите околу средината без изразена зашиленост. Нормалноста е потврдена со  $p > .20$ .

Телесната висина има просек од 163,75 cm, со минимална и максимална вредност од 156,00 до 176,00 cm, што претставува очекуван распон за девојчиња во доцна адолесценција. Стандардната девијација е 5,65 cm, а  $CV = 3,45\%$ , што означува ниска релативна варијабилност и висока хомогеност на висината во примерокот. Стандардната грешка ( $S.E = 1,20$ ) укажува на умерена прецизност на проценката на средната вредност, што е во согласност со големината на варијабилноста и веројатната големина на примерокот. Скјунисот (0,34) упатува на блага позитивна асиметрија, додека куртозата ( $-0,70$ ) сугерира порамнета форма на распределбата. Со оглед на  $p > .20$ , распределбата

на висината може да се смета за нормална, што е типично за антропометриски мерки во ваква популација.

Телесната тежина покажува просечна вредност од 58,19 kg, со распон од 40,20 до 67,00 kg, што укажува на умерено широки индивидуални разлики во телесната маса. Стандардната девијација е 6,69 kg, а  $CV = 11,50\%$ , што означува умерена релативна варијабилност и поголема хетерогеност во споредба со висината, што е очекувано бидејќи телесната тежина е посензитивна на варијации во телесниот состав и нутритивниот статус. Стандардната грешка ( $S.E = 1,43$ ) е повисока, што сигнализира нешто помала прецизност на средната вредност во однос на мерките со помала варијабилност. Скјунисот е изразено негативен ( $-0,93$ ), што укажува на лево навалена распределба, односно поголема концентрација на вредности во повисокиот интервал со мал број испитаници со пониски тежини кои ја продолжуваат „опашката“ кон левата страна. Куртозата ( $0,81$ ) е позитивна и сугерира поизразена зашиленост, односно поголема концентрација околу центарот со одредена тенденција кон екстреми. И покрај ова, K-S тестот покажува  $p > .20$ , што значи дека овие отклонувања не се доволно големи за статистички да ја нарушат нормалноста.

ВМІ има просек од 21,71 kg/m<sup>2</sup>, со распон од 16,31 до 25,56 kg/m<sup>2</sup>. Стандардната девијација е 2,41, а  $CV = 11,08\%$ , што укажува на умерена релативна варијабилност, слична на телесната тежина, но со нешто поурамнотежен профил поради корекцијата за висина. Стандардната грешка ( $S.E = 0,51$ ) е ниска, што укажува на релативно прецизна проценка на средниот ВМІ. Скјунисот ( $-0,36$ ) упатува на блага негативна асиметрија, што значи дека има дискретна тенденција резултатите да се групираат во повисоките ВМІ интервали со нешто подолга „опашка“ кон пониските вредности. Куртозата ( $-0,40$ ) сугерира благо порамнета распределба, без изразена концентрација на екстреми. Нормалноста е потврдена со  $p > .20$ , што ја поддржува употребата на ВМІ во натамошни параметарски анализи.

Обемот на половината има просек од 71,45 cm, со распон од 63,00 до 81,00 cm. Стандардната девијација е 5,17 cm, а  $CV = 7,24\%$ , што укажува на ниска до умерена релативна варијабилност и релативно хомогена распределба. Стандардната грешка ( $S.E = 1,10$ ) е умерена. Скјунисот е речиси нулта ( $0,03$ ), што значи дека распределбата е речиси симетрична, додека куртозата ( $-0,81$ ) укажува на порамнета форма со поширока распределеност на резултатите. Со  $p > .20$ , нема индикации за нарушена нормалност, што е важно бидејќи обемот на половината често може да покаже асиметрии во популации со поголема варијација во абдоминалната адипозност.

Претклонот во сед има средна вредност од 19,23 cm, со минимална и максимална вредност од 8,00 до 35,00 cm, што укажува на значајни индивидуални разлики во флексибилноста. Стандардната девијација е 8,02 cm, а CV = 41,70%, што претставува многу висока релативна варијабилност и сугерира дека флексибилноста е домен со изразена хетерогеност кај девојчињата. Стандардната грешка (S.E = 1,71) е релативно повисока, што е очекувано со оглед на големата дисперзија. Скјунисот (0,14) укажува на речиси симетрична распределба, додека куртозата (-1,10) сугерира јасно порамнета распределба, односно резултатите се распоредени пошироко низ опсегот без силна концентрација околу средината. K-S тестот  $p > .20$  укажува дека и покрај големата варијабилност, распределбата останува статистички прифатлива како нормална.

Скокот од место има просек од 151,00 cm и широк распон од 95,00 до 202,00 cm, што укажува на големи индивидуални разлики во експлозивната сила на долните екстремитети. Стандардната девијација е 28,03 cm, а CV = 18,56%, што означува умерено висока релативна варијабилност и значајна хетерогеност во моторичкиот капацитет. Стандардната грешка (S.E = 6,12) е висока во апсолутни единици, што е очекувано за моторички тестови со широка варијабилност. Скјунисот (-0,08) е речиси нулта, што значи дека распределбата е симетрична, а куртозата (-0,45) сугерира благо порамнета форма. Нормалноста е потврдена со  $p > .20$ , што укажува дека резултатите не се концентрирани во екстрими и имаат статистички уреден распоред.

Кај динамометријата на дланка, десната рака има просек од 29,08 kg со распон од 16,00 до 40,10 kg. Стандардната девијација е 5,63 kg, а CV = 19,35%, што укажува на релативно висока варијабилност во максималната сила. Стандардната грешка (S.E = 1,20) покажува умерена прецизност на средната вредност. Скјунисот е благо негативен (-0,38), што сугерира мала тенденција кон концентрација на повисоки вредности со неколку пониски резултати, додека куртозата (0,68) укажува на умерена зашиленост. Левата рака има просек 27,96 kg, со распон 16,50 до 39,50 kg, SD = 5,90 и CV = 21,11%, што означува уште поголема релативна варијабилност во споредба со десната рака. Скјунисот (0,34) е благо позитивен, а куртозата (0,29) сугерира умерена концентрација околу средината без екстремна форма. Кога силата се разгледува преку агрегирана мерка (просек), просечната вредност изнесува 28,52 kg, со SD = 5,59 и CV = 19,61%, што покажува дека комбинирањето на двете мерења донекаде ја стабилизира релативната варијабилност. Скјунисот (-0,05) е практично нула, што укажува на висока симетричност, а куртозата (0,59) сугерира умерена зашиленост. Кај сите три мерки, K-S е  $p > .20$ , што укажува на прифатлива нормалност и стабилна распределба.

Подигнувањето на трупот за 30 секунди има просек од 20,57 повторувања, со распон од 12 до 32 повторувања. Стандардната девијација е 4,46, а  $CV = 21,66\%$ , што укажува на релативно висока варијабилност во силовата издржливост на трупот. Стандардната грешка ( $S.E = 0,97$ ) е умерена и упатува на солидна прецизност на средната вредност. Скјунисот (0,39) укажува на блага позитивна асиметрија, што значи дека мал број испитаници постигнуваат повисоки резултати и ја продолжуваат распределбата во десниот дел. Куртозата (1,05) упатува на поизразена зашиленост, што сугерира дека значаен дел од резултатите се групирани околу средината со потенцијално поизразени крајни вредности. Сепак,  $p > .20$  на К-S тестот значи дека дистрибуцијата не е статистички проблематична во однос на нормалноста.

Чунестото трчање  $4 \times 10$  m има просек од 12,84 s, со распон од 10,60 до 15,01 s. Стандардната девијација е 1,32 s, а  $CV = 10,31\%$ , што означува умерена релативна варијабилност и разумна хомогеност на агилноста и брзинската промена на насока во примерокот. Стандардната грешка ( $S.E = 0,30$ ) е ниска, што укажува на прецизна проценка на средната вредност. Скјунисот (0,06) е речиси нулта, што упатува на симетрична распределба, додека куртозата (-1,21) укажува на силно порамнета форма, односно резултатите се распределени пошироко без нагласена концентрација околу средината. Нормалноста е потврдена со  $p > .20$ , што обезбедува стабилна основа за натамошно статистичко моделирање.

Табела 3. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од првото мерење кај девојчињата

Варијабли	Mean	Min	Max	SD	CV%	S.E	Skewn	Kurto	K-S
Години	16,65	15,75	17,76	0,59	3,55	0,13	0,37	-0,97	$p > .20$
Телесна висина	163,75	156,00	176,00	5,65	3,45	1,20	0,34	-0,70	$p > .20$
Телесна тежина	58,19	40,20	67,00	6,69	11,50	1,43	-0,93	0,81	$p > .20$
ВМI ( $kg/m^2$ )	21,71	16,31	25,56	2,41	11,08	0,51	-0,36	-0,40	$p > .20$
Обем на половина	71,45	63,00	81,00	5,17	7,24	1,10	0,03	-0,81	$p > .20$
Претклон во сед	19,23	8,00	35,00	8,02	41,70	1,71	0,14	-1,10	$p > .20$
Скок од место	151,00	95,00	202,00	28,03	18,56	6,12	-0,08	-0,45	$p > .20$
Динамометрија десна	29,08	16,00	40,10	5,63	19,35	1,20	-0,38	0,68	$p > .20$
Динамометрија лева	27,96	16,50	39,50	5,90	21,11	1,26	0,34	0,29	$p > .20$
Динамометрија	28,52	16,25	39,80	5,59	19,61	1,19	-0,05	0,59	$p > .20$
Поди. труп 30 s	20,57	12,00	32,00	4,46	21,66	0,97	0,39	1,05	$p > .20$
Чунесто $4 \times 10$ m	12,84	10,60	15,01	1,32	10,31	0,30	0,06	-1,21	$p > .20$

Сумирано, Табела 3 покажува дека девојчињата во првото мерење имаат дистрибуции кои се статистички усогласени со нормалната распределба, со ниска релативна варијабилност кај возраста и телесната висина, умерена варијабилност кај телесната тежина, ВМІ и обемот на половината; и очекувано повисока хетерогеност кај моторичките тестови, особено флексибилноста и скокот од место. Ваквата дистрибутивна структура е типична за адолесцентна популација и обезбедува методолошки стабилна основа за споредби со ретест мерењата и за последователни анализи на тест ретест релијабилноста и согласноста.

Во Табела 4 се прикажани основните дескриптивни статистички параметри и резултатите од тестот на нормалност (Колмогоров–Смирнов) за антропометриските мерки и моторичките тестови од второто мерење, односно ретест мерењата кај девојчињата. Во најголем дел од варијабилите се добива  $p > .20$ , што укажува дека распределбите не покажуваат статистички значајни отстапувања од нормалноста и дека податоците се соодветни за параметарски анализи. Единствено кај подигнувањето на трупот за 30 секунди се регистрира гранично отстапување ( $p < .15$ ), што, заедно со изразените вредности на скјунис и куртоза, сугерира дека во оваа варијабла распределбата е по асиметрична и поостро „заострена“ отколку кај останатите. Сепак, со оглед дека ова не е класично значајно отстапување на конвенционално ниво и дека другите дистрибутивни индикатори кај останатите тестови се умерени, целокупниот дистрибутивен профил на ретест податоците може да се оцени како статистички стабилен.

Телесната висина на ретест има просек од 163,93 cm, со распон од 156,00 до 176,00 cm, што е практично идентично со распонот од првото мерење и укажува на очекувана биолошка стабилност на мерката во краток временски интервал. Стандардната девијација изнесува 5,60 cm, а коефициентот на варијација ( $CV = 3,42\%$ ) потврдува ниска релативна варијабилност и висока хомогеност на висината во примерокот. Стандардната грешка ( $S.E = 1,19$ ) укажува на умерено прецизна проценка на средната вредност. Скјунисот (0,34) покажува блага позитивна асиметрија, што значи дека мал број испитаници имаат повисоки вредности кои ја продолжуваат распределбата кон десно, додека куртозата (-0,70) сугерира порамнета форма на распределбата, односно пошироко распоредени вредности околу средината без доминација на екстрими. К-S тестот со  $p > .20$  ја потврдува нормалноста и во ретест услови.

Телесната тежина при второто мерење има средна вредност од 58,20 kg, со минимална и максимална вредност од 40,60 до 66,80 kg. Во споредба со првото мерење, просекот останува практично непроменет, што укажува на стабилен групен профил. Стандардната девијација е 7,03 kg, а  $CV = 12,07\%$  означува умерена релативна варијабилност, што е очекувано бидејќи телесната тежина е чувствителна на дневни и краткорочни промени поврзани со хидратација, гастроинтестинален статус и други биолошки фактори. Стандардната грешка ( $S.E = 1,50$ ) е умерена и упатува на солидна прецизност на средната вредност. Скјунисот ( $-0,86$ ) е изразено негативен, што укажува на лево навалена распределба, односно поголема концентрација на вредности во повисокиот интервал со мал број пониски тежини кои ја формираат „опашката“ кон левата страна. Куртозата ( $0,25$ ) сугерира умерено приближување до мезокуртична форма без екстремна зашиленост. Нормалноста е потврдена со  $p > .20$ , што значи дека овие асиметрии не се доволно силни за да ја нарушат статистичката претпоставка за нормална распределба.

ВМІ на ретест има просек од 21,66 kg/m<sup>2</sup>, со распон од 16,37 до 25,39 kg/m<sup>2</sup>. Вредностите се блиски до оние од првото мерење, што укажува на стабилност на индексот во групата. Стандардната девијација е 2,52, а  $CV = 11,62\%$  означува умерена релативна варијабилност, што е типично за ВМІ како индикатор кој комбинира варијации во телесната маса и висината. Стандардната грешка ( $S.E = 0,54$ ) е ниска, што укажува на релативно прецизна проценка на просекот. Скјунисот ( $-0,34$ ) покажува блага негативна асиметрија, а куртозата ( $-0,80$ ) укажува на порамнета распределба со пошироко распоредени вредности. Со  $p > .20$ , распределбата е статистички усогласена со нормалноста, што го поддржува користењето на ВМІ во натамошни споредбени и инференцијални анализи.

Обемот на половината во ретест мерењето има средна вредност од 71,73 cm, со распон од 60,00 до 81,00 cm, при што просекот е малку повисок во однос на првото мерење, но распонот останува компарабилен. Стандардната девијација е 5,65 cm, а  $CV = 7,88\%$  укажува на ниска до умерена релативна варијабилност. Стандардната грешка ( $S.E = 1,20$ ) е умерена. Скјунисот е негативен ( $-0,33$ ), што значи дека дел од испитаниците имаат пониски вредности кои ја продолжуваат распределбата кон левата страна, додека куртозата ( $-0,81$ ) укажува на порамнета форма, без силна концентрација околу средината. K-S тестот со  $p > .20$  сугерира дека распределбата е нормална, што е важно бидејќи обемот на половината често може да биде подложен на асиметрии во популации со поголема варијација во абдоминалната адипозност.

Претклонот во сед на ретест има просечна вредност од 21,55 cm, со распон од 9,00 до 35,00 cm. Во однос на првото мерење се забележува повисок просек, што сугерира подобрување на флексибилноста или подобра изведбена техника при повторното тестирање, но на ниво на дескриптивната статистика тоа пред сè се манифестира како поместување на средината без драматична промена на распонот. Стандардната девијација е 7,68 cm, а  $CV = 35,63\%$  укажува на висока релативна варијабилност, што потврдува дека флексибилноста останува домен со значајни индивидуални разлики. Стандардната грешка ( $S.E = 1,64$ ) е умерено висока поради дисперзијата на резултатите. Скјунисот (0,17) укажува на речиси симетрична распределба со минимално десно навалување, додека куртозата (-0,90) означува порамнета форма со пошироко распоредени резултати. Нормалноста е потврдена со  $p > .20$ , што сугерира дека и покрај хетерогеноста, распределбата не е статистички проблематична.

Скокот од место при второто мерење има просек од 153,62 cm и распон од 115,00 до 210,00 cm. Во споредба со првото мерење, средната вредност е нешто повисока, што е конзистентно со можноста за подобрување при повторно тестирање. Стандардната девијација изнесува 25,93 cm, а  $CV = 16,88\%$  укажува на умерено висока релативна варијабилност, односно значајна хетерогеност во експлозивната сила на долните екстремитети. Стандардната грешка ( $S.E = 5,66$ ) е висока во апсолутни единици, што е карактеристично за моторички тестови со широка дисперзија. Скјунисот (0,27) сугерира блага позитивна асиметрија, а куртозата (-0,76) укажува на порамнета распределба. K-S тестот со  $p > .20$  потврдува дека резултатите се доволно близу до нормалната распределба.

Кај динамометријата на дланка, десната рака има просек од 28,87 kg со распон од 16,60 до 39,20 kg. Стандардната девијација е 6,09 kg, а  $CV = 21,08\%$  означува висока релативна варијабилност, што сугерира изразени индивидуални разлики во максималната сила. Стандардната грешка ( $S.E = 1,30$ ) укажува на умерена прецизност на средната вредност. Скјунисот (0,01) е практично нула, што значи дека распределбата е многу симетрична, додека куртозата (-0,54) упатува на порамнета форма. Левата рака има просек од 27,94 kg, со распон од 14,20 до 42,50 kg, што покажува дека во ретест услови се јавуваат и нешто повисоки максимални резултати во горниот дел од распонот. Стандардната девијација е 6,52 kg, а  $CV = 23,34\%$  укажува на уште поголема релативна варијабилност. Скјунисот (0,32) е благо позитивен, додека куртозата (0,63) сугерира умерена зашпиленост, односно поголема концентрација на резултати околу средината со

присуство на поизразени крајни вредности. Кога се разгледува просечната динамометрија, средната вредност изнесува 28,40 kg,  $SD = 6,08$ ,  $CV = 21,41\%$ , со минимална асиметрија ( $Skewn = 0,06$ ) и куртоза блиска до нула ( $-0,07$ ), што сугерира најстабилен дистрибутивен профил кога силата се резимира преку просек од двете раце. Кај сите мерки на динамометрија,  $p > .20$  укажува на прифатлива нормалност.

Подигнувањето на трупот за 30 секунди во ретест мерењето има просек од 22,76 повторувања, со распон од 13 до 40 повторувања. Во однос на првото мерење, средната вредност е повисока, што може да сугерира подобрување во изведбата при повторно тестирање или поголема мотивација и оптимизација на ритамот. Стандардната девијација е 5,51, а  $CV = 24,22\%$  означува висока релативна варијабилност и изразени индивидуални разлики во силовата издржливост на трупот. Стандардната грешка ( $S.E = 1,20$ ) е умерена. Дистрибутивните параметри, сепак, упатуваат на поспецифичен профил: скјунисот е висок и позитивен (1,11), што значи дека поголемиот дел од резултатите се концентрирани во пониските и средни интервали, додека мал број испитаници постигнуваат многу високи резултати и ја продолжуваат распределбата кон десно. Куртозата е исклучително висока (4,08), што укажува на силно зашилена распределба со нагласено присуство на екстремни вредности. K-S тестот покажува  $p < .15$ , што сугерира дека оваа варијабла е најблиску до статистички забележливо отстапување од нормалноста и дека при натамошна инференција може да биде корисно да се разгледа и робусна интерпретација или дополнителна проверка на екстремите.

Чунестото трчање 4×10 m има просек од 12,81 s со распон од 10,09 до 15,20 s. Во споредба со првото мерење, просекот е практично непроменет, што укажува на стабилен групен профил на агилност во ретест услови. Стандардната девијација е 1,37 s, а  $CV = 10,73\%$  означува умерена релативна варијабилност и разумна хомогеност во резултатите. Стандардната грешка ( $S.E = 0,31$ ) е ниска и укажува на прецизна проценка на средната вредност. Скјунисот (0,13) е минимален, што сугерира речиси симетрична распределба, а куртозата ( $-0,46$ ) упатува на благо порамнета форма. K-S тестот со  $p > .20$  ја потврдува нормалноста, што обезбедува стабилна основа за понатамошни споредби и анализи на релијабилност.

Сумирано, ретест мерењата кај девојчињата покажуваат дека антропометриските мерки имаат ниска релативна варијабилност и статистички уредни распределби, додека моторичките тестови очекувано демонстрираат поголема дисперзија и повисоки CV вредности, што рефлектира реални индивидуални разлики во моторичките способности. Нормалноста е главно задржана кај сите променливи, со посебна напомена дека

подигнувањето на трупот за 30 секунди покажува најизразена асиметрија и зашиленост и граничен сигнал за отстапување од нормална распределба. Овој дистрибутивен профил е важен за правилно толкување на ретест резултатите и претставува релевантен контекст за подоцнежните анализи на тест ретест релијабилност и согласност

Табела 4. Основни дескриптивни статистички параметри и нормалноста на дистрибуцијата на антропометриските мерки и моторичките тестови од второто мерење (ретест мерења) кај девојчињата

Варијабли	Mean	Min	Max	SD	CV%	S.E	Skewn	Kurto	K-S
Телесна висина	163,93	156,00	176,00	5,60	3,42	1,19	0,34	-0,70	$p > .20$
Телесна тежина	58,20	40,60	66,80	7,03	12,07	1,50	-0,86	0,25	$p > .20$
ВМI (kg/m <sup>2</sup> )	21,66	16,37	25,39	2,52	11,62	0,54	-0,34	-0,80	$p > .20$
Обем на половина	71,73	60,00	81,00	5,65	7,88	1,20	-0,33	-0,81	$p > .20$
Претклон во сед	21,55	9,00	35,00	7,68	35,63	1,64	0,17	-0,90	$p > .20$
Скок од место	153,62	115,00	210,00	25,93	16,88	5,66	0,27	-0,76	$p > .20$
Динамометрија десна	28,87	16,60	39,20	6,09	21,08	1,30	0,01	-0,54	$p > .20$
Динамометрија лева	27,94	14,20	42,50	6,52	23,34	1,39	0,32	0,63	$p > .20$
Динамометрија	28,40	15,40	38,95	6,08	21,41	1,30	0,06	-0,07	$p > .20$
Поди. труп 30 s	22,76	13,00	40,00	5,51	24,22	1,20	1,11	4,08	$p < .15$
Чуносто трчање 4×10	12,81	10,09	15,20	1,37	10,73	0,31	0,13	-0,46	$p > .20$

Вкупната синтеза на Табелите 1–4 покажува дека примерокот кај двата пола е старосно хомоген и дека основните антропометриски и моторички индикатори имаат дистрибутивни својства кои во најголем дел се блиски до нормална распределба. Доминирачките  $p$  вредности од Колмогоров–Смирновиот тест (претежно  $p > .20$ ) упатуваат дека статистичките претпоставки за параметарска обработка се генерално прифатливи, но паралелното читање на скјунис и куртоза сугерира дека кај поединечни варијабли може да постојат суптилни отстапувања кои се релевантни за индивидуално ниво на интерпретација. Такви индикации се забележуваат кај обемот на половината кај момчињата при второто мерење, како и кај подигнувањето на трупот за 30 секунди кај девојчињата при ретестот, каде формата на распределбата сугерира зголемена асиметрија и концентрација на резултати околу одредени вредности. Оттука, иако глобалната слика ја поддржува употребата на класични аналитички процедури, останува методолошки оправдано резултатите да се интерпретираат со претпазливост кај варијаблите со поголема дисперзија и потенцијални отстапување, бидејќи токму таму најчесто се појавуваат поголеми индивидуални разлики и пошироки граници на очекувана мерна варијација.

Во однос на централната тенденција и варијабилноста, антропометриските мерки се карактеризираат со ниски до умерени коефициенти на варијација, што укажува на релативна хомогеност и стабилност во рамки на групите. Наспроти тоа, дел од

моторичките тестови, особено флексибилноста, покажуваат значително поголема релативна варијабилност, што означува широк распон на моторичките способности и потенцијално поголем простор за индивидуални промени. Притоа, споредбата меѓу првото и второто мерење открива конзистентна тенденција на подобра изведба при ретест кај тестовите кои се чувствителни на техника, ритам и запознавање со задачата, како што се претклонот во сед, скокот од место и подигнувањето на трупот. Овој образец е типичен за повторливи моторички процедури и укажува дека покрај „вистинските“ индивидуални разлики, второто мерење може да вклучува и систематска компонента поврзана со учење или оптимизирана изведба. Истовремено, мерките кои се помалку зависни од техника, како динамометријата, како и чунестото трчање, покажуваат минимални поместувања на средните вредности, што на групно ниво сугерира поголема стабилност и помал потенцијал за ефект од повторување.

Овие наоди создаваат јасна основа за следната аналитичка фаза, бидејќи потврдуваат дека дистрибутивниот профил на податоците е соодветен за процена на тест ретест квалитетот на мерењето, но истовремено нагласуваат дека за дел од моторичките тестови не е доволно да се разгледува само стабилноста на рангирањето. Поради присуството на можни систематски поместувања меѓу T1 и T2 кај одредени способности, потребно е паралелно да се квантифицира и апсолутната мерна грешка, како и степенот на согласност на индивидуално ниво. Затоа, во продолжение резултатите логично се надополнуваат со процена на релативната релијабилност преку ICC, проверка на систематска разлика меѓу мерењата, определување на SEM и минималниот праг за сигурна реална промена, и анализа на Bland–Altman согласноста, со што се обезбедува целосна и практично релевантна интерпретација на стабилноста и заменливоста на повторните мерења во рамки на МАКФИТ батеријата.

### 5.3. ТЕСТ-РЕТЕСТ РЕЛИЈАБИЛНОСТА И СОГЛАСНОСТА НА МАКФИТ БАТЕРИЈАТА, ОДДЕЛНО ЗА МОМЧИЊА И ДЕВОЈЧИЊА

Во ова поглавје се прикажани резултатите од тест-ретест релијабилноста и согласноста на МАКФИТ батеријата, одделно за момчиња и девојчиња. Наодите се презентирани во три аналитички рамки: (i) релативна релијабилност преку ICC(2,1) со 95% интервали на доверба и процена на систематска разлика меѓу T1 и T2 (paired t-test и dz), (ii) апсолутна релијабилност преку SEM/SEM%, MDC95/SDD и CV% како прагови за индивидуална „реална“ промена, и (iii) Bland–Altman согласност преку bias и 95%

граница на согласност, со дополнителна проверка за хетероскедастичност преку корелацијата меѓу средната вредност и апсолутната разлика. Овој пристап овозможува истовремено да се оцени стабилноста на рангирањето и големината на мерната грешка во практични единици. Дополнително, се проценува согласноста меѓу T1 и T2 на индивидуално ниво, односно дали разликата кај поединец е во прифатливи граници и дали повторното мерење може да го замени првото без да ја промени интерпретацијата за истиот поединец.

Резултатите од Табела 5 покажуваат дека батеријата на МАКФИТ тестови кај момчињата генерално демонстрира висока тест ретест стабилност, со јасно доминантна слика на одлична релативна релијабилност кај повеќето антропометриски и дел од моторичките индикатори. Највисоки вредности на ICC(2,1) се регистрирани кај телесната висина, каде релативната релијабилност е речиси совршена (ICC=0.999; 95% CI: 0.998–0.999), што укажува на исклучително конзистентно рангирање на испитаниците меѓу двата мерења. Иако е утврдена мала, но статистички значајна систематска разлика во висината ( $\Delta=0.10\pm 0.35$  cm;  $t=2.38$ ;  $p=.020$ ), стандардизираната големина на разликата останува ниска до умерена ( $d_z=0.302$ ), што практично сугерира дека забележаната разлика најверојатно нема функционална важност и може да се поврзе со технички варијации на мерењето и биолошка микро варијабилност, а не со реална промена во телесниот раст во краток период.

Слично стабилна слика се забележува и кај телесната тежина и индексот на телесна маса, каде релативната релијабилност е многу висока (тежина: ICC=0.996; 95% CI: 0.993–0.997; BMI: ICC=0.992; 95% CI: 0.988–0.995). Во двата случаи, нема докази за систематска разлика меѓу мерењата, што е потврдено со незначајни t тестови (тежина:  $\Delta=-0.02\pm 0.97$  kg;  $t=0.19$ ;  $p=.850$ ;  $d_z=0.024$ ; BMI:  $\Delta=-0.03\pm 0.31$  kg/m<sup>2</sup>;  $t=0.91$ ;  $p=.365$ ;  $d_z=0.116$ ). Овие наоди се важни од практичен аспект бидејќи сугерираат дека мерењата се не само статистички стабилни, туку и практично конзистентни, односно дека евентуалните мали варијации се во рамки на очекувана мерна грешка и немаат значајни импликации за интерпретација на индивидуални промени.

Индексот на телесна маса (BMI) кај момчињата покажува многу висока релативна тест-ретест релијабилност (ICC(2,1)=0.992; 95% CI: 0.988–0.995), што укажува на речиси совршена стабилност на рангирањето на испитаниците меѓу првото и второто мерење. Истовремено, не се утврдува систематска разлика меѓу T1 и T2, бидејќи просечната промена е минимална и статистички незначајна ( $\Delta=-0.03\pm 0.31$  kg/m<sup>2</sup>;  $t=0.91$ ;  $p=.365$ ), а стандардизираната големина на разликата е мала ( $d_z=0.116$ ). Овие резултати имаат

практично значење бидејќи сугерираат дека мерењето на ВМІ е не само статистички стабилно, туку и практично конзистентно, односно дека малите варијации меѓу мерењата најверојатно се во рамки на очекуваната мерна грешка и не доведуваат до суштинска промена во интерпретацијата на индивидуалниот статус во краток временски период.

Од антропометриските индикатори, обемот на половината покажува висока, но релативно пониска релијабилност споредено со висината и тежината (ICC=0.948; 95% CI: 0.922–0.966), што е очекувано имајќи предвид дека овој индикатор е посензитивен на позиционирање на лентата, респираторен циклус и моментална постурална напнатост. Сепак, не е утврдена систематска разлика меѓу мерењата ( $\Delta=-0.05\pm 1.95$  cm;  $t=0.19$ ;  $p=.846$ ;  $dz=0.025$ ), што ја засилува практичната употребливост на мерката во мониторинг и проценка, особено кога се применуваат стандардизирани протоколи.

Најизразени систематски промени се регистрираат кај моторичките тестови, каде постои значаен потенцијал за ефект на запознавање со задачата и последователно подобрување на перформансот при повторното тестирање, односно на второто мерење (T2, ретест) во однос на првото (T1). Најголемите ефекти и најсилни статистички докази се забележуваат кај подигнувањето на трупот за 30 секунди и тестот за флексибилност претклон во сед. Подигнувањето на трупот за 30 секунди покажува значајно подобрување на второто мерење ( $\Delta=1.72\pm 2.12$  повторувања;  $t=6.28$ ;  $p<.001$ ) со голем ефект на влијание ( $dz=0.811$ ), што има јасна практична релевантност бидејќи просечното зголемување е доволно големо да влијае врз интерпретацијата на индивидуалната и групната промена. Истовремено, релативната релијабилност за овој тест е умерена до висока (ICC=0.822; 95% CI: 0.534–0.913), но поширокиот интервал на доверба укажува на поголема несигурност и потенцијална варијабилност меѓу испитаниците, што дополнително ја нагласува потребата од внимателно толкување и дополнителни индикатори за апсолутна релијабилност при оценка на реалната промена.

Сличен образец се забележува и кај тестот претклонот во сед, каде на второто мерење се регистрира значајно зголемување на резултатот ( $\Delta=2.56\pm 3.72$  cm;  $t=-5.41$ ;  $p<.001$ ) со умерено голем ефект на влијание ( $dz=0.686$ ). Релативната релијабилност е висока (ICC=0.894; 95% CI: 0.744–0.945), што укажува дека и покрај систематското подобрување, рангирањето меѓу испитаниците останува во значајна мера стабилно. Практично, ова значи дека тестот е соодветен за разликување меѓу учениците, но резултатите при повторно тестирање можат да бидат под влијание на ефект на учење

или подобрена техника, па интерпретацијата на промената мора да се поврзе со индикатори како SEM и MDC.

Кај скокот од место исто така се регистрира значајно подобрување на ретестот ( $\Delta=6.83\pm 11.15$  cm;  $t=4.74$ ;  $p<.001$ ) со умерено голем ефект на влијание ( $dz=0.612$ ), што укажува дека второто мерење во просек дава подобар резултат и дека постои систематска компонента на промена. Сепак, релативната релијабилност останува висока (ICC=0.913; 95% CI: 0.811–0.953), што сугерира добра стабилноста на рангирањето и соодветност на тестот за споредување на индивидуалните разлики меѓу испитаниците. Практичната импликација е дека скокот од место е сигурен тест за проценка на експлозивна сила, но при интервенциски следења или краткорочно мониторирање, систематското подобрување на ретестот (резултатот во второто мерење во однос на првото) треба да се земе предвид како потенцијален извор на преценување на реалниот напредок.

Тестот динамометрија на дланка, независно дали се анализира поединечно за десната и левата рака или како просек, покажува висока релативна релијабилност (десна: ICC=0.934; 95% CI: 0.901–0.956; лева: ICC=0.952; 95% CI: 0.928–0.968; просек: ICC=0.961; 95% CI: 0.941–0.974). Притоа, не се утврдени значајни систематски разлики меѓу мерењата, што е конзистентно со незначајните  $p$  вредности (десна:  $\Delta=0.20\pm 2.87$  kg;  $t=0.55$ ;  $p=.587$ ;  $dz=0.069$ ; лева:  $\Delta=-0.14\pm 2.37$  kg;  $t=-0.47$ ;  $p=.639$ ;  $dz=0.060$ ; просек:  $\Delta=0.01\pm 2.10$  kg;  $t=0.11$ ;  $p=.914$ ;  $dz=0.014$ ). Овие наоди се практично значајни бидејќи укажуваат дека тестот е стабилен и дека краткорочните разлики најверојатно се случајни и минимални, што ја поддржува неговата употреба и за проценка и за мониторинг.

Конечно, тестот чунестото трчање  $4\times 10$  m покажува исклучително висока релативна релијабилност (ICC=0.994; 95% CI: 0.991–0.996), при што не е утврдена значајна систематска разлика меѓу мерењата ( $\Delta=-0.03\pm 0.33$  s;  $t=-0.62$ ;  $p=.538$ ;  $dz=0.082$ ). Иако ефект на големината е многу мала и нема статистичка значајност, практично овој наод е особено важен бидејќи укажува дека агилноста, брзината и координацијата во овој протокол е мерена со висока стабилност и без изразен ефект на учење во примерокот, што го прави тестот чувствителен и сигурен за проценка на брзинско агилни способности во училишната популација.

Сумирано, најважниот заклучок од табелата е дека повеќето тестови покажуваат висока до речиси совршена релативна релијабилност, при што антропометриските мерки и чунестото трчање се карактеризираат со минимални или непостоечки систематски разлики, додека кај дел од моторичките тестови како флексибилност, експлозивна сила

и силова издржливост на трупот постои јасна систематска компонента на подобрување на ретестот (подобар резултат во второто мерење во однос на првото). Оваа комбинација на високи ICC вредности со истовремени статистички значајни промени кај одредени тестови имплицира дека инструментите се стабилни за рангирање на испитаниците, но дека за интерпретација на реалната индивидуална промена е неопходно резултатите да се дополнат со индикатори за апсолутна релијабилност и прагови за минимално детектибилна промена.

Табела 5. Тест-ретест релијабилност (релативна релијабилност (ICC) и систематска разлика кај момчињата)

Тест/варијабла	T1 (M±SD)	T2 (M±SD)	Δ (T2-T1) (M±SD)	t	p	ICC (2,1)	95% CI (ICC)	dz
Телесна висина (cm)	178.0 ± 6.90	178.0 ± 6.91	0.10 ± 0.35	2.38	.020	0.999	0.998 - 0.999	0.302
Телесна тежина (kg)	69.2 ± 10.2	69.2 ± 10.2	-0.02 ± 0.97	0.19	.850	0.996	0.993 - 0.997	0.024
ВМІ (kg/m <sup>2</sup> )	21.8 ± 2.51	21.7 ± 2.49	-0.03 ± 0.31	0.91	.365	0.992	0.988 - 0.995	0.116
Обем на половина (cm)	77.6 ± 5.99	77.6 ± 6.01	-0.05 ± 1.95	0.19	.846	0.948	0.922 - 0.966	0.025
Претклон во сед (cm)	18.4 ± 9.50	21.0 ± 9.64	2.56 ± 3.72	5.41	< .001	0.894	0.744 - 0.945	0.686
Скок од место (cm)	198.0 ± 29.7	205.0 ± 32.1	6.83 ± 11.15	4.74	< .001	0.913	0.811 - 0.953	0.612
Динамометрија десна (kg)	46.8 ± 7.91	47.0 ± 7.84	0.20 ± 2.87	0.55	.587	0.934	0.901 - 0.956	0.069
Динамометрија лева (kg)	44.8 ± 7.73	44.6 ± 7.51	-0.14 ± 2.37	-0.47	.639	0.952	0.928 - 0.968	0.060
Динамометрија (kg)	45.8 ± 7.52	45.8 ± 7.40	0.01 ± 2.10	0.11	.914	0.961	0.941 - 0.974	0.014
Подигнување труп 30 s	26.0 ± 4.46	27.7 ± 4.31	1.72 ± 2.12	6.28	< .001	0.822	0.534 - 0.913	0.811
Чуност 4×10 m (s)	10.9 ± 0.74	10.9 ± 0.78	-0.03 ± 0.33	-0.62	.538	0.994	0.991 - 0.996	0.082

Табела 6 ја прикажува апсолутната релијабилност на МАКФИТ батеријата кај момчињата преку индикаторите за мерна грешка и праг за реална промена, што е клучно за практичната интерпретација на резултатите на индивидуално и групно ниво. За разлика од релативната релијабилност, која првенствено ја опишува стабилноста на рангирањето, апсолутната релијабилност ја квантифицира „колку“ резултатот може да варира при повторно мерење во истите мерни единици и со тоа директно ја определува чувствителноста на тестовите за детекција на реална промена. Во овој контекст, најинформативни показатели се SEM како типична мерна грешка, SEM% како нормализирана грешка за споредливост меѓу тестовите, MDC или SDD како минимална промена што мора да се надмине за да се смета дека промената е реална, а не последица на мерна варијабилност.

Највисока прецизност, односно најмала апсолутна грешка, се забележува кај телесната висина. Иако SDpooled е умерена (6.91 cm), мерната грешка е минимална (SEM=0.249 cm), а релативната грешка практично занемарлива (SEM%=0.14). Прагот за

минимално детектибилна промена е мал ( $MDC=0.690$  cm;  $nSDD\%=0.388$ ), што значи дека промена под приближно 0.7 cm кај поединец може да се припише на мерна варијабилност, додека промена над овој праг има висока веројатност да претставува реална разлика. Ова е во согласност со претходно утврдената статистички значајна, но практично мала систематска разлика во релативната анализа ( $p=.020$ ;  $dz=0.302$ ), при што апсолутните индикатори јасно покажуваат дека откриената разлика е далеку под прагот што би имал практична импликација.

Слично поволен профил на апсолутна релијабилност се забележува кај телесната тежина и BMI, кои како индикатори се широко користени и за проценка и за мониторинг. Кај телесната тежина SEM изнесува 0.681 kg, со  $SEM\%=0.98$  и  $MDC=1.888$  kg ( $nSDD\%=2.73$ ), што укажува дека индивидуална промена мора да биде приближно 1.9 kg или повеќе за да се интерпретира како реална промена надвор од мерната грешка. Кај BMI, SEM е  $0.221$  kg/m<sup>2</sup> и  $SEM\%=1.01$ , додека MDC достигнува  $0.612$  kg/m<sup>2</sup> ( $nSDD\%=2.81$ ). Овие вредности се практично значајни затоа што покажуваат дека, иако статистички нема систематска разлика меѓу мерењата (тежина  $p=.850$ ;  $dz=0.024$ ; BMI  $p=.365$ ;  $dz=0.116$ ), детекцијата на мали промени кај поединци во краток период треба да се толкува внимателно и да се спореди со MDC, особено во интервенциски студии каде се очекуваат суптилни промени.

Кога станува збор за обемот на половината, табелата сугерира дека мерната грешка и прагот за реална промена се релативно поголеми во споредба со претходните антропометриски мерки. SEM изнесува 1.369 cm,  $SEM\%=1.76$ , а MDC е 3.795 cm со  $nSDD\%=4.89$ . Практично, ова значи дека само промени од околу 3.8 cm или повеќе можат да се сметаат за сигурна реална промена кај поединец, што е важно за клинички и училишен мониторинг, бидејќи помали флукуации може да се должат на позиционирање на мерната лента, респираторни варијации и моментални постурални фактори. Затоа, во табелата со право е нагласен CV (коефициент на варијација) како дополнителен, методолошки посоодветен критериум за оваа варијабла, бидејќи CV ја изразува мерната варијабилност во релативни (процентуални) единици и овозможува да се процени дали грешката се менува со големината на резултатот (т.е. дали кај испитаниците со поголем обем на половина се очекува и апсолутно поголема разлика T1–T2). На тој начин, CV помага интерпретацијата да биде попрецизна кога постои можност за хетероскедастичност, односно кога мерната грешка не е константна кај сите нивоа на обемот на половината.

Најизразената апсолутна варијабилност се забележува кај претклонот во сед, што е типично за тестовите на флексибилност каде техниката, загреаноста и толеранцијата на истегнување имаат значајно влијание. Иако  $SD_{pooled}$  е висок (9.57 cm), SEM е 2.633 cm, а SEM% достигнува 13.38, што ја поставува оваа мерка како најмалку прецизна во батеријата во релативна смисла. MDC изнесува 7.300 cm, додека nSDD% е многу висок (37.06), што практично значи дека потребна е голема промена за да се надмине мерната грешка и варијабилноста. Ова е особено важно ако се земе предвид дека релативната анализа покажа статистички значајна систематска разлика со умерено голема големина на влијание ( $p < .001$ ;  $dz = 0.686$ ), што упатува на изразен ефект на запознавање со тестот или подобрување на техниката при второто мерење (ретест). Оттука, иако постои статистички јасно поместување на средните вредности, апсолутните индикатори сугерираат дека интерпретацијата на индивидуални промени мора да се врши со голема претпазливост, бидејќи дел од варијацијата е очекувано висока и прагот за реална промена е релативно строг.

Кај скокот од место, апсолутната релијабилност е значително подобра и има повисока практична употребливост за мониторинг, иако грешката во апсолутни единици е очекувано поголема поради самата природа на тестот и дисперзијата на резултатите. SEM изнесува 7.892 cm, SEM%=3.92, а MDC е 21.88 cm со nSDD%=10.86. Овие резултати укажуваат дека за поединец, само подобрување од околу 22 cm или повеќе може сигурно да се интерпретира како реална промена надвор од мерната грешка. Во контекст на релативната анализа, каде беше утврдена статистички значајна систематска разлика ( $p < .001$ ) со умерено голем ефект на влијание ( $dz = 0.612$ ), апсолутните параметри сугерираат дека дел од просечната промена може да биде реална на групно ниво, но за индивидуална проценка е потребен значително поголем прираст за да се исклучи мерната варијабилност. Ова ја нагласува разликата меѓу статистичката значајност на групно ниво и практичната значајност на индивидуално ниво.

Мускулната сила мерена преку динамометрија покажува умерена апсолутна грешка и релативно поволни SEM% вредности. Кај десната рака SEM е 2.023 kg со SEM%=4.32 и MDC=5.607 kg (nSDD%=11.95), додека кај левата рака SEM е 1.666 kg со SEM%=3.72 и MDC=4.618 kg (nSDD%=10.33). Кога се користи просекот, апсолутната грешка дополнително се намалува (SEM=1.472 kg; SEM%=3.21), а прагот за реална промена е најнизок во рамки на силовите индикатори (MDC=4.081 kg; nSDD%=8.91). Овој образец е практично значаен, бидејќи покажува дека пресметувањето на просечна вредност ја намалува случајната варијабилност и ја подобрува чувствителноста за

детекција на промена. Во релативната анализа, динамометријата немаше значајни систематски разлики ( $p > .05$ ;  $dz$  многу мал), што во комбинација со умерени MDC прагови ја поддржува употребата на овој тест за следење, особено кога се користи просекот од двете раце.

Кај подигнувањето на трупот за 30 секунди се забележува SEM од 1.496 повторувања и  $SEM\%=5.57$ , со  $MDC=4.147$  повторувања и  $nSDD\%=15.45$ . Овие вредности покажуваат дека индивидуална промена треба да биде околу 4 повторувања за да се смета за сигурна реална промена. Во релативната анализа беше утврдено статистички значајно подобрување со голем ефект на влијание ( $p < .001$ ;  $dz=0.811$ ), што имплицира дека на групно ниво промената е и статистички и практично релевантна. Сепак, апсолутните индикатори нагласуваат дека кај поединец, мал прираст од 1 до 2 повторувања може да не биде доволен за да се заклучи реална промена, додека прираст над 4 повторувања е многу поверодостојно да ја надмине мерната варијабилност.

На крај, тестот чунестото трчање  $4 \times 10$  m покажува релативно ниска апсолутна грешка во однос на средната вредност, што го прави тестот особено употреблив за следење на перформансите во времетраење. SEM изнесува 0.230 s,  $SEM\%=2.11$ , а MDC е 0.638 s со  $nSDD\%=5.85$ . Иако релативната анализа не покажува систематска разлика ( $p=.538$ ;  $dz=0.082$ ), апсолутните параметри прецизно дефинираат дека индивидуално подобрување мора да биде приближно 0.64 секунди или повеќе за да се смета за реално. Ова е практично важно во педагошки и тренажни услови, каде често се очекуваат мали подобрувања, а без праг за реална промена постои ризик да се преинтерпретираат случајни варијации што се во рамки на мерната грешка.

Табела 6. Апсолутна релијабилност (SEM, SEM%, MDC/SDD, nSDD, CV% кај момчињата)

Тест/варијабла	Mean (T1,T2)	SDpooled	SEM (cm)	SEM%	MDC/SDD (cm)	nSDD%	CV%	Метод (SEM/MDC или CV)
Телесна висина (cm)	178.0	6.91	0.249	0.14	0.690	0.388	0.14	SEM/MDC
Телесна тежина (kg)	69.2	10.2	0.681	0.98	1.888	2.73	0.98	SEM/MDC
BMI (kg/m <sup>2</sup> )	21.75	2.50	0.221	1.01	0.612	2.81	1.02	SEM/MDC
Обем на половина (cm)	77.6	6.00	1.369	1.76	3.795	4.89	1.76	CV
Претклон во сед (cm)	19.7	9.57	2.633	13.38	7.300	37.06	13.38	SEM/MDC
Скок од место (cm)	201.5	30.92	7.892	3.92	21.88	10.86	3.92	SEM/MDC
Динамометрија десна (kg)	46.9	7.88	2.023	4.32	5.607	11.95	4.32	SEM/MDC
Динамометрија лева (kg)	44.7	7.62	1.666	3.72	4.618	10.33	3.72	SEM/MDC
Динамометрија просек (kg)	45.8	7.46	1.472	3.21	4.081	8.91	3.21	SEM/MDC
Подигнување труп 30 s	26.85	4.39	1.496	5.57	4.147	15.45	5.58	SEM/MDC
Чунесто $4 \times 10$ m (s)	10.9	0.760	0.230	2.11	0.638	5.85	2.20	SEM/MDC

Севкупно, табелата ја потврдува високата прецизност на антропометриските мерки и временскиот тест за агилност, каде SEM% е ниска и MDC праговите се релативно мали во однос на средните вредности, што овозможува сигурна детекција на релативно мали реални промени. Наспроти тоа, тестовите кои се почувствителни на техника и моментални функционални фактори, особено претклонот во сед, покажуваат значително поголема релативна мерна грешка и високи прагови за реална промена, што ограничува дел од нивната чувствителност за индивидуален мониторинг. Во практична смисла, овие наоди упатуваат дека за интервенциски или лонгитудинални процени кај момчињата, интерпретацијата на промените мора да се темели не само на статистичката значајност на групно ниво од t тестовите и ефект големините, туку и на MDC или SDD праговите кои дефинираат кога промената е доволно голема за да биде сметана за реална и клинички или педагошки значајна.

Табела 7 ја прикажува согласноста меѓу првото и второто мерење кај момчињата преку Bland–Altman пристапот, кој обезбедува најдиректен увид во практичната заменливост на тестот во повторени услови. Во оваа рамка, клучни се две компоненти: просечната пристрасност (bias), која покажува дали постои систематско поместување меѓу T1 и T2 и границите на согласност (limits of agreement, LoA), кои ја дефинираат очекуваната варијабилност на индивидуалните разлики. Дополнително, корелацијата меѓу средната вредност и апсолутната разлика  $r(\text{MEAN}, \text{ADIFF})$  служи како проверка за хетероскедастичност, односно дали мерната грешка се зголемува со поголеми резултати, што има директни импликации за толкување и избор на метриката.

Во антропометриските варијабли се забележува најповолна слика на согласност, со минимални систематски поместувања и тесни граници на согласност. Телесната висина покажува мала негативна пристрасност (bias =  $-0.10$  cm), при што 95% интервалот на доверба за bias ( $-0.19$  до  $-0.01$ ) укажува дека поместувањето е статистички детектибилно во насока на нешто повисоки вредности на ретестот, но неговиот апсолутен износ е практично занемарлив. Уште поважно, границите на согласност се во релативно тесен опсег (Lower LoA =  $-0.79$  cm; Upper LoA =  $0.583$  cm), што значи дека кај најголемиот дел од поединците разликата меѓу мерењата се очекува да остане во рамки од приближно  $\pm 0.8$  cm. Отсуството на значајна зависност меѓу нивото на резултат и големината на апсолутната разлика ( $r = 0.149$ ;  $p = 0.249$ ) дополнително потврдува дека грешката е стабилна низ целиот опсег на висини, што ја прави мерката високо заменлива за индивидуално следење.

Сличен профил се забележува и кај телесната тежина, каде пристрасноста практично е нула ( $\text{bias} = 0.02 \text{ kg}$ ; 95% CI:  $-0.22$  до  $0.27$ ), што укажува на отсуство на систематски тренд меѓу T1 и T2. Границите на согласност (LoA) се во релативно тесен опсег ( $-1.88$  до  $1.93 \text{ kg}$ ), што практично значи дека кај поединец разликата при повторното мерење најчесто ќе биде во рамки од приближно  $\pm 2 \text{ kg}$ . Оваа варијација е очекувана со оглед на физиолошки флукуации (хидратација, гастроинтестинален статус) и стандардната мерна реалност, но сепак останува доволно контролирана за групни споредби и за мониторинг на поголеми промени. И тука нема доказ за хетероскедастичност ( $r = 0.147$ ;  $p = 0.254$ ), што значи дека мерната варијабилност не се зголемува со поголема телесна тежина.

BMI ја покажува најстабилната согласност во релативна смисла, со многу мала пристрасност ( $\text{bias} = 0.03 \text{ kg/m}^2$ ; 95% CI:  $-0.04$  до  $0.11$ ) и тесни граници на согласност ( $-0.57$  до  $0.649 \text{ kg/m}^2$ ). Практично, тоа имплицира дека, кај поединец, варијации помали од приближно  $0.6 \text{ kg/m}^2$  можат да се сметаат како очекувана мерна и биолошка варијабилност при повторно тестирање. Корелацијата MEAN–ADIFF е мала и незначајна ( $r = 0.080$ ;  $p = 0.537$ ), што ја поддржува стабилноста на грешката низ целата распределба.

Обемот на половината, иако покажува минимална пристрасност ( $\text{bias} = 0.04 \text{ cm}$ ; 95% CI:  $-0.44$  до  $0.54$ ), има пошироки граници на согласност ( $-3.77$  до  $3.874 \text{ cm}$ ), што укажува дека индивидуалната варијација може да достигне приближно  $\pm 4 \text{ cm}$ . Тоа е важен практичен наод, затоа што сугерира дека мали промени во обемот на половината, особено во кратки интервали, треба да се толкуваат претпазливо бидејќи лесно можат да паднат во рамки на очекуваната варијабилност од позиционирање на лентата, постаура и респираторни варијации. За разлика од претходните антропометриски мерки, тука се појавува статистички значајна зависност меѓу средната вредност и апсолутната разлика ( $r = 0.277$ ;  $p = 0.030$ ), што укажува на умерена хетероскедастичност: кај испитаниците со поголем обем на половината разликите меѓу мерењата имаат тенденција да бидат поголеми. Оваа карактеристика е методолошки релевантна бидејќи оправдува дополнителна проверка со релативни индикатори (на пр. CV) или евентуална трансформација при анализа на варијабилност, со цел да се обезбеди споредливост низ различни нивоа на мерената варијабла.

Кај тестовите на физички способности, Bland–Altman резултатите јасно ја издвојуваат улогата на систематски ефект на повторно мерење кај дел од задачите, како и пошироки граници на согласност во споредба со антропометријата. Најизразена

систематска пристрасност се забележува кај тестот претклонот во сед ( $\text{bias} = -2.56 \text{ cm}$ ; 95% CI:  $-3.50$  до  $-1.61$ ), што укажува дека ретестот (резултатот во второто мерење) во просек е подобар од тестот (резултатот во првото мерење), со јасна статистички конзистентна насока на промена. Практично, ова најверојатно ја рефлектира комбинацијата на ефект на запознавање со протоколот и акутни промени во толеранцијата на истегнување и загреаност. Иако границите на согласност се релативно широки ( $-9.86$  до  $4.74 \text{ cm}$ ), што значи дека кај поединци може да се јават значајни флукуации, отсуството на хетероскедастичност ( $r = 0.044$ ;  $p = 0.733$ ) сугерира дека варијабилноста е приближно константна низ целиот опсег на флексибилност. Од аспект на практична значајност, овој образец имплицира дека тестот може да биде корисен за групни споредби, но за индивидуален мониторинг на мали промени е потребна поголема внимателност, бидејќи систематското поместување и широките LoA ја намалуваат интерпретабилноста на суптилни разлики.

Тестот скокот во далечина од место исто така покажува јасно систематско поместување во насока на подобар ретест- резултат во второто мерење ( $\text{bias} = -6.83 \text{ cm}$ ; 95% CI:  $-9.71$  до  $-3.95$ ), што укажува на значаен ефект на учење или подобрена мотивација и техника при повторното изведување. Границите на согласност се многу пошироки ( $-28.71$  до  $15.04 \text{ cm}$ ), што значи дека индивидуалните разлики можат да варираат во широк распон, па затоа за практична интерпретација на индивидуална промена се потребни поголеми прагови. Иако MEAN-ADIFF корелацијата не е статистички значајна ( $r = 0.182$ ;  $p = 0.164$ ), тенденцијата е во насока на поголема варијабилност кај повисоки резултати, што методолошки заслужува внимание при интерпретација, иако со овие податоци не се потврдува како сигурна појава. Во целина, наодот упатува дека скокот е релевантен и чувствителен на тренажни ефекти, но токму поради изразениот ретест ефект и широките LoA, најсоодветен е за следење на поголеми промени и за групни процени, додека малите индивидуални варијации може да бидат во рамки на очекуваната мерна варијабилност.

Динамометријата на дланка, без разлика дали се анализира десната, левата или просечната вредност, покажува добра симетрија околу нулта пристрасност и умерено широки LoA во апсолутни единици, што е очекувано за силов тест со биолошка варијација и зависност од моментален напор. Кај десната рака  $\text{bias}$  е  $-0.20 \text{ kg}$  (95% CI:  $-0.93$  до  $0.53$ ), кај левата  $0.14 \text{ kg}$  (95% CI:  $-0.46$  до  $0.74$ ), а кај просекот  $-0.02 \text{ kg}$  (95% CI:  $-0.56$  до  $0.50$ ), што заедно укажува на отсуство на систематска разлика меѓу мерењата. Границите на согласност се приближно симетрични и релативно стабилни: за

десната рака (−5.83 до 5.44 kg), за левата (−4.505 до 4.788 kg), и за просекот (−4.143 до 4.085 kg). Практично, ова значи дека поединечни промени од неколку килограми може да се појават и без реална физиолошка промена, па затоа во мониторинг е разумно да се бара промена која го надминува типичниот распон на LoA или претходно дефинираниот MDC праг. Дополнително, корелациите MEAN–ADIFF се ниски и незначајни (десна  $r = -0.084$ ;  $p = 0.516$ ; лева  $r = 0.020$ ;  $p = 0.880$ ; просек  $r = -0.140$ ;  $p = 0.279$ ), што потврдува дека мерната варијабилност не зависи од нивото на силата, односно грешката е хомогена низ спектарот на резултати.

Подигнувањето на трупот за 30 секунди се издвојува како тест со јасна систематска пристрасност во насока на подобар ретест ( $bias = -1.72$  повторувања; 95% CI: −2.26 до −1.17). Овој наод е важен затоа што укажува на конзистентен ефект на запознавање со задачата или подобрена моторна координација и темпо при повторното изведување. Границите на согласност (−5.86 до 2.43 повторувања) покажуваат дека кај дел од испитаниците може да се јават релативно големи индивидуални разлики, што ја ограничува интерпретацијата на мали промени на индивидуално ниво. Сепак, отсуството на хетероскедастичност ( $r = -0.058$ ;  $p = 0.661$ ) сугерира дека варијабилноста е стабилна низ различни нивоа на изведба, што ја олеснува примената на униформни прагови за реална промена во практиката.

Тестот чунестото трчање  $4 \times 10$  m покажува мала пристрасност ( $bias = 0.02$  s; 95% CI: −0.06 до 0.11) и релативно тесни граници на согласност (−0.61 до 0.668 s), што, за временски тест, претставува поволен резултат за употреба во мониторинг. Практично, ова значи дека варијации од околу половина секунда може да се појават без реална промена, но истовремено тестот има доволна стабилност за да детектира значајни подобрувања кога тие се поголеми од распонот на LoA. Корелацијата MEAN–ADIFF е на граница на статистичка значајност ( $r = 0.252$ ;  $p = 0.056$ ), што сугерира можна, но недоволно потврдена тенденција кон зголемување на варијабилноста кај повисоки средни времиња. Иако со овие податоци не може да се тврди сигурна хетероскедастичност, практично е оправдано да се остане внимателен при интерпретација кај послабите изведувачи и, доколку во проширени анализи се потврди зависност, да се разгледа употребата на релативни мерки на грешка или дополнителни проверки.

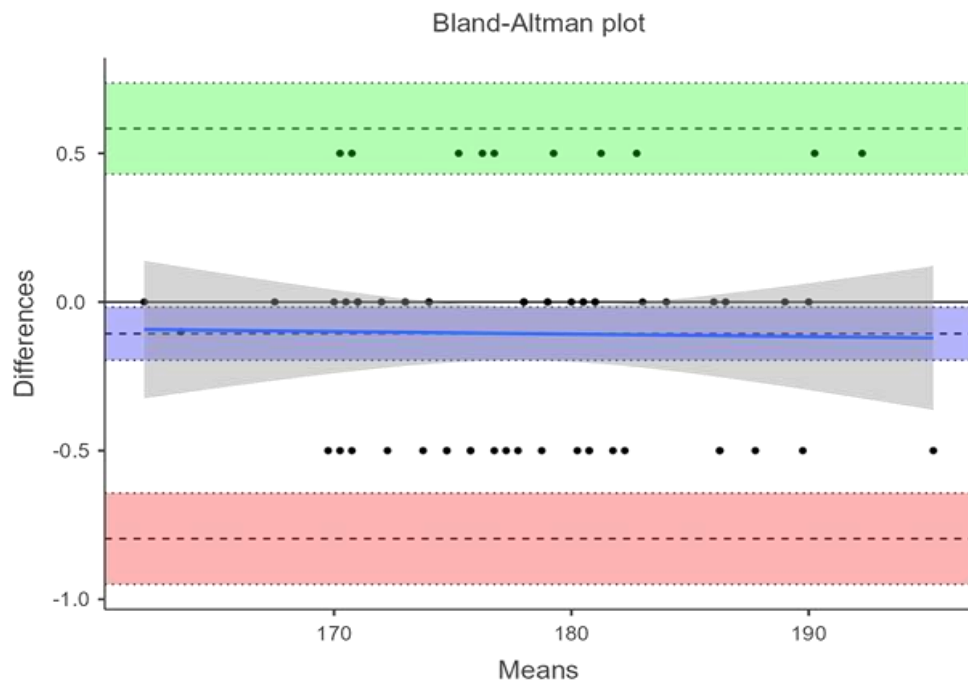
Сумирано, Bland–Altman анализата кај момчињата покажува дека антропометриските мерки и времињата во тестот за процена на агилноста, брзината и координацијата имаат висока согласност и минимална пристрасност, што ги прави

погодни за повторени мерења и за споредба низ време, особено на групно ниво. Од друга страна, тестовите кои се по осетливи на техника, мотивација и ефект на учење, како претклонот во сед, скокот од место и подигнувањето на трупот, покажуваат јасни систематски поместувања во насока на подобар ретест (подобар резултат во второто мерење), што е статистички и практично значајно за интерпретација на резултатите. Во такви случаи, статистичката значајност на bias не треба да се толкува како „проблем“ сама по себе, туку како сигнал дека во протоколот треба да се вклучи адекватна стандардизација, пробни обиди или претходно запознавање со тестот, за да се минимизира ефектот на учење и да се зголеми заменливоста на мерењата. Воедно, појавата на значајна MEAN–ADIFF зависност кај обемот на половината и граничната тенденција кај чунестото трчање укажуваат дека кај одредени варијабли, варијабилноста може да зависи од нивото на резултат, што има директна методолошка вредност при избор на индикатори и при планирање на анализа на промена во интервенциски и лонгитудинални дизајни.

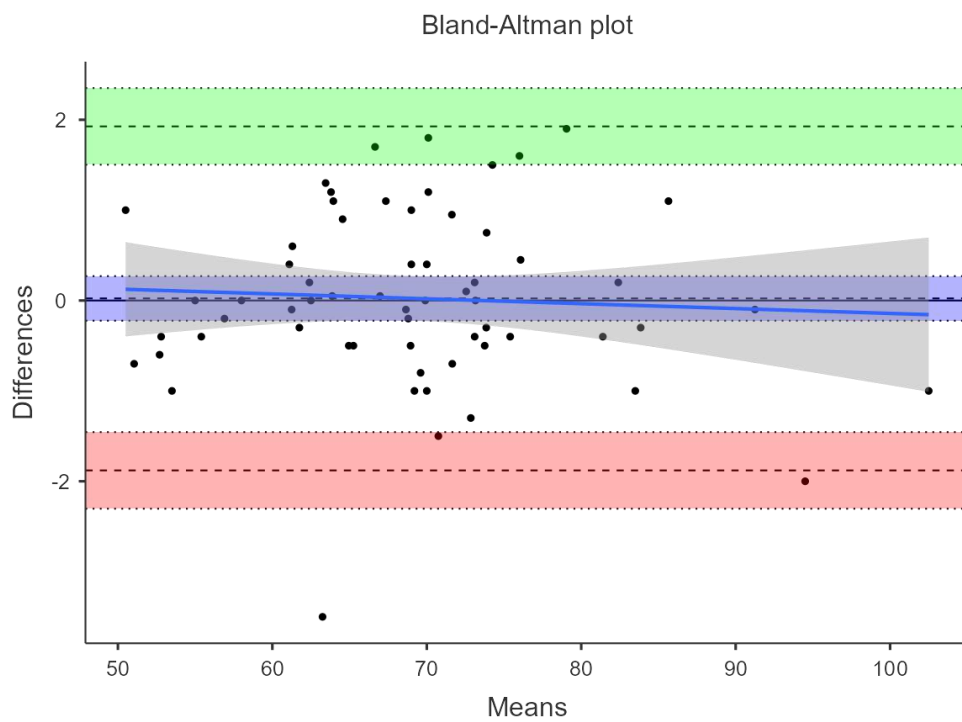
Табела 7. Bland–Altman согласност на варијаблите кај момчињата

Тест/варијабла	Bias (T1–T2) (kg)	95% CI (Bias)	Lower LoA (kg)	95% CI (Lower LoA)	Upper LoA (kg)	95% CI (Upper LoA)	r(MEAN, ADIFF)	p
Телесна висина (cm)	-0.10	-0.19 - -0.01	-0.79	-0.95 - -0.64	0.583	0.43 - 0.73	0.149	0.249
Телесна тежина (kg)	0.02	-0.22 - 0.27	-1.88	-2.30 - -1.45	1.927	1.50 - 2.35	0.147	0.254
ВМИ (kg/m <sup>2</sup> )	0.03	-0.04 - 0.11	-0.57	-0.71 - -0.44	0.649	0.51 - 0.78	0.080	0.537
Обем на половина (cm)	0.04	-0.44 - 0.54	-3.77	-4.62 - -2.92	3.874	3.02 - 4.72	0.277	0.030
Претклон во сед (cm)	-2.56	-3.50 - -1.61	-9.86	-11.48 - -8.23	4.74	3.12 - 6.37	0.044	0.733
Скок од место (cm)	-6.83	-9.71 - -3.95	-28.71	-33.66 - -23.75	15.04	10.09 - 20.00	0.182	0.164
Динамометрија десна (kg)	-0.20	-0.93 - 0.53	-5.83	-7.09 - -4.58	5.440	4.18 - 6.69	-0.084	0.516
Динамометрија лева (kg)	0.14	-0.46 - 0.74	-4.505	-5.53 - -3.40	4.788	3.75 - 5.82	0.020	0.880
Динамометрија (kg)	-0.02	-0.56 - 0.50	-4.143	-5.05 - -3.22	4.085	3.16 - 5.00	-0.140	0.279
Подигнување груп 30 s	-1.72	-2.26 - -1.17	-5.86	-6.80 - -4.92	2.43	1.49 - 3.37	-0.058	0.661
Чуесто 4×10 m (s)	0.02	-0.06 - 0.11	-0.61	-0.76 - -0.46	0.668	0.51 - 0.81	0.252	0.056

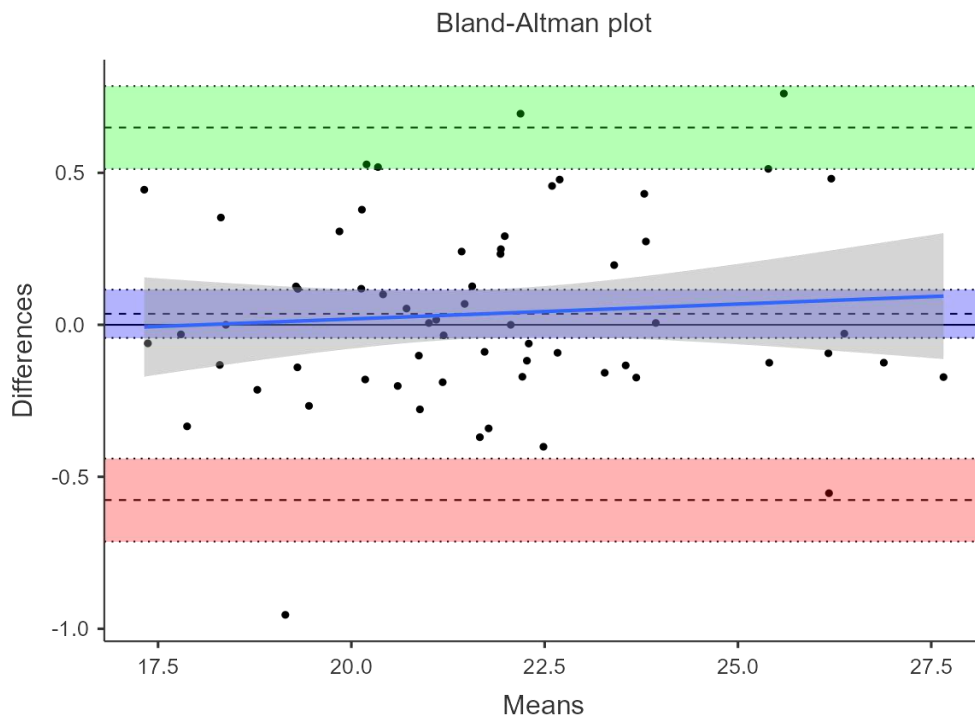
Графикон 1. Bland-Altman дијаграм на телесната висина кај момчињата



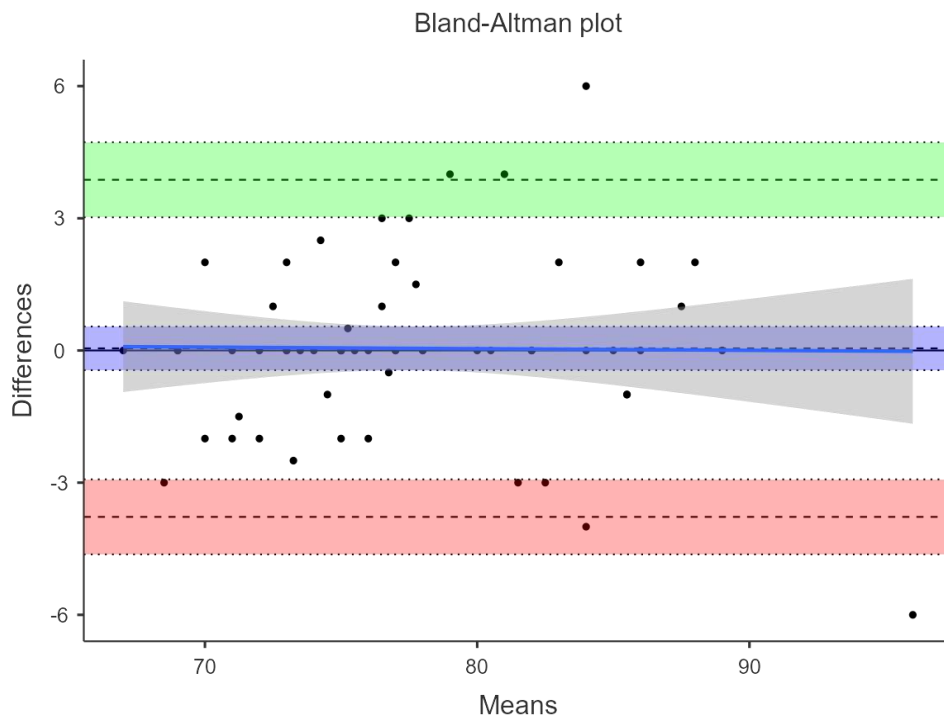
Графикон 2. Bland-Altman дијаграм на телесна тежина кај момчињата



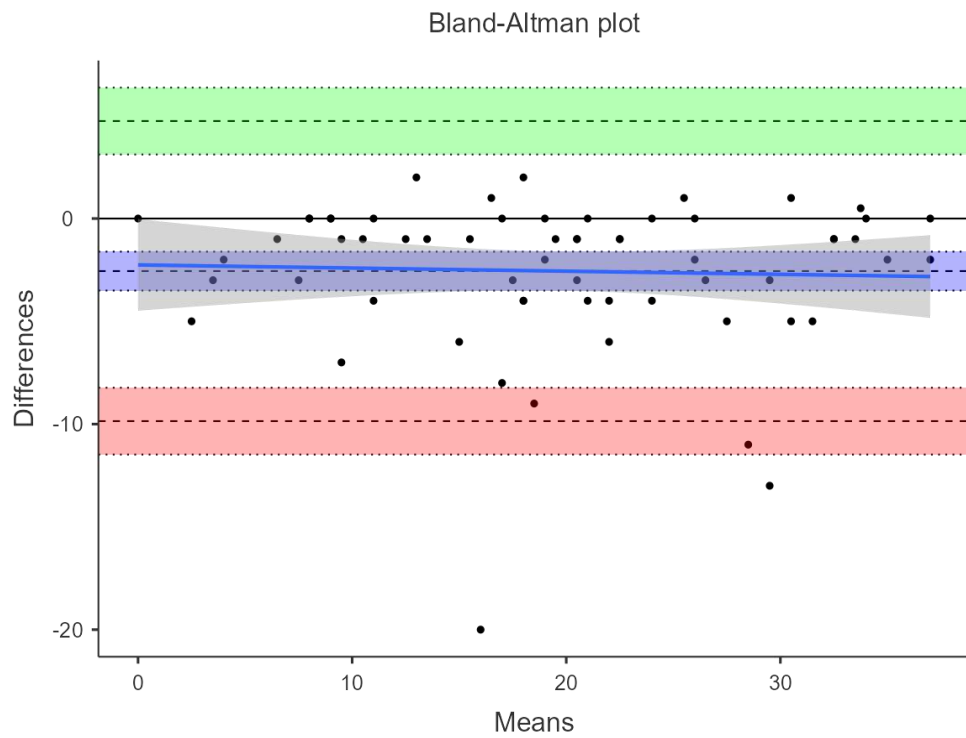
Графикон 3. Bland-Altman дијаграм на индексот на телесна маса (BMI) кај момчињата



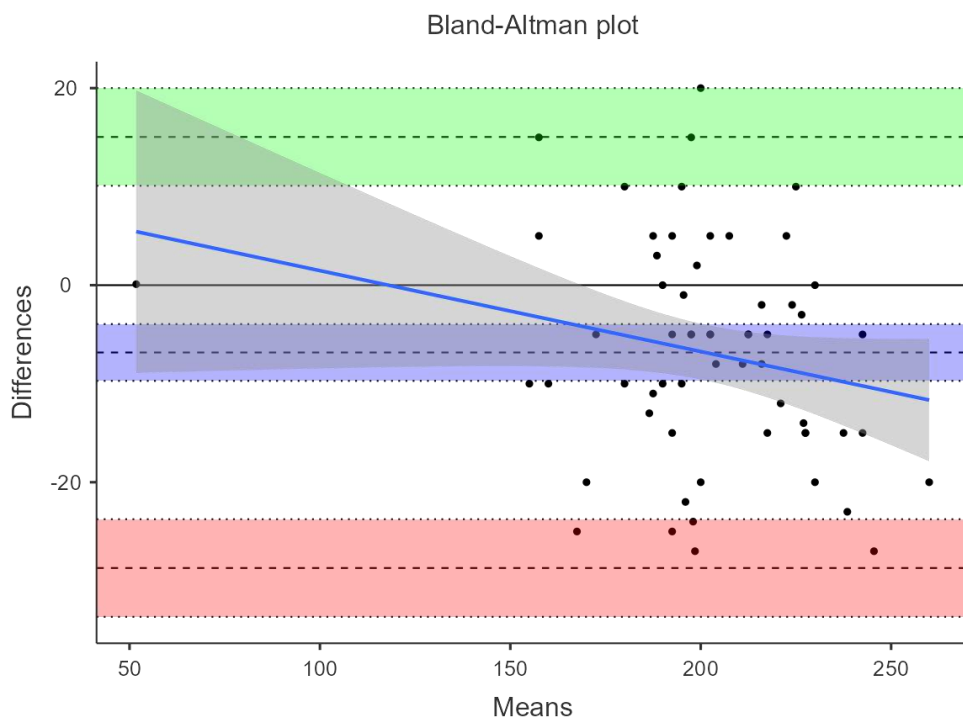
Графикон 4. Bland-Altman дијаграм на обем на половината (WC) кај момчињата



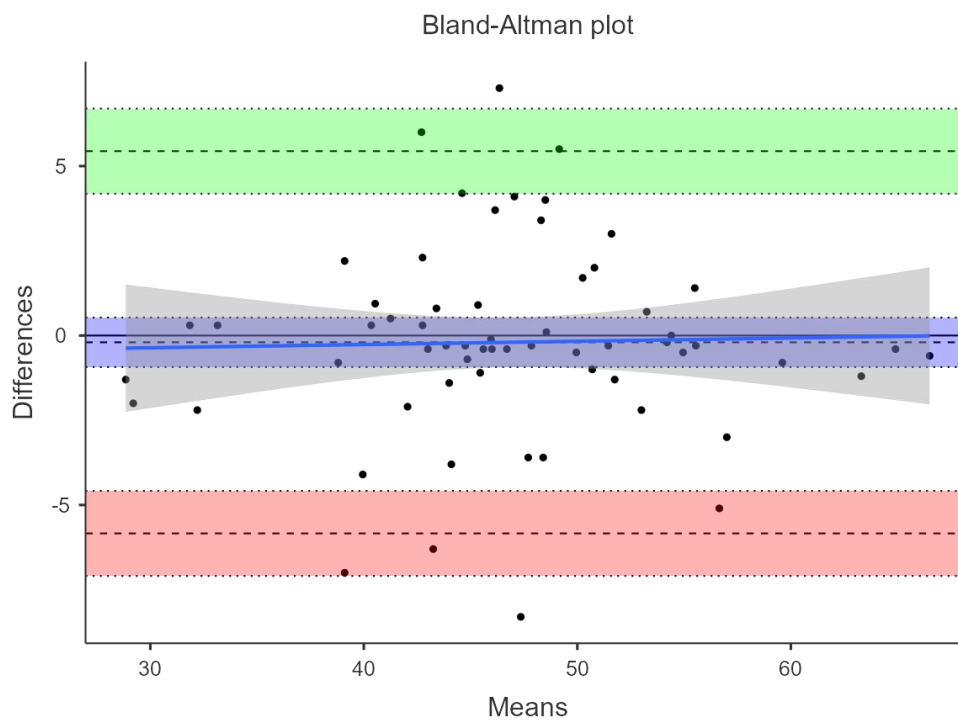
Графикон 5. Bland-Altman дијаграм на тестот претклон во сед кај момчињата



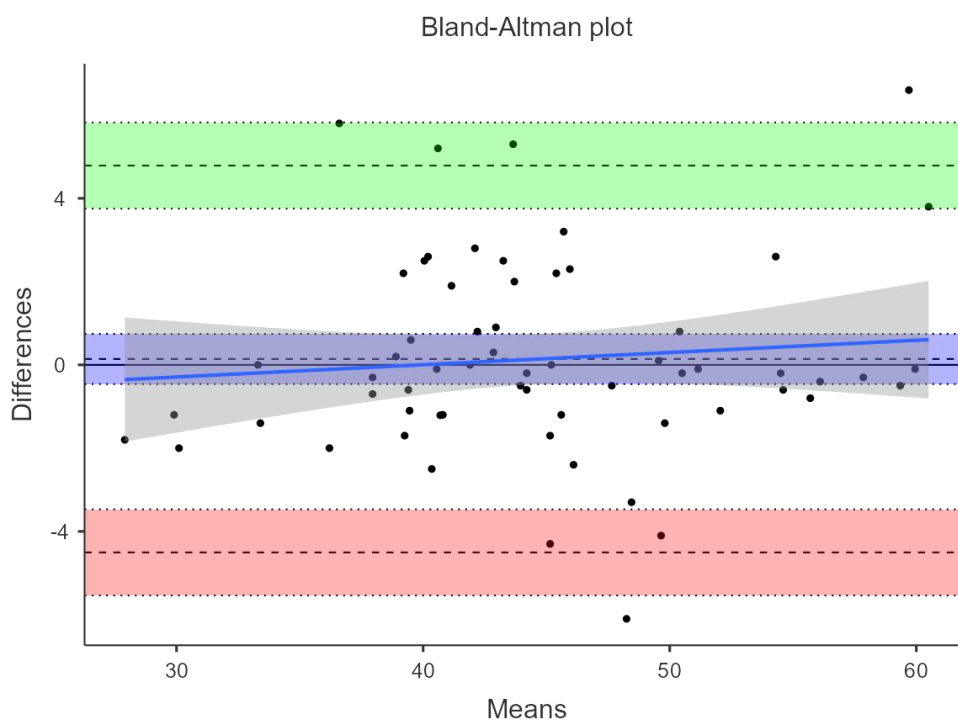
Графикон 6. Bland-Altman дијаграм на тестот скок во далечина од место кај момчињата



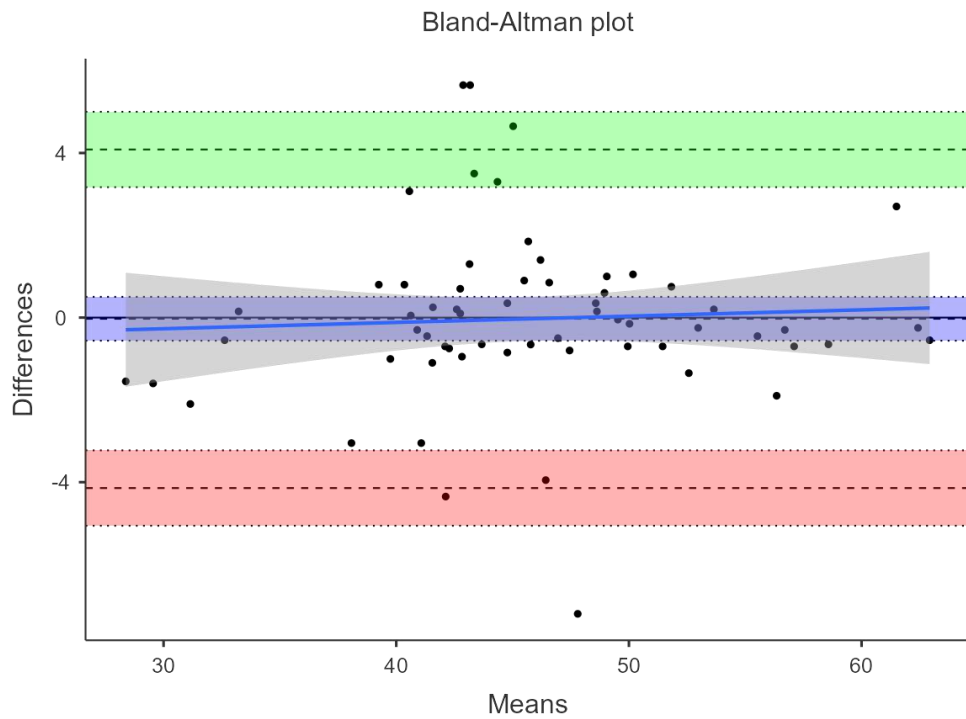
Графикон 7. Bland-Altman дијаграм на тестот динамометрија на дланка на десна рака кај момчињата



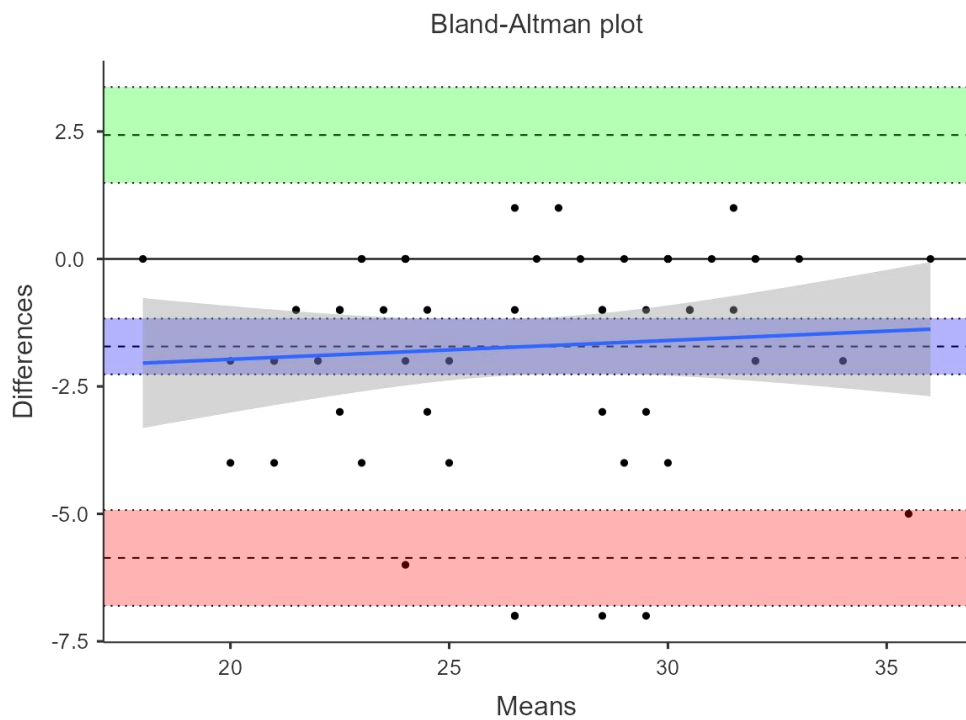
Графикон 8. Bland-Altman дијаграм на тестот динамометрија на дланка на лева рака кај момчињата



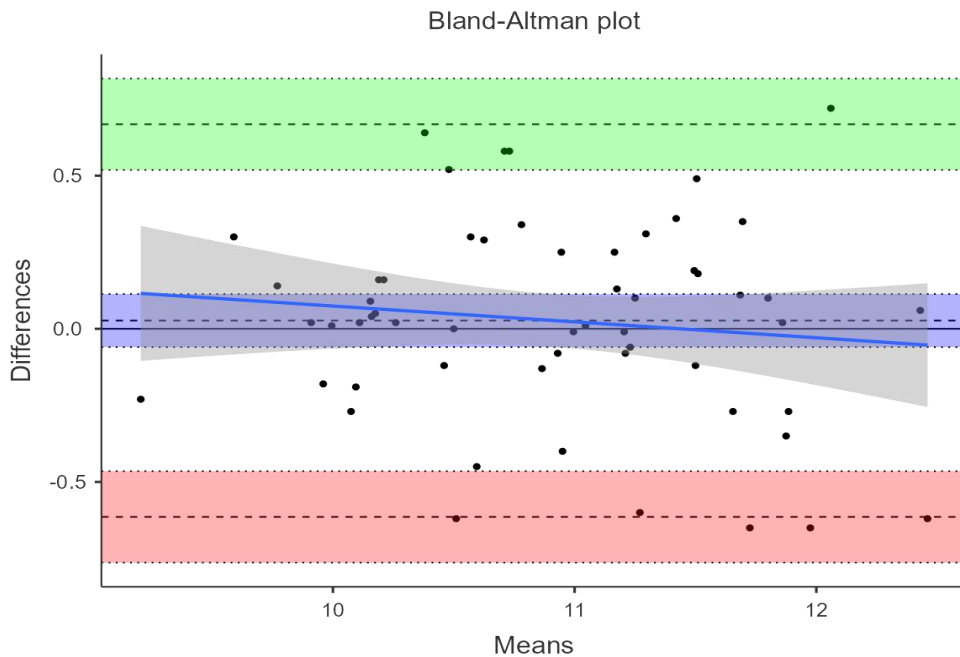
Графикон 9. Bland-Altman дијаграм на тестот динамометрија на дланка просечна вредност од десна и лева рака кај момчињата



Графикон 10. Bland-Altman дијаграм на тестот подигнување труп 30 секунди кај момчињата



Графикон 11. Bland-Altman дијаграм на тестот Чунесто трчање 4×10 m кај момчињата



Табела 8 ја прикажува тест ретест релијабилноста кај девојчињата преку комбинација од релативни индикатори (ICC со 95% интервали на доверба) и процена на систематска разлика меѓу T1 и T2 (t тест и ефект големината на влијание dz). Во целина, резултатите укажуваат на висок степен на стабилност на мерењата кај антропометриските варијабли и на дел од моторичките тестови, но истовремено откриваат дека кај одредени задачи постои изразен ретест ефект, што ја нагласува потребата јасно да се разликува статистичката од практичната значајност при толкувањето.

Највпечатливи и најрелевантни за практичната интерпретација се наодите за тестовите каде се детектира систематска промена меѓу мерењата. Подигнувањето на трупот за 30 секунди покажува најголем систематски поместувачки ефект во насока на подобар резултат на второто мерење. Просечната разлика изнесува 2.19 повторувања, при што разликата е статистички значајна ( $t = 4.33$ ,  $p < .001$ ) и придружена со многу голема ефект на големина на влијание ( $dz = 0.946$ ). Оваа комбинација на голем ефект и висок практичен импакт укажува дека повтореното тестирање ја активира компонентата на запознавање со задачата, подобрување на темпото и оптимизација на техниката, па затоа резултатот од ретестот не треба механички да се толкува како „реална“ промена во краток период без дополнителна стандардизација. Истовремено, релативната релијабилност е висока (ICC = 0.819; 95% CI: 0.401–0.929), но широкиот интервал на доверба укажува на поголема неизвесност, што е очекувано со оглед на малиот

примерок. Затоа, иако самата ICC вредност сугерира добра стабилност, заклучокот треба да се формулира претпазливо.

Сличен, иако помалку изразен, систематски ретест ефект се забележува и кај тестот претклонот во сед. Второто мерење е повисоко за 2.32 cm, разлика која е статистички значајна ( $t = 3.50$ ,  $p = .002$ ) и има голем ефект на влијание ( $dz = 0.746$ ). Практичната важност тука е јасна затоа што флексибилноста е особено чувствителна на загревање, толеранција на истегнување и искуство со протоколот, па систематското подобрување сугерира дека најверојатно дел од разликата произлегува од ефект на учење и адаптација, а не од трајна физиолошка промена. Релативната релијабилност е добра (ICC = 0.886, 95% CI: 0.668–0.952), но повторно интервалот е релативно широк, што имплицира дека прецизноста на проценката е ограничена од големината на примерокот. Во овој контекст, статистичката значајност на разликата е важна за методолошкиот заклучок дека тестот не е неутрален на повторување, додека практичната значајност налага во идни протоколи да се обезбеди стандардизиран пробен обид или претходно запознавање за да се минимизира ретест ефектот.

Кај телесната висина, иако вредностите на T1 и T2 на прв поглед се идентични на една децимала, се детектира мала но статистички значајна систематска разлика од 0.186 cm ( $t = -2.62$ ,  $p = .016$ ), со умерена ефект на влијание големина ( $dz = 0.558$ ). Ова е типичен пример каде статистичката значајност не мора да имплицира практична важност. Разликата е практично занемарлива (0.1 cm), па веројатно произлегува од мерна варијабилност (позиционирање и читање на мерката), а не од реален раст во периодот меѓу првото и второто мерење. Токму тука доаѓа до израз високата релативна релијабилност (ICC = 0.998, 95% CI: 0.994–0.999), која покажува дека рангирањето на испитаниците е речиси совршено стабилно, и дека и покрај статистички детектибилното поместување, мерката практично останува извонредно конзистентна за употреба во истражувачки и мониторинг контексти.

Кај останатите антропометриски показатели, систематските разлики не се статистички значајни и се практично занемарливи, но релативната релијабилност останува висока. Телесната тежина е речиси идентична меѓу мерењата, со занемарлива разлика ( $\Delta = 0.002$  kg), без статистичка значајност ( $t = -0.01$ ,  $p = .991$ ) и минимален ефект на влијание ( $dz = 0.002$ ), додека ICC е многу висок (0.991, 95% CI: 0.982–0.996). Ова укажува дека мерењето на тежината кај девојчињата е стабилно и дека евентуалните многу мали разлики најверојатно се должат на нормални дневни промени (хидратација, внес на храна, празнење) или на начинот на кој вагата ја заокружува вредноста, а не на

систематска грешка. ВМІ исто така не покажува значајна систематска промена ( $t = 0.62$ ,  $p = .544$ ), со мал ефект на влијание ( $dz = 0.131$ ) и висока релијабилност (ICC = 0.988, 95% CI: 0.976–0.994), што значи дека иако ВМІ е изведена мерка, неговата стабилност во повторени мерења останува многу добра. Обемот на половината не покажува значајна разлика меѓу T1 и T2 ( $t = -0.57$ ,  $p = .572$ ), со мал ефект на влијание ( $dz = 0.122$ ), но ICC е понизок во однос на висината и тежината (ICC = 0.918, 95% CI: 0.838–0.959). Практично, ова е очекувано бидејќи обемот на половината е посензитивен на позиционирање на лентата, постурални варијации и респирација; сепак, вредностите и понатаму укажуваат на добра релативна релијабилност за групни споредби и за следење на поголеми промени низ време.

Кај тестовите на сила, систематски разлики не се детектираат, а релативната релијабилност е генерално добра до многу добра. Динамометријата на десната рака покажува мала негативна разлика ( $\Delta = -0.21$  kg) без статистичка значајност ( $t = 0.42$ ,  $p = .679$ ) и мал ефект на влијание ( $dz = 0.090$ ), со добар ICC (0.923, 95% CI: 0.849–0.962). Левата рака има практично нулта разлика ( $\Delta = -0.02$  kg;  $t = 0.04$ ,  $p = .964$ ;  $dz = 0.010$ ) и дури повисок ICC (0.934, 95% CI: 0.868–0.967), што упатува на висока конзистентност во рангирањето на испитаниците. Просечната динамометрија дополнително ја стабилизира мерката преку редуција на случајната варијација, што се рефлектира во висок ICC (0.941, 95% CI: 0.883–0.971) и незначајна разлика меѓу мерењата ( $t = 0.26$ ,  $p = .793$ ;  $dz = 0.057$ ). Практично, овие наоди сугерираат дека рачната динамометрија кај девојчињата е погодна за повторени мерења, а отсуството на систематска промена значи дека тестот е релативно „неутрален“ на повторување, барем во однос на просечниот резултат.

Скокот во далечина од место претставува интересен случај бидејќи релативната релијабилност е добра (ICC = 0.884, 95% CI: 0.771–0.943), но не се забележува статистички значајна систематска разлика ( $t = -0.91$ ,  $p = .370$ ), при што ефектот на влијание е мал до умерен ( $dz = 0.200$ ). Ова може да се толкува така што девојчињата, како група, не покажуваат конзистентен ретест тренд во подобрување или влошување, но поединечните варијации и понатаму постојат, што ја одржува потребата за внимателно толкување на индивидуални промени. Во контекст на целта на тест ретест процената, оваа комбинација е позитивна за групна употреба, но за мониторинг на мали индивидуални разлики, дополнителните индикатори на апсолутна релијабилност ќе бидат од суштинско значење.

Конечно, тестот чунестото трчање 4×10 m покажува практично нулта систематска промена ( $\Delta = -0.03$  s;  $t = 0.39$ ,  $p = .695$ ) со мал ефект на влијание ( $dz = 0.089$ ), но извонредно висок ICC (0.995, 95% CI: 0.991–0.998). Ова е особено значајно затоа што брзинско агилносните тестови често се чувствителни на мотивација и варијации во стартна техника; тука, сепак, податоците укажуваат дека рангирањето на испитаниците е речиси совршено стабилно, а просечната промена е занемарлива. Практичниот заклучок е дека тестот, во оваа група, обезбедува многу конзистентни резултати за споредба меѓу испитаници и низ време, при што секоја поголема промена би можела со поголема сигурност да се припише на реална промена, а не на систематска мерна грешка.

Сумирано, кај девојчињата се издвојуваат две групи наоди. Прво, антропометриските показатели и агилноста покажуваат многу висока релативна релијабилност и минимални систематски разлики, при што единствената статистички значајна разлика кај висината е практично тривијална. Второ, тестовите кои зависат од техника, темпо и моментална подготвеност, како претклонот и подигнувањето труп, покажуваат јасен и практично релевантен ретест ефект со големи ефект големини и значајни  $p$  вредности, што треба да се земе предвид при дизајн на протокол и при интерпретација на промените. Во целина, и покрај ограничувањето од помалиот примерок, профилот на ICC вредностите сугерира дека повеќето мерки имаат добра до одлична стабилност, додека систематските разлики кај дел од моторичките тестови го нагласуваат значењето на стандардизацијата, пробните обиди и внимателното разграничување меѓу статистичка детекција и реална, практично значајна промена.

Табела 8. Тест-ретест релијабилност (релативна релијабилност (ICC) и систематска разлика кај девојчињата)

Тест/варијабла	T1 (M±SD)	T2 (M±SD)	$\Delta$ (T2-T1) (M±SD)	t	p	ICC (2,1)	95% CI (ICC)	dz
Телесна висина (cm)	164 ± 5.65	164 ± 5.60	0.18 ± 0.33	2.62	.016	0.998	0.994 - 0.999	0.558
Телесна тежина (kg)	58.2 ± 6.69	58.2 ± 7.03	0.00 ± 0.93	0.01	.991	0.991	0.982 - 0.996	0.002
BMI (kg/m <sup>2</sup> )	21.7 ± 2.41	21.7 ± 2.52	-0.05 ± 0.38	-0.62	.544	0.988	0.976 - 0.994	0.131
Обем на половина (cm)	71.5 ± 5.17	71.7 ± 5.65	0.27 ± 2.22	0.57	.572	0.918	0.838 - 0.959	0.122
Претклон во сед (cm)	19.2 ± 8.02	21.5 ± 7.68	2.32 ± 3.10	3.50	.002	0.886	0.668 - 0.952	0.746
Скок од место (cm)	151 ± 28.0	154 ± 25.9	2.62 ± 13.06	0.91	.370	0.884	0.771 - 0.943	0.200
Динамометрија десна (kg)	29.1 ± 5.63	28.9 ± 6.09	-0.21 ± 2.34	-0.42	.679	0.923	0.849 - 0.962	0.090
Динамометрија лева (kg)	28.0 ± 5.90	27.9 ± 6.52	-0.02 ± 2.32	-0.04	.964	0.934	0.868 - 0.967	0.010
Динамометрија (kg)	28.5 ± 5.59	28.4 ± 6.08	-0.11 ± 2.04	-0.26	.793	0.941	0.883 - 0.971	0.057
Подигнување труп 30 s	20.6 ± 4.46	22.8 ± 5.51	2.19 ± 2.31	4.33	<.001	0.819	0.401 - 0.929	0.946
Чунесто 4×10 m (s)	12.8 ± 1.32	12.8 ± 1.37	-0.03 ± 0.38	-0.39	.695	0.995	0.991 - 0.998	0.089

Табела 9 ја прикажува апсолутната релијабилност кај девојчињата преку повеќе индикатори на мерна грешка и прагови на детекција на реална промена. За разлика од релативната релијабилност која ја опишува стабилноста на рангирањето меѓу испитаниците, апсолутните показатели директно одговараат на прашањето колкава варијација во единиците на мерката се очекува само поради мерење и повторување. Во таа смисла, SEM ја претставува типичната грешка на мерење изразена во истите единици како и варијаблата, SEM% ја прикажува таа грешка релативно во однос на просечната вредност, додека MDC (или SDD) го дефинира минималниот праг над кој индивидуалната промена може да се смета за реална (вистинска) со висока веројатност, а не само мала варијација во рамки на мерната несигурност. CV% дополнително ја сумира варијабилноста во релативни термини и е особено корисен кога грешката е пропорционална на големината на резултатот (хетероскедастичност) или кога практичното толкување најмногу се темели на проценти.

Најпрецизни, со најмала апсолутна и релативна грешка, се утврдени кај антропометриските индикатори, особено телесната висина. Кај висината SEM е 0.236 cm и SEM% 0.14%, а MDC изнесува 0.655 cm, што имплицира дека индивидуалната промена е помала од приближно две третини на сантиметар не може сигурно да се разликува од мерна грешка. Иако во тест ретест споредбата кај девојчињата е регистрирана статистички значајна систематска разлика ( $t = -2.62$ ,  $p = .016$ ,  $dz = 0.558$ ), таа просечна разлика е од ред на 0.186 cm и останува јасно под прагот на MDC, па практичната импликација е дека станува збор за минимално поместување без реално значење за интерпретација на раст, најверојатно поврзано со позиционирање или микроваријации при мерењето, а не со вистинска промена на телесната димензија. Слично, телесната тежина и BMI покажуваат ниска мерна грешка: за тежината SEM е 0.647 kg (SEM% 1.11%) со MDC 1.794 kg, а за BMI SEM е 0.267 kg/m<sup>2</sup> (SEM% 1.23%) со MDC 0.739 kg/m<sup>2</sup>. Ова значи дека за индивидуален мониторинг на краток рок, само промени поголеми од овие прагови би биле интерпретирани како сигурни, додека помалите промени, иако можат да бидат статистички видливи во големи примероци, на индивидуално ниво остануваат во рамките на очекуваната мерна несигурност. Во согласност со тоа, систематски разлики за тежина и BMI не се детектираат (тежина:  $t = -0.01$ ,  $p = .991$ ,  $dz = 0.002$ ; BMI:  $t = 0.62$ ,  $p = .544$ ,  $dz = 0.131$ ), што ја зацврстува практичната проценка дека овие мерки се стабилни и погодни за повторени мерења.

Во средната зона на апсолутна грешка се наоѓа обемот на половината, каде SEM е 1.552 cm и SEM% 2.17%, а MDC 4.302 cm. Овие вредности се очекувано повисоки од

висината и тежината, поради поголемата чувствителност на постурални варијации, респирација и точна локација на лентата. Практично, тоа значи дека индивидуална промена од два до три сантиметри може да изгледа „видлива“, но сепак да не го надмине прагот за сигурна детекција на реална промена. Воедно, отсуството на статистички значајна систематска разлика кај девојчињата ( $t = -0.57$ ,  $p = .572$ ,  $dz = 0.122$ ) сугерира дека во просек нема конзистентно поместување меѓу T1 и T2, туку доминира случајната компонента на мерната варијабилност, што дополнително ја нагласува важноста на MDC при индивидуална интерпретација.

Најголемите вредности на SEM%, CV% и nSDD% се регистрираат кај тестовите кои зависат од техника, толеранција на истегнување, мотивација и моментална подготвеност, каде мерната грешка и варијабилноста по природа се повисоки. Тестот претклонот во сед има SEM 2.197 cm и SEM% 10.79%, со MDC 6.090 cm и nSDD 29.93%. Овие индикатори јасно покажуваат дека, иако тестот може да биде корисен за групни споредби, индивидуалните промени мора да бидат релативно големи за да се интерпретираат како сигурни. Во контекст на детектираната систематска разлика во табелата за релативна релијабилност ( $t = -3.50$ ,  $p = .002$ ,  $dz = 0.746$ ), просечното подобрување од околу 2.32 cm е статистички и методолошки важно бидејќи укажува на ретест ефект, но практично останува под MDC, што значи дека на индивидуално ниво најголем дел од испитаниците веројатно нема да ја надминат границата на сигурна „вистинска“ промена без значително поголемо подобрување. Ова е типичен пример каде статистичката значајност сигнализира систематичност на ефектот, но апсолутните индекси предупредуваат дека реалната детекција на промена кај поединец бара праг кој е повисок од просечната разлика.

Сличен образец се гледа и кај подигнувањето на трупот за 30 секунди, каде SEM е 1.637 повторувања, SEM% 7.55% и nSDD 20.92%, а методот во табелата е означен преку CV, со CV% 7.56, што имплицира дека процентуалната варијабилност е особено информативна за оваа варијабла. Практично, MDC од 4.538 повторувања укажува дека индивидуално подобрување помало од приближно 4 до 5 повторувања може да биде во рамки на мерната несигурност. Истовремено, тест ретест разликата е статистички многу значајна и со многу голем ефект на влијание ( $t = -4.33$ ,  $p < .001$ ,  $dz = 0.946$ ), што ја нагласува постоењето на силен ефект на запознавање со задачата и оптимизација на темпо. Оттука, на групно ниво ретест ефектот е очигледен и практично релевантен за интерпретација на резултатите, додека за индивидуални заклучоци е неопходно

подобрувањето да го надмине MDC за да се тврди дека е надвор од очекуваната мерна варијација.

Кај тестот скокот во далечина од место, апсолутната грешка е релативно поголема во споредба со антропометриските мерки и брзинскиот тест, што е типично за експлозивни задачи каде техниката и моменталната изведба варираат. SEM е 9.210 cm, SEM% 6.04%, а MDC 25.53 cm со nSDD 16.74%. Практичната импликација е строга: индивидуална промена од 10 до 15 cm може да изгледа значајна во практиката, но според MDC таа сè уште може да биде недоволна за сигурна детекција на „вистинска“ промена. Ова се вклопува со фактот дека систематска разлика не е детектирана ( $t = -0.91$ ,  $p = .370$ ,  $dz = 0.200$ ), што сугерира дека доминира варијабилноста на изведбата, а не конзистентен тренд на подобрување на ретест.

Кај динамометријата, апсолутната релијабилност е умерено добра, со SEM околу 1.4 до 1.6 kg и SEM% околу 5%, што е очекувано за мерки на максимална волева сила. Десната рака има SEM 1.620 kg и SEM% 5.59% со MDC 4.490 kg, левата SEM 1.602 kg и SEM% 5.73% со MDC 4.438 kg, а просечната вредност покажува нешто пониска релативна грешка (SEM 1.415 kg, SEM% 4.97%) и понизок MDC од 3.922 kg, што ја потврдува практичната предност на агрегирање на мерења за редукција на случајната варијација. Во согласност со тоа, тест ретест анализите не покажуваат значајни систематски разлики и ефектите на влијание се многу мали (десна:  $t = 0.42$ ,  $p = .679$ ,  $dz = 0.090$ ; лева:  $t = 0.04$ ,  $p = .964$ ,  $dz = 0.010$ ; просек:  $t = 0.26$ ,  $p = .793$ ,  $dz = 0.057$ ), што имплицира дека грешката е претежно случајна, а не систематска. Практично, за индивидуална интерпретација на напредок кај девојчињата, само промени поголеми од приближно 4 до 4.5 kg за поединечна рака или околу 4 kg за просечната вредност би биле сигурен индикатор за реална промена.

Најстабилна меѓу моторичките мерки во апсолутна смисла, со најниски релативни грешки, е чунестото трчање  $4 \times 10$  m. SEM е 0.269 s, SEM% 2.10%, CV% 2.20 и MDC 0.745 s, со nSDD 5.82%. Ова значи дека индивидуални промени од неколку стотинки или една две десетинки од секундата не треба да се преценуваат, додека подобрување поголемо од приближно 0.75 s може да се интерпретира како сигурна промена над мерната несигурност. Отсуството на статистички значајна систематска разлика ( $t = 0.39$ ,  $p = .695$ ,  $dz = 0.089$ ) дополнително сугерира дека тестот не продуцира конзистентен ретест тренд, туку дава стабилни вредности на групно ниво, при што апсолутните индекси обезбедуваат јасни прагови за индивидуална проценка.

Сумирано, табелата покажува јасна хиерархија на апсолутната релијабилност кај девојчињата. Антропометриските мерки имаат најмала грешка и најниски MDC прагови, што ги прави високо погодни за мониторинг, при што статистички значајните поместувања, како кај висината, остануваат практично тривијални кога се споредат со MDC. Наспроти тоа, во тестовите претклон во сед и подигнувањето труп за 30 секунди покажуваат најголема релативна варијабилност и највисоки прагови за сигурна детекција на индивидуална промена, а истовремено и значајни ретест ефекти со големи ефекти на влијание, што методолошки упатува на потреба од стандардизирано запознавање и контрола на условите за тестирање. Останатите моторички тестови, како скокот во далечина од место и динамометријата на дланка, имаат умерена апсолутна грешка и релативно високи MDC прагови, па нивната интерпретација на индивидуално ниво треба секогаш да се врзува за тоа дали забележаната промена го надминува MDC, а не само дали е „видлива“ или статистички значајна во групна анализа

Табела 9. Апсолутна релијабилност (SEM, SEM%, MDC/SDD, nSDD, CV% кај девојчињата)

Тест/варијабла	Mean (T1,T2)	SDpooled	SEM (cm)	SEM%	MDC/SDD (cm)	nSDD%	CV%	Метод (SEM/MDC или CV)
Телесна висина (cm)	164.0	5.63	0.236	0.14	0.655	0.40	0.14	SEM/MDC
Телесна тежина (kg)	58.2	6.86	0.647	1.11	1.794	3.08	1.11	SEM/MDC
BMI (kg/m <sup>2</sup> )	21.7	2.47	0.267	1.23	0.739	3.41	1.23	SEM/MDC
Обем на половина (cm)	71.6	5.42	1.552	2.17	4.302	6.01	2.17	SEM/MDC
Претклон во сед (cm)	20.35	7.85	2.197	10.79	6.090	29.93	10.78	SEM/MDC
Скок од место (cm)	152.5	26.97	9.210	6.04	25.53	16.74	6.05	SEM/MDC
Динамометрија десна (kg)	29.0	5.86	1.620	5.59	4.490	15.48	5.59	SEM/MDC
Динамометрија лева (kg)	27.95	6.22	1.602	5.73	4.438	15.88	5.73	SEM/MDC
Динамометрија (kg)	28.45	5.84	1.415	4.97	3.922	13.79	4.97	SEM/MDC
Подигнување труп 30 s	21.7	5.01	1.637	7.55	4.538	20.92	7.56	CV
Чуносто 4×10 m (s)	12.8	1.35	0.269	2.10	0.745	5.82	2.20	SEM/MDC

Табела 10 ја прикажува Bland–Altman согласноста кај девојчињата и овозможува директна процена на тоа дали двата наврати на мерење се доволно усогласени на индивидуално ниво. За разлика од корелациските и ICC индикаторите, кои главно ја опишуваат стабилноста на рангирањето меѓу испитаниците, Bland–Altman пристапот ја квантифицира систематската пристрасност (bias) меѓу T1 и T2 и ширината на границите на согласност (limits of agreement, LoA), односно интервалот во кој реално ќе се движи разликата кај најголем дел од поединците. Оттаму, практичната вредност на оваа табела

е во тоа што покажува колкави индивидуални отстапувања се очекувани „нормално“ при повторено мерење, дури и кога просечната разлика е мала.

Најконзистентна согласност, со минимална пристрасност и тесни LoA, се забележува кај антропометриските индикатори на телесна големина, особено кај телесната висина. Пристрасноста е негативна и статистички јасно дефинирана, со bias =  $-0.18$  cm и 95% CI од  $-0.33$  до  $-0.03$ , што укажува дека T1, во просек, е малку повисок од T2. Сепак, апсолутната големина на оваа разлика е многу мала и практично занемарлива во контекст на антропометрија, а и границите на согласност се тесни (Lower LoA =  $-0.841$  cm; Upper LoA =  $0.469$  cm). Ова значи дека, на индивидуално ниво, разликите меѓу двете мерења ретко би надминале приближно 1 cm во негативна насока или половина сантиметар во позитивна насока, што претставува одлична повторливост. Дополнително, не е евидентна пропорционална грешка, бидејќи корелацијата меѓу средната вредност и апсолутната разлика е негативна но статистички несигурна ( $r = -0.364$ ,  $p = 0.096$ ), што сугерира дека евентуалната тенденција за поголеми отстапувања кај повисоки/пониски девојчиња не е доволно стабилна за да се смета за систематска.

Телесната тежина покажува уште поочигледно отсуство на систематска пристрасност, со bias практично нула ( $-0.00$  kg) и широк интервал на доверба кој симетрично ја опфаќа нулата (95% CI:  $-0.41$  до  $0.41$ ). Иако LoA се пошироки отколку кај висината, што е очекувано за маса поради биолошки и методски флукутации, нивниот распон ( $-1.839$  до  $1.834$  kg) укажува дека индивидуалните разлики од редот на околу  $\pm 2$  kg може да се појават без да имплицираат реална промена. Отсуството на пропорционална грешка е потврдено и со многу ниска и статистички незначајна корелација ( $r = -0.080$ ,  $p = 0.724$ ). Сличен профил се гледа и кај BMI, каде bias е мал и позитивен ( $0.05$  kg/m<sup>2</sup>) со 95% CI од  $-0.11$  до  $0.22$ , што повторно сугерира дека нема конзистентно поместување меѓу мерењата. Границите на согласност ( $-0.700$  до  $0.801$  kg/m<sup>2</sup>) се релативно тесни за оваа мера и го поддржуваат заклучокот дека BMI може сигурно да се користи за повторени процени, при што корелацијата MEAN-ADIFF останува незначајна ( $r = 0.106$ ,  $p = 0.640$ ), без индикација за пропорционална варијабилност.

Обемот на половината покажува умерена пристрасност во негативна насока (bias =  $-0.27$  cm), но со интервал на доверба кој јасно ја вклучува нулата (95% CI:  $-1.26$  до  $0.71$ ), што значи дека на групно ниво не може да се тврди постоење стабилна систематска разлика. Сепак, практично најважен аспект тука е ширината на LoA, кои се значително пошироки (Lower LoA =  $-4.642$  cm; Upper LoA =  $4.097$  cm). Ова упатува дека кај

поединец, разликите меѓу две мерења на обем можат реално да достигнат околу  $\pm 4$  cm и тоа без вистинска промена во состојбата, што е методолошки очекувано поради варијации во позиционирање на лентата, респирација и постурални промени. Корелацијата  $r(\text{MEAN}, \text{ADIFF})$  е негативна и блиску до статистичка значајност ( $r = -0.379$ ,  $p = 0.082$ ), што сугерира можен тренд на поголеми отстапувања кај повисоки средни вредности, но ова останува индикација, не и доказ за пропорционална грешка, особено со оглед на малата големина на примерокот.

Кај тестовите кои зависат од техника, толеранција и мотивација, Bland–Altman профилот ја нагласува комбинацијата од систематски ретест ефект и широка индивидуална варијабилност. Претклонот во сед покажува јасна негативна пристрасност ( $\text{bias} = -2.32$  cm) со 95% CI од  $-3.70$  до  $-0.94$ , што значи дека во второто мерење резултатите, во просек, се повисоки од првото, односно постои изразен ефект на запознавање и подобрување при повторување. Практично, ова е важен наод бидејќи укажува дека еднакви услови на тестирање не гарантираат идентична изведба ако испитаничките немаат доволно претходна проба. LoA се широки ( $-8.41$  до  $3.77$  cm), што имплицира дека кај дел од поединците може да се појават значителни негативни разлики, односно многу подобри резултати на T2, но и дека кај некои може да се јави и обратно, иако во помала мера. Во исто време, не се детектира пропорционална грешка ( $r = -0.021$ ,  $p = 0.926$ ), што значи дека големината на отстапувањето не зависи систематски од тоа дали девојчето има повисока или пониска флексибилност, туку се должи на индивидуални флукуации и процесот на учење.

Тестот скок во далечина од место има мала просечна пристрасност ( $\text{bias} = -2.62$  cm) со интервал на доверба кој ја опфаќа нулата ( $-8.57$  до  $3.33$ ), што укажува дека на групно ниво нема стабилен систематски ретест ефект. Сепак, LoA се многу широки ( $-28.25$  до  $23.01$  cm), што има силна практична импликација: на индивидуално ниво, разликите меѓу два наврати може да бидат од редот на две до три десетици сантиметри, што ја рефлектира варијабилноста на експлозивната изведба и техничката конзистентност. Отсуството на пропорционална грешка ( $r = 0.093$ ,  $p = 0.687$ ) сугерира дека оваа варијабилност е приближно слична низ цел спектар на перформанс, но сепак останува методолошки предизвик за индивидуално следење на промени, бидејќи умерено подобрување може да се „изгуби“ во рамките на очекуваните LoA.

Кај динамометријата, просечните пристрасности се мали и клинички/практично скромни, без индикации за систематско поместување. За десната рака  $\text{bias}$  е  $0.21$  kg со 95% CI од  $-0.826$  до  $1.24$ , за левата  $\text{bias}$  е  $0.02$  kg ( $-1.01$  до  $1.05$ ), а за просечната вредност

bias е 0.12 kg (−0.791 до 1.02), при што во сите случаи интервалите ја вклучуваат нулата. LoA се релативно широки, но типични за максимална волева сила: околу  $\pm 4.4$  до  $\pm 4.8$  kg. Ова значи дека кај поединечна испитаничка, разлика од 2–3 kg меѓу два наврати не мора да претставува реален напредок или пад, туку може да биде резултат на дневна варијација, мотивација и техника на фаќање. Исто така, корелациите меѓу средната вредност и апсолутната разлика се незначајни (десна:  $r = 0.076$ ,  $p = 0.738$ ; лева:  $r = -0.025$ ,  $p = 0.910$ ; просек:  $r = -0.017$ ,  $p = 0.941$ ), што сугерира дека варијабилноста не е систематски поголема кај посилните или послабите девојчиња.

Најметодолошки индикативниот наод во табелата се однесува на подигнувањето на трупот за 30 секунди. Пристрасноста е изразено негативна и статистички јасна (bias = −2.19 повторувања; 95% CI: −3.24 до −1.14), што значи дека второто мерење е, во просек, повисоко и дека постои силен ретест ефект. LoA (−6.73 до 2.35) дополнително покажуваат дека кај дел од девојчињата подобрувањето може да биде големо, додека кај мал дел може да се појави и намалување, но во помал опсег. Особено важно е што овде е детектирана статистички значајна позитивна поврзаност меѓу средната вредност и апсолутната разлика ( $r = 0.466$ ,  $p = 0.033$ ), што претставува сигнал за можен „пропорционален“ аспект на грешката или хетероскедастичност: девојчињата со повисоки просечни перформанси имаат тенденција да покажуваат поголеми апсолутни разлики меѓу повторните мерења. Практично, ова значи дека интерпретацијата на промените во овој тест треба да биде повнимателна, бидејќи варијабилноста не е еднаква низ целиот перформансен спектар, а ретест ефектот може да биде поизразен кај оние со повисок капацитет или подобра способност за оптимизација на темпо при повторување. Оттука, иако на групно ниво резултатите јасно укажуваат на подобрување, за индивидуална проценка е неопходно да се земат предвид LoA и можната зависност на грешката од нивото на перформанс.

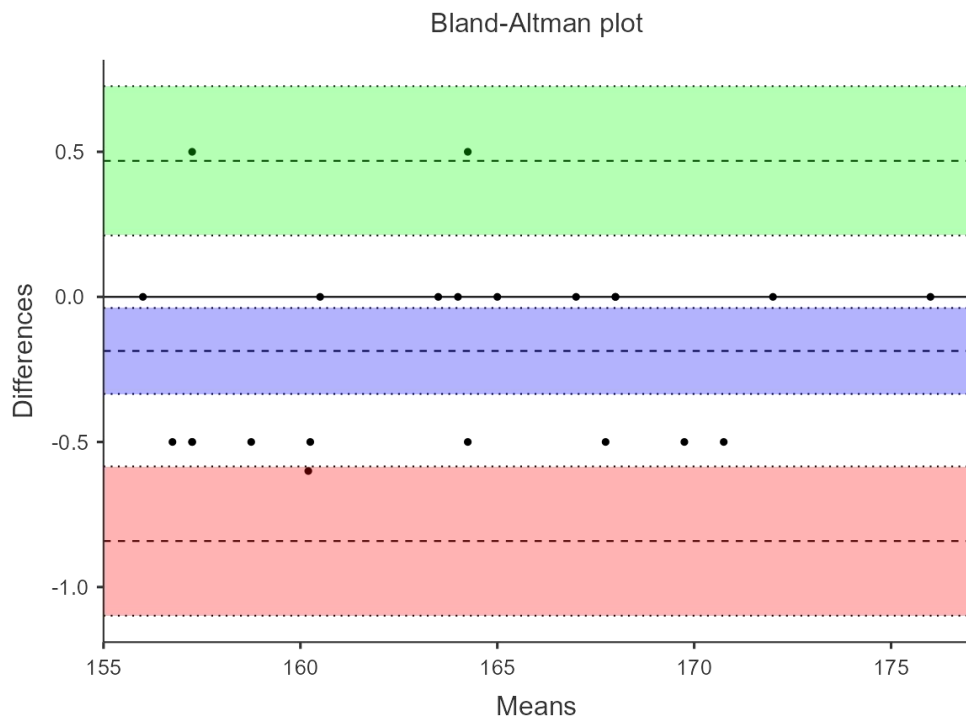
Конечно, тестотот чунестото трчање 4×10 m покажува мала пристрасност и релативно тесни LoA во споредба со повеќето моторички тестови. Bias е 0.03 s со 95% CI од −0.14 до 0.21, што сугерира отсуство на систематска разлика. LoA (−0.726 до 0.795 s) покажуваат дека индивидуалните отстапувања се најчесто под една секунда, што е методолошки поволно за повторени мерења. Корелацијата меѓу средната вредност и апсолутната разлика е мала и незначајна ( $r = 0.164$ ,  $p = 0.488$ ), без доказ за пропорционална грешка, што ја потврдува стабилноста на тестот во различни нивоа на агилност.

Севкупно, Bland–Altman резултатите кај девојчињата ја потврдуваат одличната согласност кај телесната висина, тежина и ВМІ, каде пристрасностите се минимални, а LoA се релативно тесни и практично прифатливи. Кај обемот на половината, иако нема јасна систематска пристрасност, ширината на LoA упатува на значајна индивидуална варијабилност која треба да се има предвид при мониторинг. Најизразени методолошки ефекти се забележуваат кај тестот претклонот во сед и особено кај подигнувањето на трупот за 30 секунди, каде постојат јасни ретест подобрувања, а кај вториот тест дополнително се сугерира и зависност на грешката од нивото на перформанс. Скокот од место и динамометријата покажуваат мала пристрасност, но пошироки граници на согласност, што ја нагласува потребата индивидуалните промени да се интерпретираат внимателно и секогаш во контекст на очекуваните LoA, а не само врз основа на просечни разлики или корелациски индикатори.

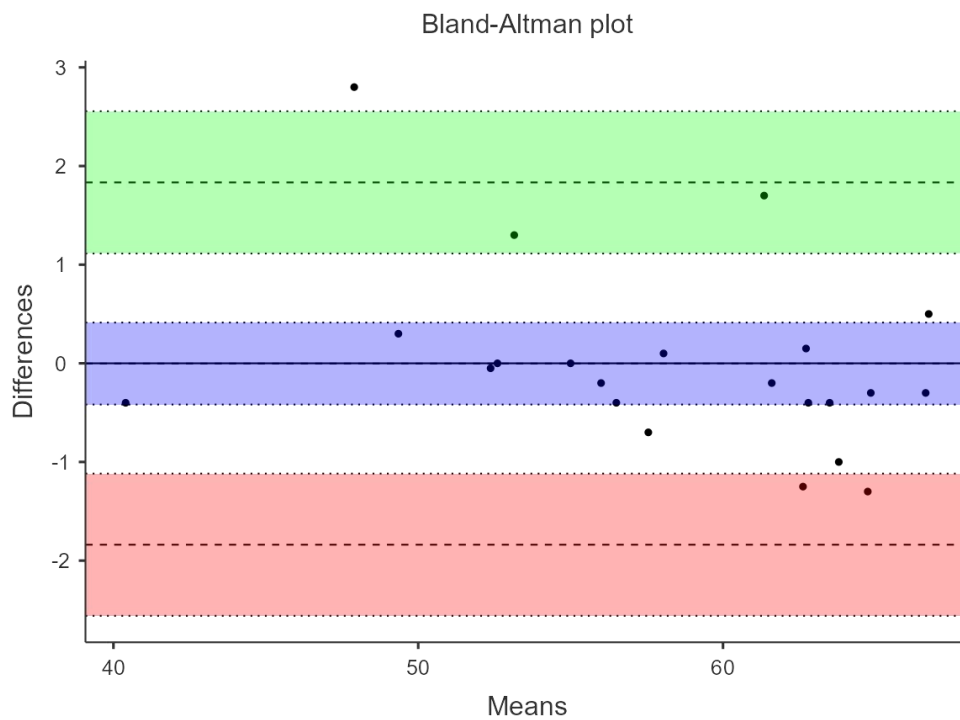
Табела 10. Bland–Altman согласност на варијаблите кај девојчињата

Тест/варијабла	Bias (T1–T2) (kg)	95% CI (Bias)	Lower LoA (kg)	95% CI (Lower LoA)	Upper LoA (kg)	95% CI (Upper LoA)	r(MEAN, ADIFF)	p
Телесна висина (cm)	-0.18	-0.33 – -0.03	-0.841	-1.09 – -0.58	0.469	0.21 – 0.72	-0.364	0.096
Телесна тежина (kg)	-0.00	-0.41 – 0.41	-1.839	-2.56 – -1.11	1.834	1.11 – 2.55	-0.080	0.724
ВМІ (kg/m <sup>2</sup> )	0.05	-0.11 – 0.22	-0.700	-0.99 – -0.40	0.801	0.50 – 1.09	0.106	0.640
Обем на половина (cm)	-0.27	-1.26 – 0.71	-4.642	-6.36 – -2.92	4.097	2.38 – 5.81	-0.379	0.082
Претклон во сед (cm)	-2.32	-3.70 – -0.94	-8.41	-10.80 – -6.01	3.77	1.38 – 6.16	-0.021	0.926
Скок од место (cm)	-2.62	-8.57 – 3.33	-28.25	-38.58 – -17.91	23.01	12.67 – 33.34	0.093	0.687
Динамометрија десна (kg)	0.21	-0.826 – 1.24	-4.368	-6.16 – -2.57	4.786	2.98 – 6.58	0.076	0.738
Динамометрија лева (kg)	0.02	-1.01 – 1.05	-4.521	-6.30 – -2.74	4.567	2.78 – 6.35	-0.025	0.910
Динамометрија (kg)	0.12	-0.791 – 1.02	-3.893	-5.46 – -2.32	4.125	2.55 – 5.70	-0.017	0.941
Подигнување труп 30 s	-2.19	-3.24 – -1.14	-6.73	-8.56 – -4.90	2.35	0.51 – 4.18	0.466*	0.033
Чунесто 4×10 m (s)	0.03	-0.14 – 0.21	-0.726	-1.04 – -0.41	0.795	0.47 – 1.111	0.164	0.488

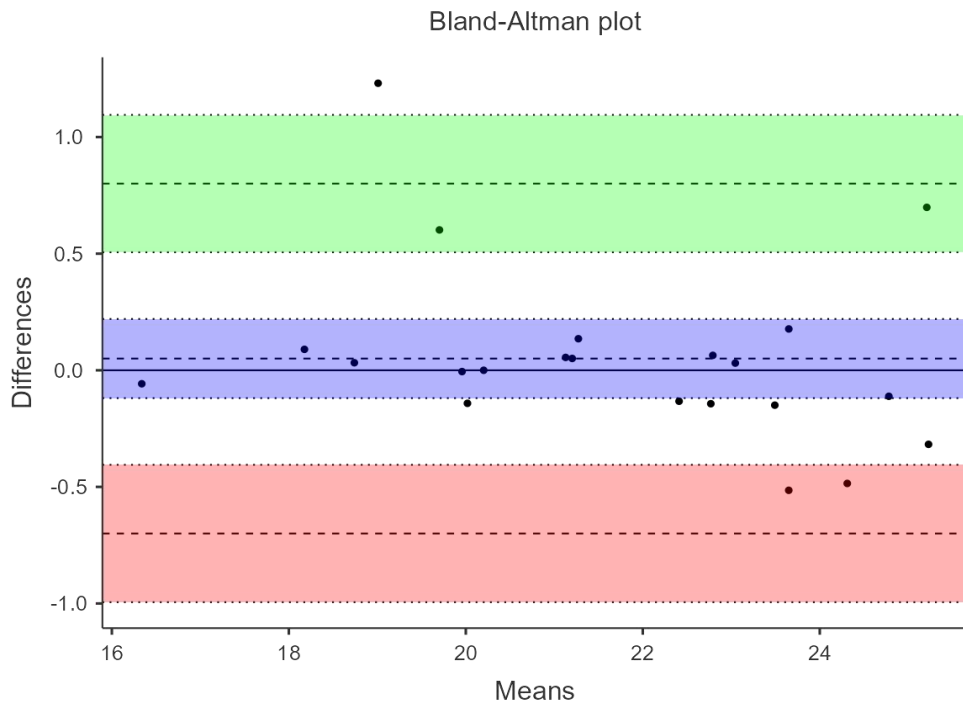
Графикон 12. Bland-Altman дијаграм на телесната висина кај девојчињата



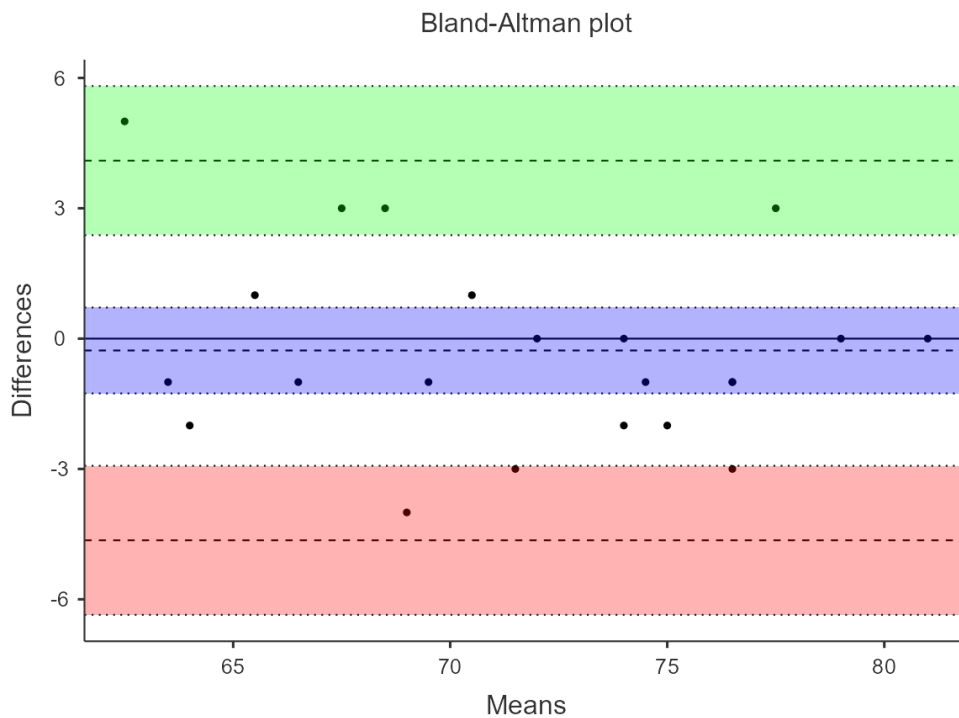
Графикон 13. Bland-Altman дијаграм на телесна тежина кај девојчињата



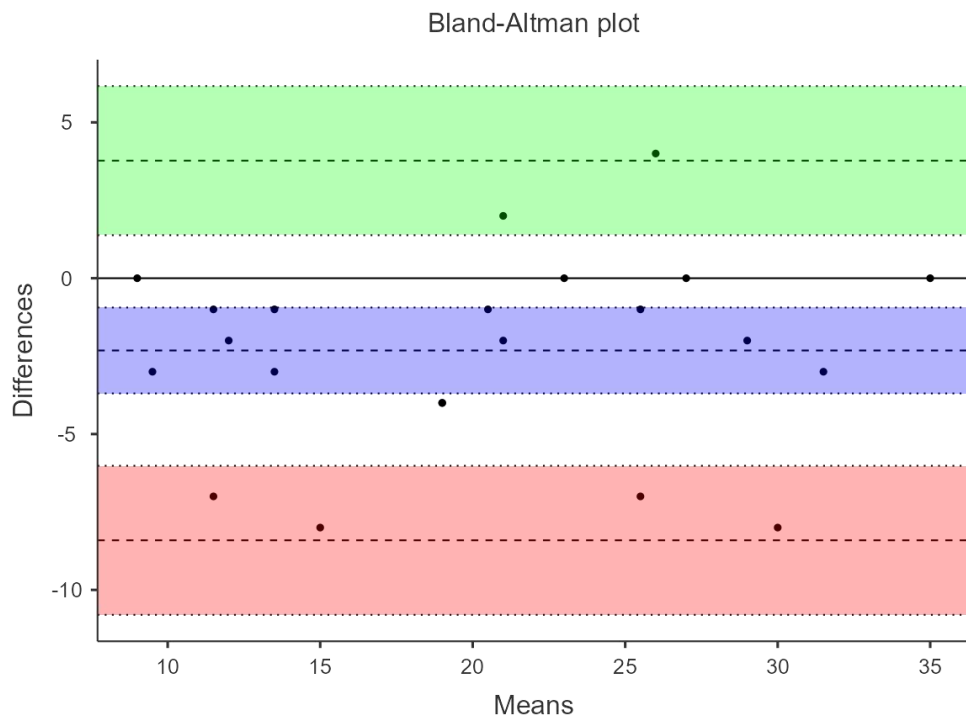
Графикон 14. Bland-Altman дијаграм на индексот на телесна маса (BMI) кај девојчињата



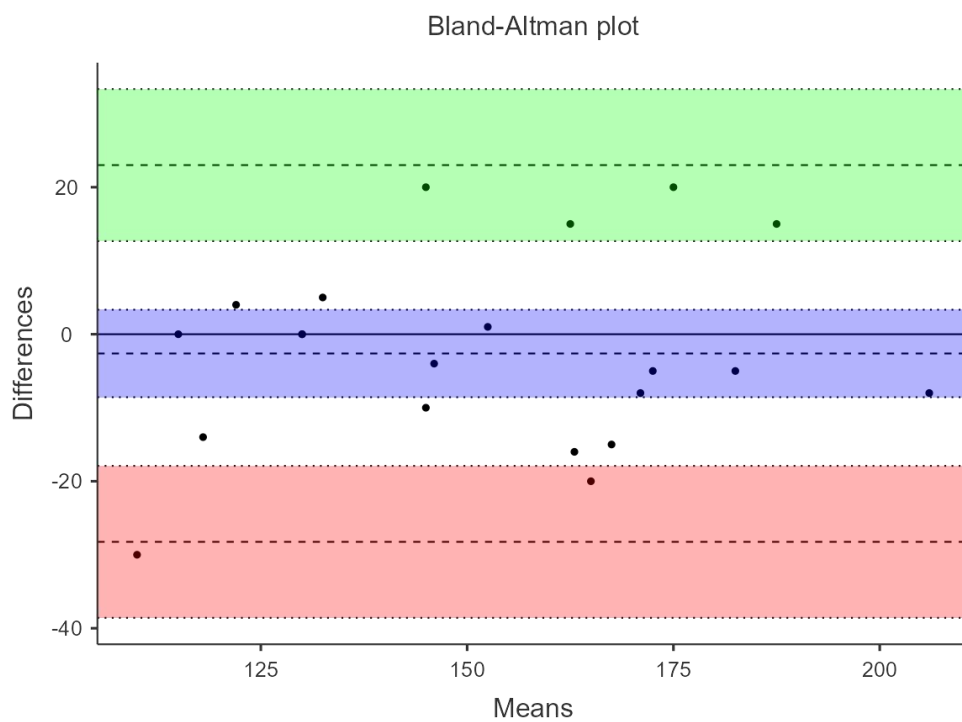
Графикон 15. Bland-Altman дијаграм на обем на половината (WC) кај девојчињата



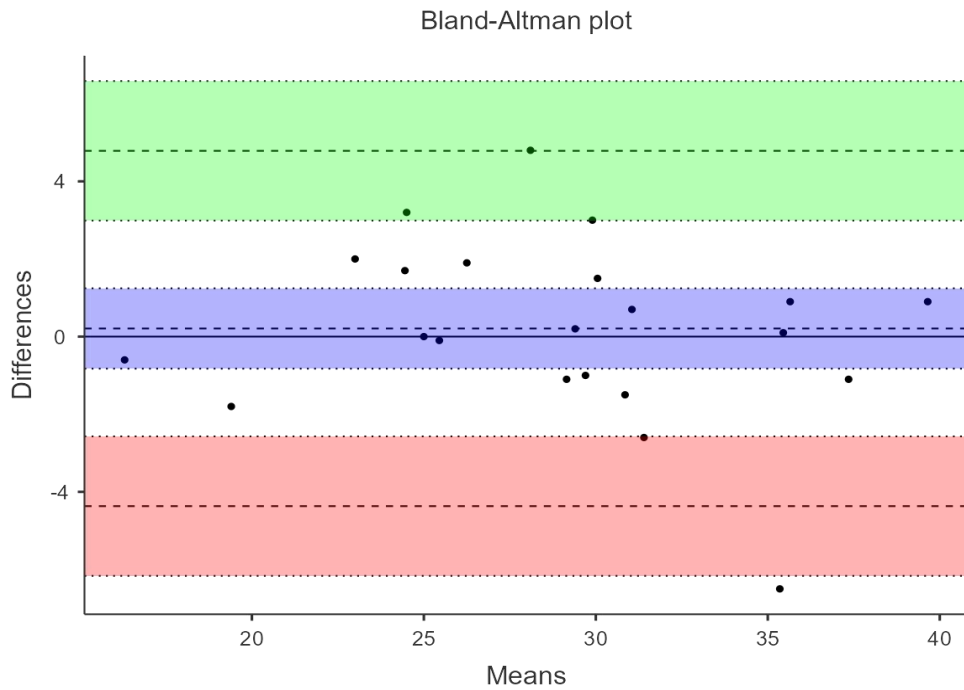
Графикон 16. Bland-Altman дијаграм на тестот претклон во сед кај девојчињата



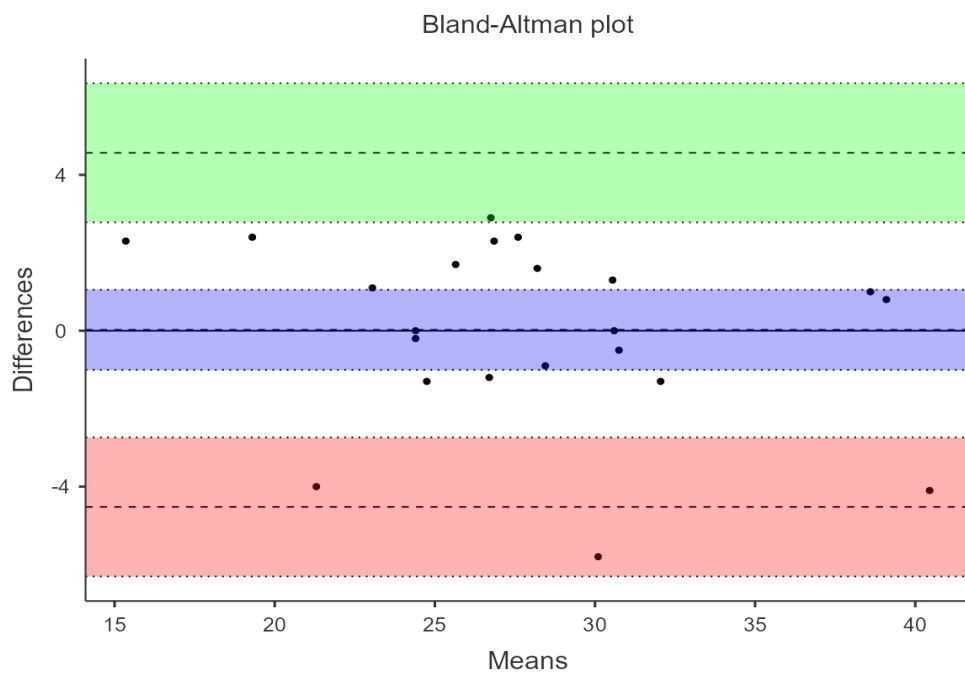
Графикон 17. Bland-Altman дијаграм на тестот скок во далечина од место кај девојчињата



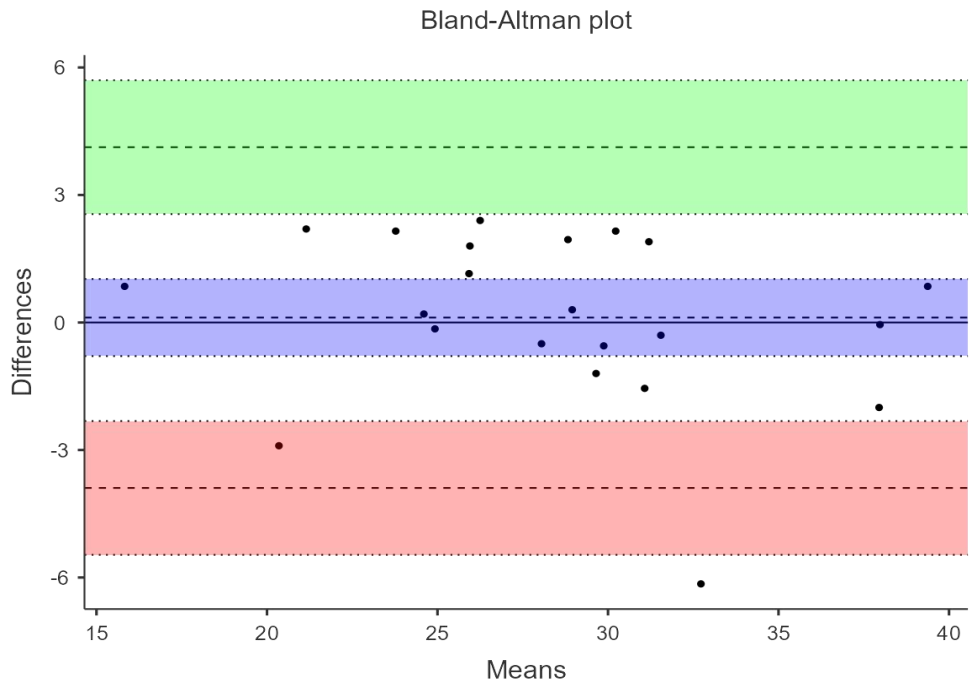
Графикон 18. Bland-Altman дијаграм на тестот динамометрија на дланка десна рака кај девојчињата



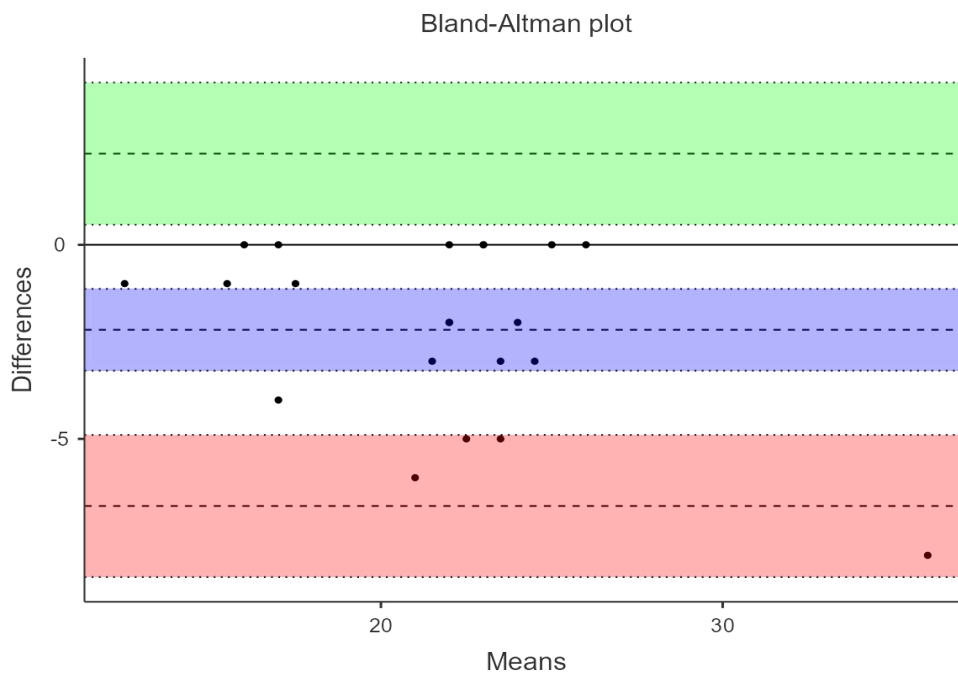
Графикон 19. Bland-Altman дијаграм на тестот динамометрија на дланка на лева рака кај девојчињата



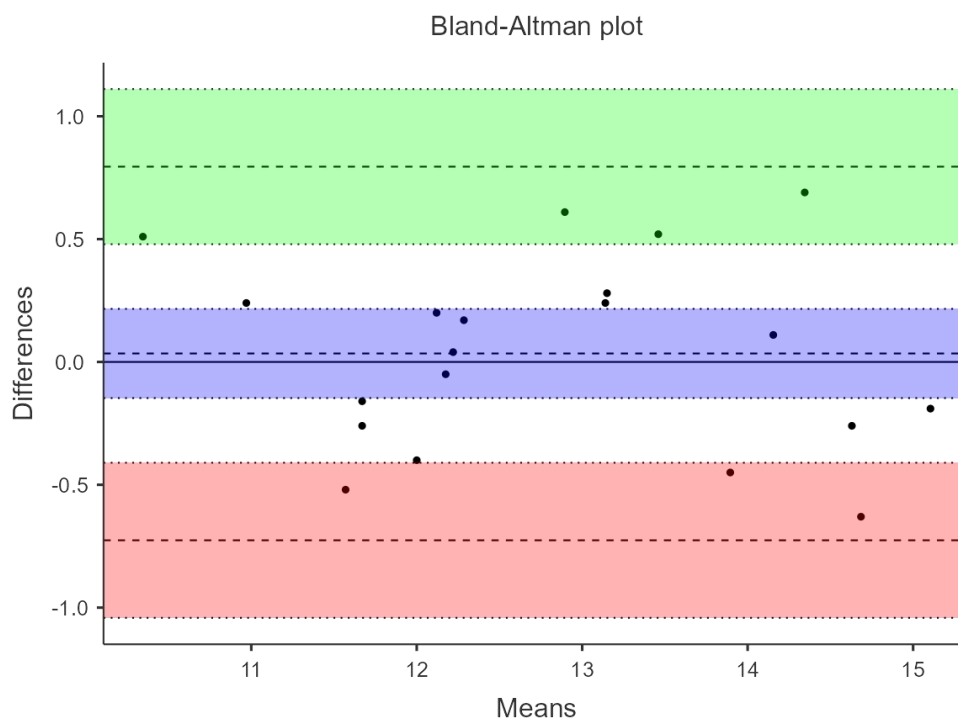
Графикон 20. Bland-Altman дијаграм на тестот динамометрија на дланка просечна вредност од десна и лева рака кај девојчињата



Графикон 21. Bland-Altman дијаграм на тестот подигнување труп 30 секунди кај девојчињата



Графикон 22. Bland-Altman дијаграм на тестот Чунесто трчање 4×10 m кај девојчињата



#### 5.4. КОМПАРАТИВЕН УВИД МЕЃУ МОМЧИЊАТА И ДЕВОЈЧИЊАТА НИЗ ТРИТЕ АНАЛИТИЧКИ РАМКИ

Компаративниот увид меѓу момчињата и девојчињата низ трите аналитички рамки ја прикажува релијабилноста како конзистентно висока кај антропометриските мерки, умерено висока кај повеќето моторички тестови и поосетлива на ретест ефект кај тестовите каде учењето, темпото и тактичката изведба имаат поголема улога. Кај двата пола, релативната релијабилност мерена со ICC(2,1) е највисока кај телесната висина, телесната тежина и ВМИ, додека кај тестовите за флексибилност и силова издржливост се забележуваат поизразени систематски разлики меѓу првото и второто мерење и поголема индивидуална варијабилност, што најјасно се потврдува преку Bland–Altman границите на согласност и индикаторите за апсолутна релијабилност.

Во однос на релативната релијабилност и систематската разлика, двата пола покажуваат речиси „плафонска“ стабилност за телесната висина, но со статистички значајна разлика меѓу навратите. Кај момчињата се добива  $t = 2.38$ ,  $p = .020$  со мал ефект ( $d_z = 0.302$ ) при ICC = 0.999 (95% CI: 0.998–0.999), а кај девојчињата  $t = -2.62$ ,  $p = .016$  со умерено поголем ефект ( $d_z = 0.558$ ) при ICC = 0.998 (95% CI: 0.994–0.999). Иако статистички значајни, овие разлики се практично минимални ( $\Delta$  од редот на десетинки сантиметар) и најверојатно одразуваат микроваријации на позиционирање и мерење, а

не реална промена. Телесната тежина и ВМІ кај двата пола немаат систематски разлики (момчиња: тежина  $p = .850$ ,  $dz = 0.024$ ; ВМІ  $p = .365$ ,  $dz = 0.116$ ; девојчиња: тежина  $p = .991$ ,  $dz = 0.002$ ; ВМІ  $p = .544$ ,  $dz = 0.131$ ) и остануваат со исклучително високи ICC вредности (момчиња 0.996 и 0.992, девојчиња 0.991 и 0.988), што укажува дека рангирањето на испитаниците е стабилно и дека мерките се многу погодни за мониторинг и споредби во рамките на популацијата.

Клучните полови разлики во релативната релијабилност се појасни кај тестовите каде постои ретест подобрување. Претклонот во сед е значајно подобрен и кај момчињата и кај девојчињата, но интензитетот на ефектот е висок во двата пола и малку поголем кај девојчињата (момчиња:  $t = -5.41$ ,  $p < .001$ ,  $dz = 0.686$ ,  $ICC = 0.894$ ; девојчиња:  $t = -3.50$ ,  $p = .002$ ,  $dz = 0.746$ ,  $ICC = 0.886$ ). Подигнувањето на трупот за 30 секунди исто така покажува најизразен систематски ретест ефект, повторно во двата пола, но со понагласена практична големина кај девојчињата (момчиња:  $t = -6.28$ ,  $p < .001$ ,  $dz = 0.811$ ,  $ICC = 0.822$ ; девојчиња:  $t = -4.33$ ,  $p < .001$ ,  $dz = 0.946$ ,  $ICC = 0.819$ ). Скокот од место е значајно подобрен кај момчињата ( $t = -4.74$ ,  $p < .001$ ,  $dz = 0.612$ ,  $ICC = 0.913$ ), но кај девојчињата не покажува статистички значајна разлика ( $t = -0.918$ ,  $p = .370$ ,  $dz = 0.200$ ), иако ICC останува висок (0.884), што сугерира дека кај девојчињата варијацијата е повеќе случајна и индивидуална, додека кај момчињата постои појасно групно подобрување. За чунестото  $4 \times 10$  m, кај двата пола нема систематска разлика (момчиња  $p = .538$ ; девојчиња  $p = .695$ ), а ICC е исклучително висок (0.994 и 0.995), што укажува на многу стабилно рангирање и практична употребливост за повторени мерења.

Апсолутната релијабилност ги нијансира овие наоди преку големината на грешката на мерење и прагот на реална промена. Кај телесната висина SEM% е идентичен и минимален во двата пола (0.14%), со многу мал MDC (момчиња 0.690 cm; девојчиња 0.655 cm), што практично значи дека секоја промена помала од околу две третини сантиметар може разумно да се припише на мерна варијабилност. За телесната тежина SEM% е ниска и слична (момчиња 0.98%, девојчиња 1.11%), а MDC е околу 1.89 kg кај момчињата и 1.79 kg кај девојчињата, што поставува јасен праг за интерпретација на индивидуални промени. Кај ВМІ, SEM% е околу 1% во двата пола, со MDC 0.612 кај момчињата и 0.739 kg/m<sup>2</sup> кај девојчињата, што имплицира малку послаба апсолутна прецизност кај девојчињата. Разликите стануваат поочигледни кај перформансните тестови: скокот од место има значајно поголем SEM% кај девојчињата (6.04%) отколку кај момчињата (3.92%), со повисок MDC (25.53 наспроти 21.88 cm), што практично значи дека за девојчињата е потребна поголема промена за да се смета за „реална“ надвор

од грешката. Кај претклонот во сед, пак, момчињата покажуваат поголема релативна грешка (SEM% 13.38%) отколку девојчињата (10.79%), што сугерира дека флексибилноста кај момчињата е понестабилна на краток рок или е посензитивна на моменталната состојба и изведба. Кај подигнувањето труп 30 s, грешката е повисока кај девојчињата (SEM% 7.55% наспроти 5.57%), што е во согласност со посилниот ретест ефект и пошироката индивидуална варијабилност; тука и изборот на индикатор во табелата сигнализира различен акцент, бидејќи кај девојчињата е означен метод „CV“, што во пракса упатува дека релативната варијабилност преку CV може да биде поинформативна за интерпретација отколку апсолутниот SEM сам по себе. За динамометријата, момчињата конзистентно имаат помали SEM% и CV (приближно 3.2–4.3%) отколку девојчињата (околу 5.0–5.7%), што укажува дека мерењата на сила кај девојчињата се нешто помалку прецизни на индивидуално ниво и дека праговите за реална промена се релативно повисоки.

Bland–Altman анализата ја потврдува групната слика со најдиректна интерпретација на индивидуалната согласност. Кај телесната висина, пристрасноста е мала во двата пола, но малку поголема кај девојчињата (момчиња bias =  $-0.10$  cm; девојчиња bias =  $-0.18$  cm), при тесни граници на согласност, што повторно ја нагласува одличната повторливост. Кај телесната тежина и BMI, пристрасностите се близу нула во двата пола и LoA се умерени и споредливи, што значи дека индивидуалните разлики генерално остануваат во очекуван опсег без доказ за систематско поместување. Кај обемот на половината, девојчињата имаат пошироки LoA (приближно  $-4.64$  до  $4.10$  cm) отколку момчињата (приближно  $-3.77$  до  $3.87$  cm), што е во склад со поголемиот SEM% и сугерира поголема чувствителност на методските услови. Дополнително, кај момчињата се детектира статистички значајна поврзаност меѓу средната вредност и апсолутната разлика за обем на половина ( $r = 0.277$ ,  $p = 0.030$ ), што е сигнал за потенцијална хетероскедастичност, односно тенденција грешката да се зголемува со поголем обем. Кај девојчињата истиот индикатор е негативен но незначаен ( $r = -0.379$ ,  $p = 0.082$ ), што повеќе упатува на тренд отколку на стабилен ефект.

Најважната полова разлика во Bland–Altman рамката се појавува кај подигнувањето труп 30 секунди. И кај момчињата и кај девојчињата пристрасноста е негативна и значајна, што значи дека T2 има подобри резултати, но кај девојчињата овој ефект е поголем (момчиња bias =  $-1.72$ ; девојчиња bias =  $-2.19$ ) и, уште поважно, само кај девојчињата се јавува значајна корелација меѓу средната вредност и апсолутната разлика ( $r = 0.466$ ,  $p = 0.033$ ), што укажува дека варијабилноста зависи од нивото на

перформанс, односно кај посилените или поуспешните девојчиња флукуациите меѓу навратите се поголеми. Кај момчињата ова не се потврдува ( $r = -0.058$ ,  $p = 0.661$ ). Практично, ова значи дека интерпретацијата на индивидуален напредок кај девојчињата во овој тест треба да биде пооптимизирана, со поголем акцент на прагови на реална промена и на стандарди за протокол и мотивација. Кај претклонот во сед, и двата пола покажуваат јасен ретест ефект, но границите на согласност се нешто пошироки кај момчињата, што е во согласност со повисок SEM% и поголема варијабилност. За скокот од место, иако просечната пристрасност е поголема кај момчињата, кај девојчињата се забележува многу широк опсег на LoA во позитивна насока, што е компатибилно со поголем SEM% и поголем MDC и упатува дека индивидуалните разлики се потешко предвидливи кај девојчињата. Кај чунестото  $4 \times 10$  m, двата пола имаат тесни LoA и минимален bias, што ја потврдува неговата висока практична стабилност.

Сумирано, и момчињата и девојчињата демонстрираат многу висока релативна релијабилност за антропометрија и агилност, но апсолутната релијабилност и Bland–Altman согласноста откриваат дека кај девојчињата грешката е поголема за скокот од место, силовите мерења и подигнувањето на труп, додека кај момчињата поголема релативна варијабилност се забележува кај претклонот во сед. Ретест подобрувањето е конзистентно присутно кај флексибилноста и абдоминалната силова издржливост во двата пола, но е поизразено кај девојчињата за подигнувањето труп, при што кај нив дополнително се јавува и пропорционална варијабилност. Овие разлики се методолошки значајни бидејќи сугерираат дека при интерпретација на индивидуални промени, особено во моторичките тестови, не е доволно да се потпреме на ICC и  $p$  вредности, туку мора да се користат SEM и MDC, како и LoA, со посебно внимание на тестовите со изразен ретест ефект и на случаите каде грешката зависи од нивото на перформанс.

## 5.5. ПРАГ НА „РЕАЛНА ПРОМЕНА“ (MDC95/SDD) КАЈ МОМЧИЊА И ДЕВОЈЧИЊА

Табела 11 прикажува праг на „реална промена“ (MDC95/SDD) за секој тест, одделно за момчиња и девојчиња. Овој праг претставува минимална промена што мора да се надмине за да можеме со приближно 95% сигурност да кажеме дека забележаната разлика меѓу две мерења не е само последица на вообичаена мерна грешка, дневни осцилации или варијација на изведбата, туку дека најверојатно станува збор за вистинска промена (на пример, по тренинг-програма, интервенција или систематско подобрување).

Во колоната „Насока на подобрување“ е наведено дали подобрувањето се очекува како зголемување на резултатот ( $\uparrow$ ), како што е случај со претклонот, скокот и силата на стисок, или како намалување ( $\downarrow$ ), како кај чунестото трчање каде помалото време значи подобар резултат. Затоа, правилото за читање е едноставно: промената е „реална“ само ако е поголема од MDC95 во правилната насока. Ако промената е помала од овој праг, тогаш разликата може целосно да се објасни со нормална варијабилност на мерењето и не треба да се толкува како сигурен ефект на интервенција.

Конкретно, кај телесната висина се гледа дека прагот е многу мал (0.690 cm кај момчиња и 0.655 cm кај девојчиња), што укажува дека мерењето е многу прецизно и дека и релативно мала разлика поголема од околу две третини сантиметар веќе може да се смета за „реална“. Кај телесната тежина прагот е околу 1.9 kg кај момчиња и 1.8 kg кај девојчиња, што значи дека промени помали од овие вредности (на пример  $\pm 0.5$  kg) најчесто се во рамки на нормални дневни осцилации и мерна варијација. Слично, кај ВМІ реална промена се смета разлика поголема од 0.612 kg/m<sup>2</sup> кај момчиња и 0.739 kg/m<sup>2</sup> кај девојчиња, што практично значи дека „мали“ поместувања во ВМІ често не се доволни за сигурен заклучок за вистинска промена.

За обемот на половина, прагот за реално намалување е 3.795 cm кај момчиња и 4.302 cm кај девојчиња. Ова е важна практична информација: ако интервенцијата доведе до намалување од, на пример, 2 cm, тоа е охрабрувачко, но од гледна точка на мерната сигурност сè уште не е доволно за да се тврди дека промената е надвор од мерната грешка. Кај претклон во сед, праговите се релативно високи (7.300 cm кај момчиња и 6.090 cm кај девојчиња), што упатува дека флексибилноста има поголема варијабилност и дека само изразени подобрувања може сигурно да се припишат на реален напредок.

Кај моторичките тестови со поголема биолошка варијабилност, праговите се уште поизразени. На пример, за скок од место потребно е подобрување од најмалку 21.88 cm кај момчиња и 25.53 cm кај девојчиња за да се смета за реален напредок. Ова не значи дека помало подобрување нема вредност, туку дека со две мерења не можеме сигурно да кажеме дека напредокот не е делумно објаснив со варијабилност. Кај динамометријата, праговите за реална промена се околу 4–6 kg зависно од раката и полот (на пример, 5.607 kg за десна кај момчиња и 4.490 kg кај девојчиња), што дава јасен критериум кога промената на силата е доволно голема за да се смета за сигурен ефект.

За подигнување труп за 30 секунди, реален напредок се смета ако зголемувањето е најмалку околу 4 повторувања (4.147 кај момчиња и 4.538 кај девојчиња). Тоа значи дека разлика од 1–2 повторувања може да биде само нормална варијација на изведбата.

Конечно, за चुнето трчање 4×10 m, бидејќи подобрувањето е намалување на времето, реална промена е кога времето ќе се намали најмалку за 0.638 s кај момчиња и 0.745 s кај девојчиња; помали „подобрувања“ може да бидат во рамки на очекуваната мерна варијабилност.

Сумирано, Табела 11 обезбедува практичен „филтер“ за интерпретација на промените: таа помага да се разграничи дали разликата меѓу две мерења најверојатно е вистински ефект или нормална варијација. Ова е особено корисно во евалуација на интервенции, бидејќи спречува преценување на мали промени кои статистички или визуелно изгледаат како напредок, но можат да бидат во рамки на очекуваната грешка на мерењето.

Табела 11. Праг на „реална промена“ (MDC95/SDD) кај момчиња и девојчиња

Тест/варијабла	Насока на подобрување	MDC95 момчиња	MDC95 девојчиња	Како да се чита
Телесна висина (cm)	↑/↓ (зависи од контекст)	0.690 cm	0.655 cm	Реална промена ако
Телесна тежина (kg)	↑ или ↓	1.888 kg	1.794 kg	Реална промена ако
BMI (kg/m <sup>2</sup> )	↑ или ↓	0.612	0.739	Реална промена ако
Обем на половина (cm)	↓ (обично)	3.795 cm	4.302 cm	Реална промена ако намалување ≥ 3.80 / 4.30 cm
Претклон во сед (cm)	↑	7.300 cm	6.090 cm	Реално подобрување ако зголемување ≥ 7.30 / 6.09 cm
Скок од место (cm)	↑	21.88 cm	25.53 cm	Реално подобрување ако зголемување ≥ 21.9 / 25.5 cm
Динамометрија десна (kg)	↑	5.607 kg	4.490 kg	Реално подобрување ако зголемување ≥ 5.61 / 4.49 kg
Динамометрија лева (kg)	↑	4.618 kg	4.438 kg	Реално подобрување ако зголемување ≥ 4.62 / 4.44 kg
Динамометрија просек (kg)	↑	4.081 kg	3.922 kg	Реално подобрување ако зголемување ≥ 4.08 / 3.92 kg
Подигнување труп 30 s (повт.)	↑	4.147 повт.	4.538 повт.	Реално подобрување ако зголемување ≥ 4.15 / 4.54 повторувања
Чунето 4×10 m (s)	↓ (помало = подобро)	0.638 s	0.745 s	Реално подобрување ако резултат е <b>намален</b>

## 5.6. ICC SUMMARY (ICC(2,1), 95% CI) – МОМЧИЊА ВО ОДНОС НА ДЕВОЈЧИЊА (РЕЛАТИВНА РЕЛИЈАБИЛНОСТ)

Релативната релијабилност се однесува на тоа колку стабилно тестот ги задржува истите разлики меѓу испитаниците кога мерењето се повторува. Таа не кажува само дали просекот на групата е сличен од првото до второто мерење, туку дали рангирањето на поединците останува конзистентно, односно дали оние кои биле меѓу најдобрите при првото мерење остануваат меѓу најдобрите и при второто. За оваа цел се користи интракласниот корелациски коефициент, ICC, кој се движи од 0 до 1, при што повисоките вредности означуваат поголема стабилност и сигурност на тестот во повторени услови. Во табелата е прикажан ICC(2,1) со 95% доверлив интервал, кој ја покажува веројатната граница во која се наоѓа вистинската вредност на релијабилноста. Потесен доверлив интервал упатува на поголема прецизност на проценката, додека поширок интервал најчесто се појавува кај помали примероци или кај тестови со поголема варијабилност во изведбата.

Општата слика од ICC summary табелата укажува дека кај момчињата и девојчињата најголемиот дел од мерките имаат одлична релативна релијабилност. Тоа практично значи дека тестовите многу добро ги разликуваат учениците меѓусебно и дека таа меѓусебна разлика останува стабилна при повторно тестирање. Со други зборови, овие инструменти и протоколи се доволно конзистентни за да се користат во проценка и следење, особено кога целта е да се утврди кои испитаници се посилни, побрзи, повисоки или со поголема телесна маса во однос на останатите.

Кај антропометриските варијабли, релативната релијабилност е исклучително висока. Телесната висина покажува речиси совршена стабилност и кај момчињата и кај девојчињата, што значи дека мерењето е практично безначајно чувствително на случајни варијации при повторување и дека рангирањето на испитаниците според висина останува непроменето. Слично, телесната тежина и ВМІ имаат одлични ICC вредности во двете групи, што упатува дека овие показатели многу конзистентно ги одразуваат индивидуалните разлики при тест ретест. Обемот на половина исто така е во зоната на одлична релијабилност кај двата пола, иако кај девојчињата доверливиот интервал е нешто поширок, што е очекувано во услови на помал примерок и потенцијално поголема чувствителност на мерењето на техниката и поставувањето на мерната лента. Сепак, и во тој случај, заклучокот останува дека мерката е доволно стабилна за групни анализи и практична примена.

Кај моторичките тестови, релативната релијабилност останува висока, но се забележуваат очекувани разлики зависно од природата на тестот. Претклонот во сед има добра релијабилност кај двата пола, што значи дека тестот стабилно ги рангира испитаниците, но сепак може да биде подложен на влијанија како загревање, моментална напнатост, техника на изведба и варијации во истегнатоста од ден на ден. Скокот од место кај момчињата покажува одлична релијабилност, додека кај девојчињата е во рамките на добра релијабилност, што сугерира дека тестот е стабилен и корисен и кај двете групи, но кај девојчињата може да постои нешто поголема варијабилност во повторувањето, веројатно поради техника, мотивациски фактори или природна флукуација во изведбата. Подигнувањето на трупот за 30 секунди покажува добра релијабилност кај двата пола, но со пошироки доверливи интервали, што е типично за тестови со повторувања каде на резултатот влијаат ритмот, стратегијата на изведба, краткотрајниот замор и степенот на поттик при тестирањето. Затоа, иако тестот е доволно стабилен, разумно е да се очекува нешто поголема варијација во споредба со чисто антропометриските мерки.

Динамометријата, и за десна и за лева рака, како и пресметаниот просек, покажува одлична релативна релијабилност и кај момчињата и кај девојчињата. Ова има јасна практична импликација, бидејќи означува дека мерењето на силата на стисок многу конзистентно ги разликува испитаниците и дека резултатите се повторливи при повторно тестирање, што ја прави динамометријата сигурен индикатор за проценка на мускулната сила во теренски и училишни услови.

Кај тестот за агилност, чунестото 4×10 m, се добива одлична релативна релијабилност во двете групи. Тоа значи дека тестот е исклучително стабилен во поглед на тоа кои испитаници се побрзи и кои се побавни при повторување, односно рангирањето е речиси идентично од првото до второто мерење. Ваквата стабилност е важна затоа што овозможува сигурно следење на перформансот низ време и споредба меѓу испитаници, при што се намалува ризикот резултатите да се должат на случајни варијации.

Табела 12. ICC summary (ICC(2,1), 95% CI) – момчиња vs девојчиња (релативна релијабилност)

Тест/варијабла	Момчиња ICC (95% CI)	Кат.	Девојчиња ICC (95% CI)	Кат.
Телесна висина (cm)	0.999 (0.998–0.999)	Одлична	0.998 (0.994–0.999)	Одлична
Телесна тежина (kg)	0.996 (0.993–0.997)	Одлична	0.991 (0.982–0.996)	Одлична
ВМИ (kg/m <sup>2</sup> )	0.992 (0.988–0.995)	Одлична	0.988 (0.976–0.994)	Одлична
Обем на половина (cm)	0.948 (0.922–0.966)	Одлична	0.918 (0.838–0.959)	Одлична
Претклон во сед (cm)	0.894 (0.744–0.945)	Добра	0.886 (0.668–0.952)	Добра
Скок од место (cm)	0.913 (0.811–0.953)	Одлична	0.884 (0.771–0.943)	Добра
Динамометрија десна (kg)	0.934 (0.901–0.956)	Одлична	0.923 (0.849–0.962)	Одлична
Динамометрија лева (kg)	0.952 (0.928–0.968)	Одлична	0.934 (0.868–0.967)	Одлична
Динамометрија просек (kg)	0.961 (0.941–0.974)	Одлична	0.941 (0.883–0.971)	Одлична
Подигнување труп 30 s (повт.)	0.822 (0.534–0.913)	Добра	0.819 (0.401–0.929)	Добра
Чуносто 4×10 m (s)	0.994 (0.991–0.996)	Одлична	0.995 (0.991–0.998)	Одлична

Интерпретација (најчесто користена): <0.50 слаба, 0.50–0.75 умерена, 0.75–0.90 добра, >0.90 одлична релијабилност.

Во целина, табелата 12 покажува дека применетите антропометриски и фитнес тестови имаат висока до многу висока релативна релијабилност кај двата пола. Тоа значи дека тие се соодветни за проценка на индивидуалните разлики и за мониторинг на учениците, особено кога целта е да се идентификуваат поединци со повисоки или пониски перформанси во однос на врсниците. Истовремено, кај тестовите кои се почувствителни на техника и моментална состојба, како претклонот и подигнувањето труп, релијабилноста останува добра, што е прифатливо и очекувано, но укажува дека во практична примена треба да се води сметка за стандардизирани услови и доследна процедура при тестирање.

Дополнително, важно е да се нагласи дека ICC ја опишува релативната релијабилност, односно колку стабилно останува рангирањето на испитаниците од првото до второто мерење. Практично, висок ICC значи дека учениците кои имале повисоки резултати на првото мерење, во најголем дел остануваат повисоко ранжирани и на второто мерење, а оние со пониски резултати остануваат пониско. Сепак, ICC сам по себе не кажува колкава е мерната грешка изразена во истите единици како тестот, ниту кажува колкава промена е потребна за да се тврди дека станува збор за реален напредок кај поединец, а не за варијација поради мерењето. Поради тоа, резултатите од ICC се најдобро да се читаат заедно со показателите за апсолутна релијабилност, како SEM, CV и особено MDC, кои директно дефинираат праг на „реална промена“ што треба да се надмине за да се интерпретира ефект од интервенција или тренинг.

Вториот клучен аспект е прецизноста на проценката, која се гледа преку 95% доверлив интервал. Кога доверливиот интервал е тесен, тоа укажува дека проценката на релијабилноста е стабилна и малку варира во рамките на примерокот, што ја зголемува довербата дека тестот навистина има конзистентно повторливи резултати. Кога интервалот е поширок, тоа не значи дека тестот е „лош“, туку дека постои поголема несигурност во проценката, најчесто поради поголема варијабилност во изведбата, влијание на техника, мотивација или дневни флукуации, а понекогаш и поради помал број испитаници. Затоа, тестовите кои бараат прецизна техника и максимален ангажман често имаат поумерена до добра релативна релијабилност во споредба со антропометриските мерки, кои се пообјективни и помалку зависат од моментална изведба.

Од практична перспектива, ICC вредностите во оваа табела покажуваат дека инструментариумот е соодветен за сигурно споредување ученици меѓусебно и за групно следење на промените низ време и кај момчиња и кај девојчиња. Сепак, кога целта е да се процени дали конкретен ученик навистина напредувал, потребно е промената да се спореди со MDC праговите: само промени поголеми од тој праг се интерпретираат како реални и клинички или практично значајни, додека помалите промени најверојатно се во рамките на очекуваната мерна грешка. На тој начин, ICC обезбедува потврда дека тестот е стабилен во рангирањето, а MDC обезбедува јасен критериум за тоа кога промената е доволно голема за да се смета за вистински ефект, а не за случајна варијација.

#### 5.7. РАЗЛИКИ ВО АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ПОМЕЃУ МОМЧИЊАТА И ДЕВОЈЧИЊАТА ВО ПРВОТО МЕРЕЊЕ

Мултиваријантната анализа на коваријанса (MANCOVA) беше применета за да се испитаат половите разлики во антропометриските мерки и моторичките тестови при првото мерење, при што возраста беше вклучена како коваријата со цел да се контролира нејзиното потенцијално влијание врз испитуваните показатели. Мултиваријантниот тест покажа статистички значаен глобален ефект на полот врз збирната варијација на вклучените зависни променливи ( $Wilks' \lambda = 0,21$ ;  $F(10,64) = 23,53$ ;  $p < 0,001$ ;  $\eta^2 = 0,79$ ). Ниската вредност на  $Wilks' \lambda$  укажува на изразено раздвојување меѓу половите на мултиваријантно ниво, додека многу високата парцијална  $\eta^2$  сугерира дека полот објаснува исклучително голем дел од варијансата во заедничкиот простор на

антропометриските и моторичките исходи. Овој наод ја оправдува понатамошната униваријатна анализа на поединечните променливи.

Униваријатните резултати укажаа дека кај дел од основните антропометриски показатели се присутни силно изразени полови разлики. Телесната висина беше значајно поголема кај момчињата ( $M = 178,21$  cm;  $SD = 6,48$ ) во споредба со девојчињата ( $M = 163,90$  cm;  $SD = 5,61$ ), со многу висок ефект на полот ( $F = 75,64$ ;  $p < 0,001$ ;  $\eta^2 = 0,51$ ), што укажува дека половата припадност објаснува повеќе од половина од варијансата на овој показател во рамки на контролирана возраст. Слично, телесната тежина беше статистички значајно повисока кај момчињата ( $M = 69,03$  kg;  $SD = 10,27$ ) отколку кај девојчињата ( $M = 57,87$  kg;  $SD = 6,71$ ), со умерено до високо влијание на полот ( $F = 20,54$ ;  $p < 0,001$ ;  $\eta^2 = 0,22$ ). Наспроти тоа, индексот на телесна маса не покажува значајна полова диференцијација (момчиња:  $M = 21,68$  kg/m<sup>2</sup>;  $SD = 2,52$ ; девојчиња:  $M = 21,55$  kg/m<sup>2</sup>;  $SD = 2,47$ ), што е потврдено со занемарливиот F-тест и ефект ( $F = 0,01$ ;  $p = 0,939$ ;  $\eta^2 = 0,00$ ). Обемот на половината, пак, беше значајно поголем кај момчињата ( $M = 77,44$  cm;  $SD = 6,03$ ) во однос на девојчињата ( $M = 71,20$  cm;  $SD = 5,36$ ), со умерен ефект на полот ( $F = 16,26$ ;  $p < 0,001$ ;  $\eta^2 = 0,18$ ), што упатува на полого специфични разлики во абдоминалната димензија, независно од возраста.

Кај моторичките тестови, резултатите покажаа дека разликите не се униформни по сите способности. Флексибилноста проценета преку претклон во сед не се разликуваше статистички значајно меѓу половите (момчиња:  $M = 17,73$  cm;  $SD = 9,33$ ; девојчиња:  $M = 19,50$  cm;  $SD = 8,30$ ), при што ефектот беше мал и без практична релевантност ( $F = 0,68$ ;  $p = 0,411$ ;  $\eta^2 = 0,01$ ). Спротивно на тоа, експлозивната сила на долните екстремитети, мерена преку скок од место, беше значајно повисока кај момчињата ( $M = 200,34$  cm;  $SD = 23,35$ ) во однос на девојчињата ( $M = 150,80$  cm;  $SD = 28,74$ ), со многу висок ефект на полот ( $F = 57,55$ ;  $p < 0,001$ ;  $\eta^2 = 0,44$ ), што укажува на нагласена полова диференцијација во овој моторички домен.

Најизразени разлики беа регистрирани кај показателите на мускулна сила на горните екстремитети. Динамометријата на десната рака беше значајно поголема кај момчињата ( $M = 47,12$  kg;  $SD = 8,02$ ) отколку кај девојчињата ( $M = 28,95$  kg;  $SD = 5,52$ ), со исклучително висок ефект ( $F = 91,02$ ;  $p < 0,001$ ;  $\eta^2 = 0,56$ ). Идентичен тренд беше добиен и за левата рака (момчиња:  $M = 44,93$  kg;  $SD = 8,05$ ; девојчиња:  $M = 27,78$  kg;  $SD = 5,39$ ), со висок ефект ( $F = 78,14$ ;  $p < 0,001$ ;  $\eta^2 = 0,52$ ), како и за просечната динамометрија (момчиња:  $M = 46,03$  kg;  $SD = 7,75$ ; девојчиња:  $M = 28,36$  kg;  $SD = 5,27$ ), повторно со многу висок ефект ( $F = 90,98$ ;  $p < 0,001$ ;  $\eta^2 = 0,56$ ). Овие резултати

покажуваат дека во рамки на контролирана возраст, полот е доминантен фактор во објаснувањето на варијансата на рачната сила.

Функционалната сила и силовата издржливост на трупот, проценета преку подигнување труп за 30 секунди, исто така беше значајно повисока кај момчињата ( $M = 26,04$ ;  $SD = 4,41$ ) во однос на девојчињата ( $M = 20,85$ ;  $SD = 4,38$ ), со умерено висок ефект ( $F = 20,21$ ;  $p < 0,001$ ;  $\eta^2 = 0,22$ ). Во тестот за агилност  $4 \times 10$  m беше утврдена значајна разлика во корист на момчињата, кои постигнаа пониско време (подобар резултат) ( $M = 10,92$  s;  $SD = 0,74$ ) во споредба со девојчињата ( $M = 12,84$  s;  $SD = 1,32$ ), со висок ефект на полот ( $F = 64,13$ ;  $p < 0,001$ ;  $\eta^2 = 0,47$ ), што укажува на изразено полово раздвојување во брзинско агилните капацитети.

Сумирано, резултатите од Табела 13 покажуваат дека иако не постојат полови разлики кај ВМІ и флексибилноста, полот има значајно и во повеќето случаи силно влијание врз телесната висина, телесната тежина и обемот на половината, како и врз клучни моторички способности, особено експлозивната сила, рачната динамометрија, силовата издржливост на трупот и агилноста. Големините на ефект ( $\eta^2$ ) дополнително потврдуваат дека утврдените разлики не се само статистички значајни, туку и суштински изразени, при што најголемата дисперзија меѓу половите се регистрира кај силата на горните екстремитети и кај телесната висина, а потоа и кај експлозивната сила и агилноста.

Табела 13. Разлики во антропометриските мерки и моторичките тестови помеѓу момчињата и девојчињата во првото мерење

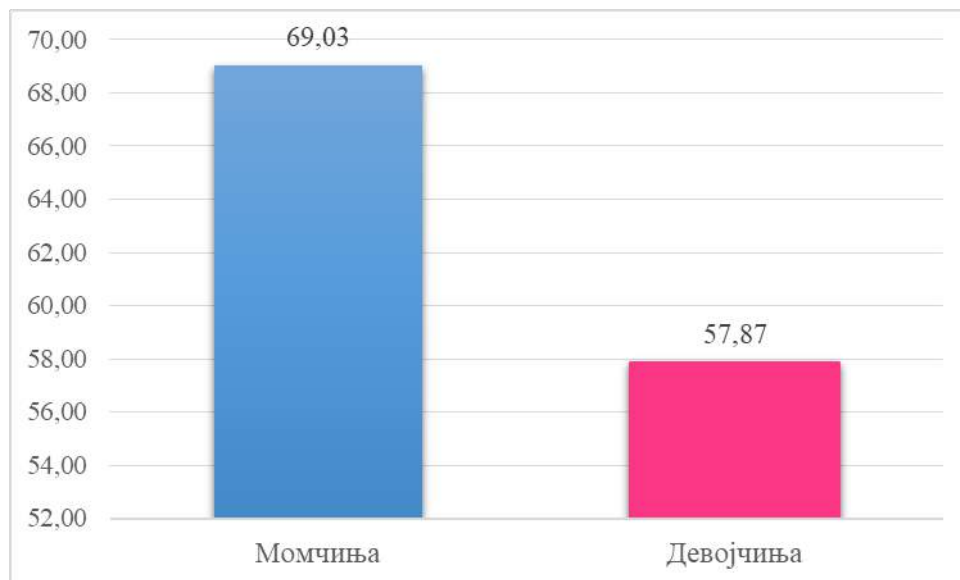
	Value	F	Hypothesis df	Error df	Sig.	$\eta^2$
Wilks' lambda	0,21	23,53	10	64	,000	0,79

Варијабли	Момчиња		Девојчиња		F	p.	$\eta^2$
	Mean	SD	Mean	SD			
Телесна висина (cm)	178,21	6,48	163,90	5,61	75,64	<b>,000</b>	0,51
Телесна тежина (kg)	69,03	10,27	57,87	6,71	20,54	<b>,000</b>	0,22
ВМІ (kg/m <sup>2</sup> )	21,68	2,52	21,55	2,47	0,01	,939	0,00
Обем на половина (cm)	77,44	6,03	71,20	5,36	16,26	<b>,000</b>	0,18
Претклон во сед (cm)	17,73	9,33	19,50	8,30	0,68	,411	0,01
Скок од место (cm)	200,34	23,35	150,80	28,74	57,55	<b>,000</b>	0,44
Динамометрија десна (kg)	47,12	8,02	28,95	5,52	91,02	<b>,000</b>	0,56
Динамометрија лева (kg)	44,93	8,05	27,78	5,39	78,14	<b>,000</b>	0,52
Динамометрија просек (kg)	46,03	7,75	28,36	5,27	90,98	<b>,000</b>	0,56
Подигнување труп 30 s (повт.)	26,04	4,41	20,85	4,38	20,21	<b>,000</b>	0,22
Чунесто $4 \times 10$ m (s)	10,92	0,74	12,84	1,32	64,13	<b>,000</b>	0,47

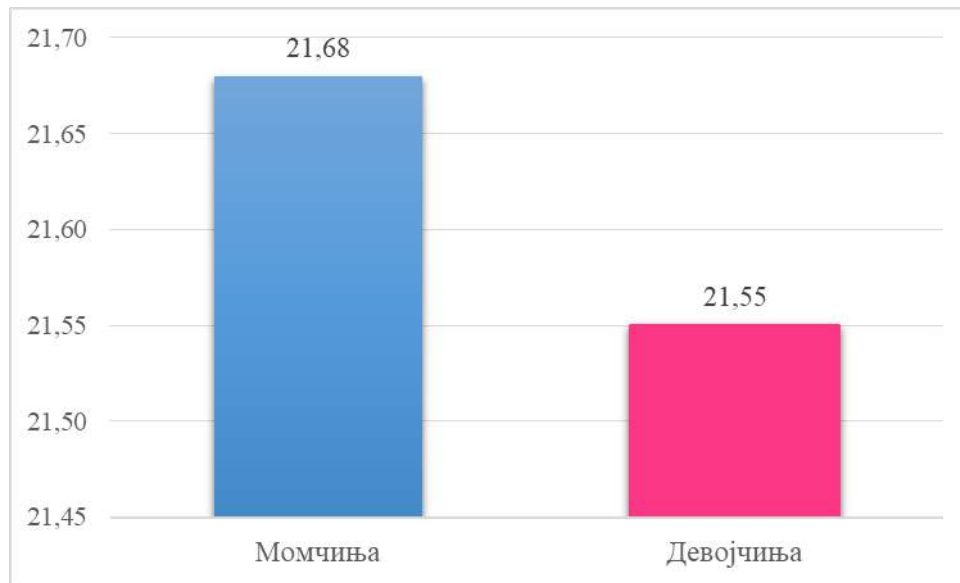
Графикон 23. Преглед на аритметичките средини на антропометриската мерка телесна висина помеѓу момчињата и девојчињата во првото мерење



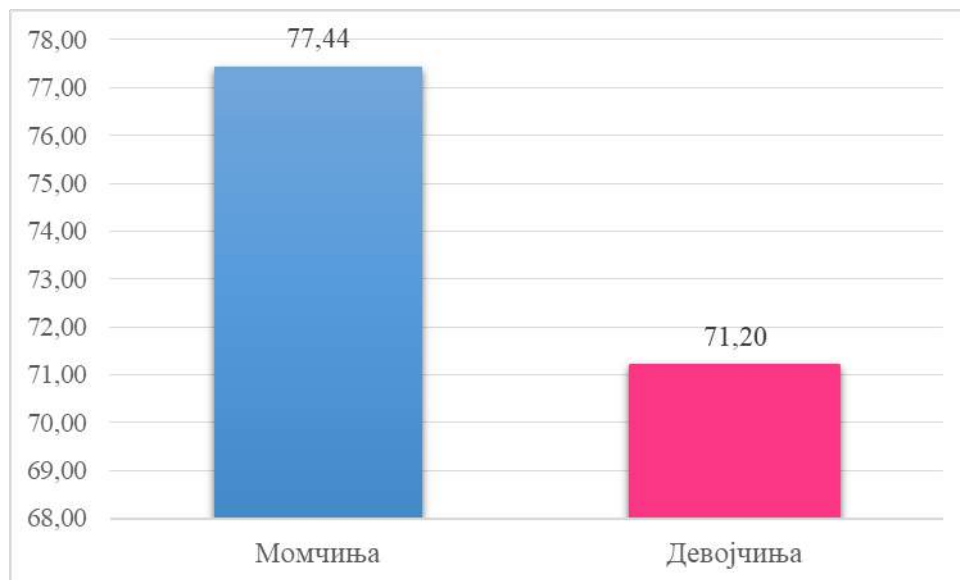
Графикон 24. Преглед на аритметичките средини на антропометриската мерка телесна тежина помеѓу момчињата и девојчињата во првото мерење



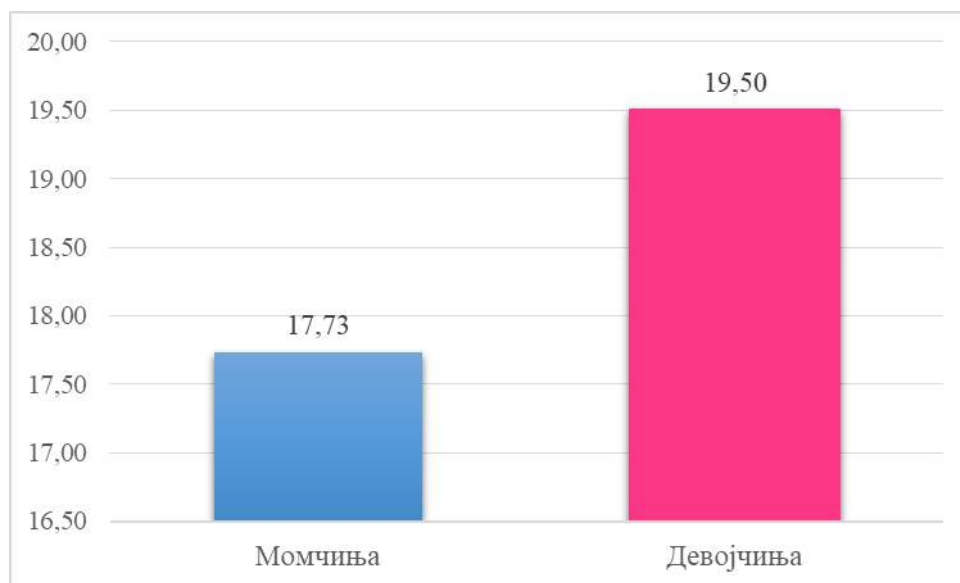
Графикон 25. Преглед на аритметичките средини на антропометриската мерка индекс на телесна маса помеѓу момчињата и девојчињата во првото мерење



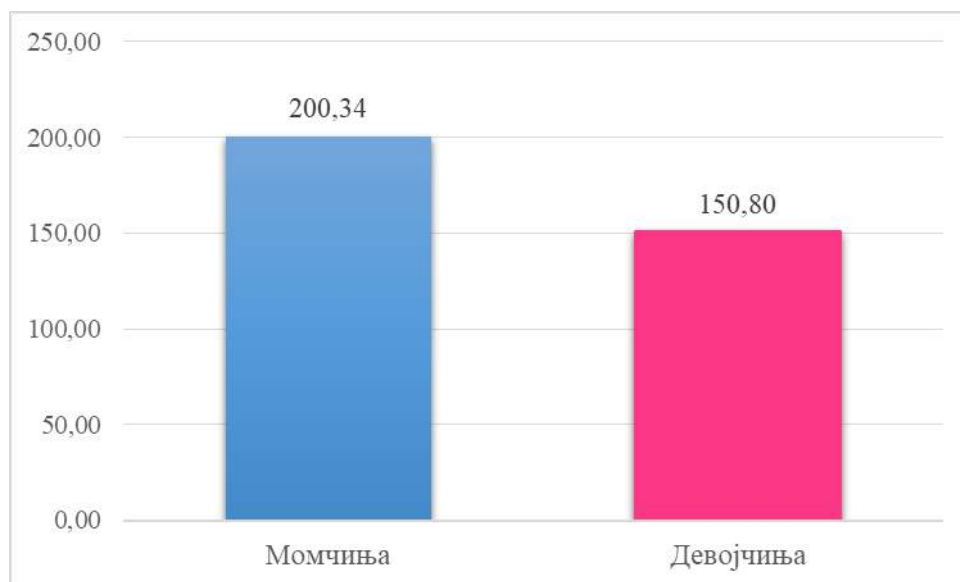
Графикон 26. Преглед на аритметичките средини на антропометриската мерка обем на половината помеѓу момчињата и девојчињата во првото мерење



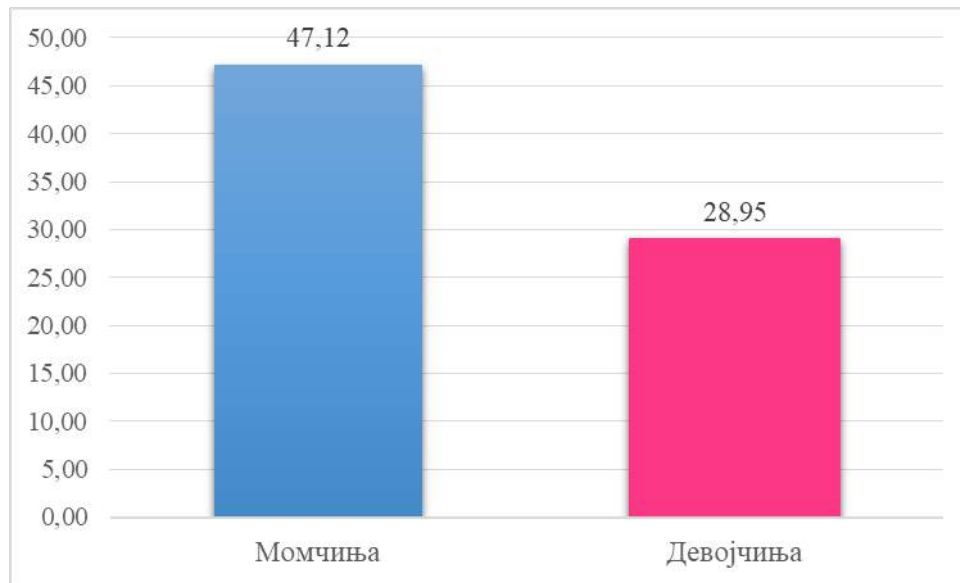
Графикон 27. Преглед на аритметичките средини на моторичкиот тест претклон во сед помеѓу момчињата и девојчињата во првото мерење



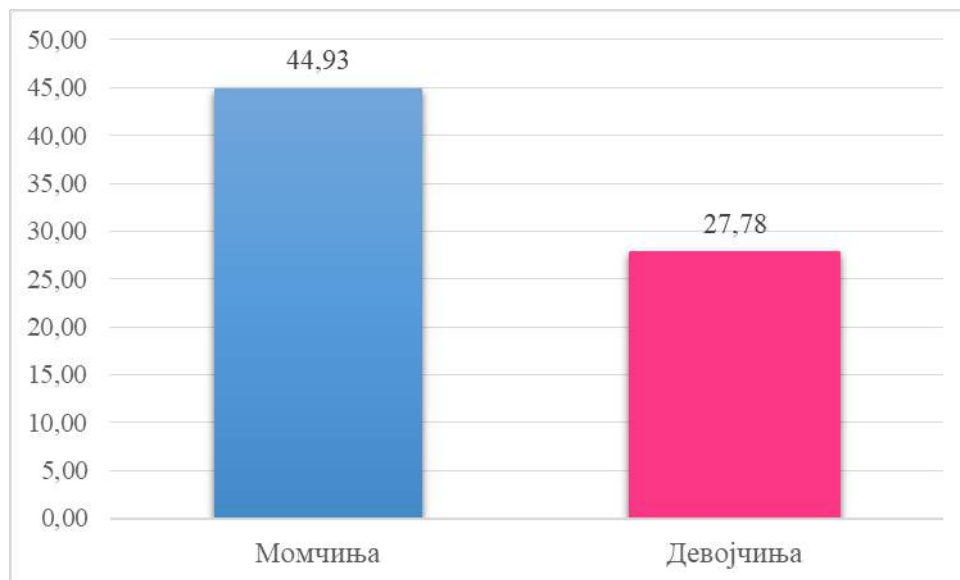
Графикон 28. Преглед на аритметичките средини на моторичкиот тест скок во далечина помеѓу момчињата и девојчињата во првото мерење



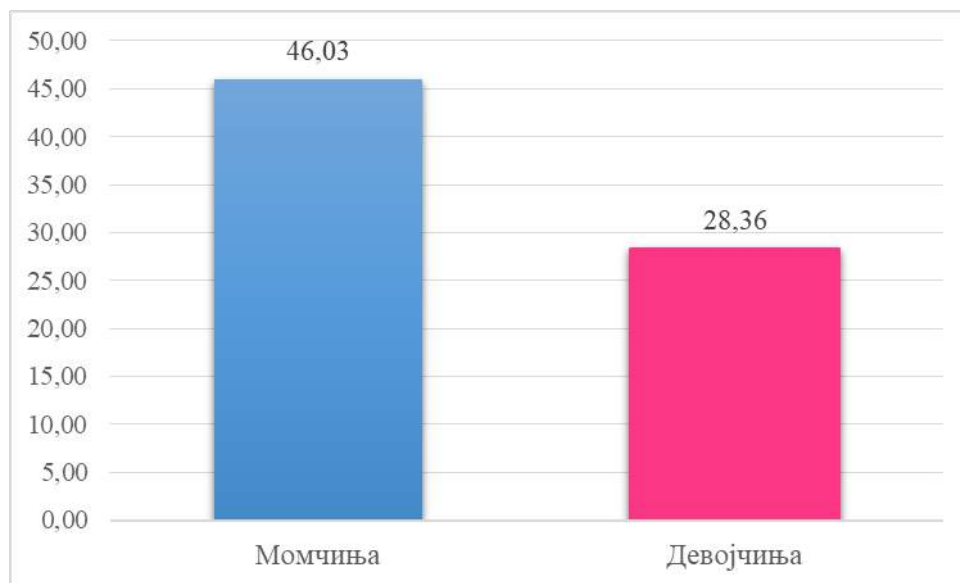
Графикон 29. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на десна рака помеѓу момчињата и девојчињата во првото мерење



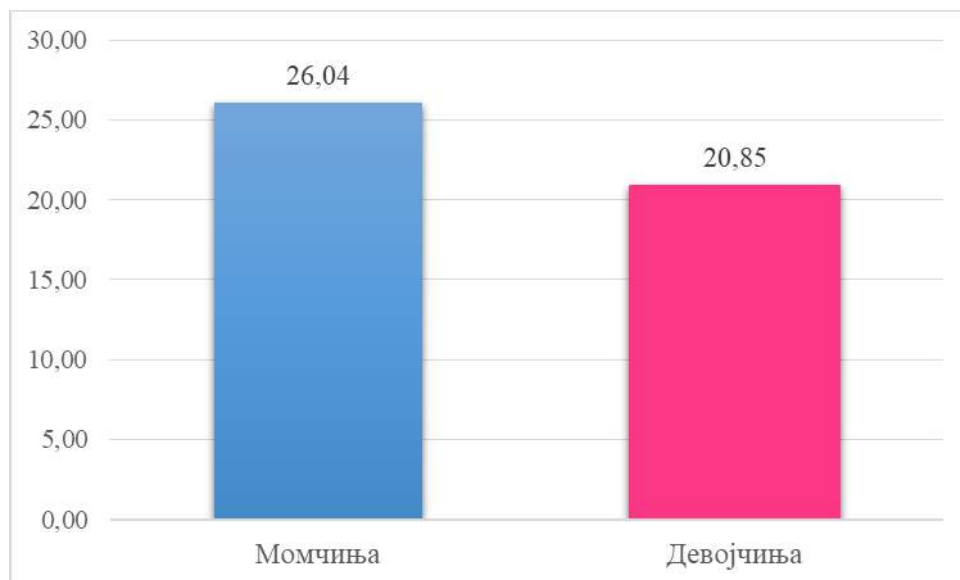
Графикон 30. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на лева рака помеѓу момчињата и девојчињата во првото мерење



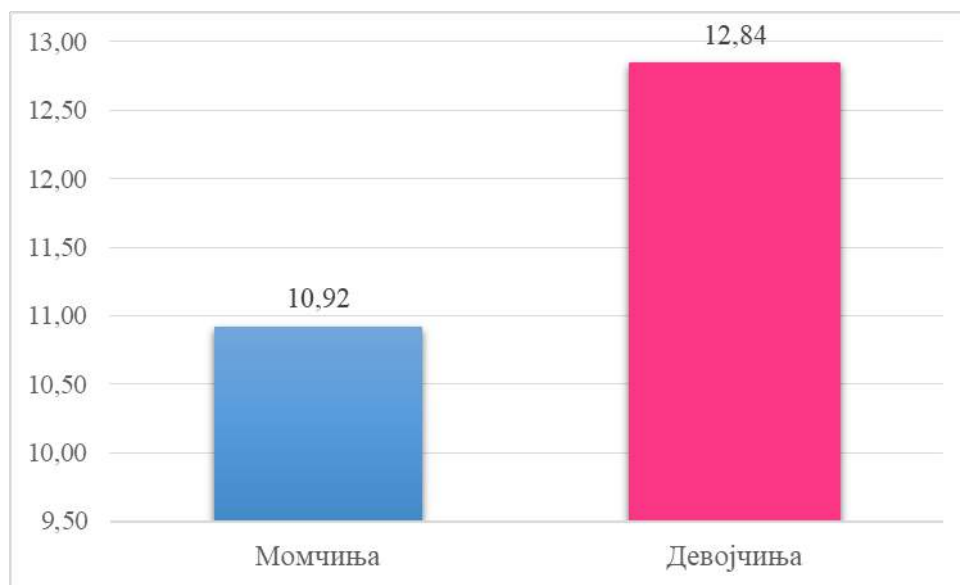
Графикон 31. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка просечна вредност од десна и лева рака помеѓу момчињата и девојчињата во првото мерење



Графикон 32. Преглед на аритметичките средини на моторичкиот тест подигнување труп 30 s помеѓу момчињата и девојчињата во првото мерење



Графикон 33. Преглед на аритметичките средини на моторичкиот тест чуносто 4×10 m помеѓу момчињата и девојчињата во првото мерење



## 5.8. РАЗЛИКИ ВО АНТРОПОМЕТРИСКИТЕ МЕРКИ И МОТОРИЧКИТЕ ТЕСТОВИ ПОМЕЃУ МОМЧИЊАТА И ДЕВОЈЧИЊАТА ВО ВТРОТО (РЕТЕСТ) МЕРЕЊЕ

Мултиваријантната анализа на коваријанса (MANCOVA) беше применета за да се испитаат половите разлики во антропометриските мерки и моторичките тестови во второто мерење (ретест по 7 дена), при што возраста беше вклучена како коваријата со цел да се контролира нејзиното потенцијално влијание врз испитуваните показатели. Мултиваријантниот тест (табела 14) покажа статистички значаен глобален ефект на полот врз збирната варијација на сите вклучени зависни варијабли (Wilks'  $\lambda = 0,26$ ;  $F(10,67) = 19,33$ ;  $p < 0,001$ ;  $\eta^2 = 0,74$ ). Вредноста на Wilks'  $\lambda$ , која е значително пониска од 1, укажува на јасно мултиваријантно раздвојување меѓу момчињата и девојчињата и во ретест услови, додека високата парцијалн ефект  $\eta^2$  (74%) потврдува дека полот и по контрола на возраста објаснува голем дел од варијансата во комбинираниот простор на антропометриските и моторичките показатели. Овој резултат ја оправдува последователната униваријатна анализа на секоја поединечна варијабла.

Униваријатните анализи покажаа дека половите разлики во основните морфолошки показатели остануваат изразени и на второто мерење. Телесната висина повторно беше статистички значајно поголема кај момчињата ( $M = 178,17$  cm;  $SD = 6,72$ ) во споредба со девојчињата ( $M = 164,26$  cm;  $SD = 5,52$ ), со многу висок ефект на полот ( $F = 70,96$ ;  $p < 0,001$ ;  $\eta^2 = 0,48$ ), што упатува дека речиси половина од варијансата во висината може да се припише на половата припадност во рамките на контролирана возраст. Телесната тежина исто така беше значајно повисока кај момчињата ( $M = 69,18$  kg;  $SD = 10,22$ ) отколку кај девојчињата ( $M = 58,30$  kg;  $SD = 7,18$ ), со умерен до висок ефект на влијание ( $F = 20,63$ ;  $p < 0,001$ ;  $\eta^2 = 0,21$ ). Наспроти тоа, индексот на телесна маса не покажа статистички значајни разлики меѓу половите (момчиња:  $M = 21,74$  kg/m<sup>2</sup>;  $SD = 2,54$ ; девојчиња:  $M = 21,61$  kg/m<sup>2</sup>;  $SD = 2,57$ ), при што и F-вредноста и големината на ефектот остануваат занемарливи ( $F = 0,01$ ;  $p = 0,913$ ;  $\eta^2 = 0,00$ ), што укажува дека и покрај повисоката висина и тежина кај момчињата, релативната телесна маса во однос на висината е споредлива меѓу половите. Обемот на половината беше значајно поголем кај момчињата ( $M = 77,60$  cm;  $SD = 6,01$ ) во однос на девојчињата ( $M = 71,43$  cm;  $SD = 5,61$ ), со умерен ефект на полот ( $F = 16,89$ ;  $p < 0,001$ ;  $\eta^2 = 0,18$ ), што укажува на стабилна полова диференцијација на абдоминалната адипозност и во ретест мерењето.

Во доменот на флексибилноста, претклонот во сед не регистрираше статистички значајни разлики меѓу половите (момчиња:  $M = 20,70$  cm;  $SD = 9,71$ ; девојчиња:  $M = 21,90$  cm;  $SD = 7,67$ ), при што ефектот на полот е практично занемарлив ( $F = 0,32$ ;  $p = 0,571$ ;  $\eta^2 = 0,00$ ). Овој наод сугерира дека во услови на краткорочен ретест интервал,

флексибилноста не покажува систематски полово специфичен образец во рамките на анализираниот примерок, за разлика од повеќето силиви и брзинско агилни компоненти.

Кај моторичките тестови поврзани со силата и експлозивноста се задржаа изразени и статистички многу значајни полови разлики. Скокот од место беше значајно подобар кај момчињата ( $M = 204,58$  cm;  $SD = 32,45$ ) споредено со девојчињата ( $M = 153,62$  cm;  $SD = 25,93$ ), со висок ефект на полот ( $F = 41,64$ ;  $p < 0,001$ ;  $\eta^2 = 0,35$ ), што укажува на конзистентна супериорност на момчињата во експлозивната сила на долните екстремитети и во повтореното мерење. Уште понагласени разлики се регистрираа кај мускулната сила на горните екстремитети мерена преку динамометрија. Десната рака покажа највисока F-вредност и најголем ефект во целокупниот сет варијабли (момчиња:  $M = 47,29$  kg;  $SD = 7,92$ ; девојчиња:  $M = 29,20$  kg;  $SD = 6,04$ ;  $F = 96,34$ ;  $p < 0,001$ ;  $\eta^2 = 0,56$ ), што значи дека половата припадност објаснува повеќе од половина од варијансата на оваа мерка. Слично, левата рака (момчиња:  $M = 44,72$  kg;  $SD = 7,72$ ; девојчиња:  $M = 28,41$  kg;  $SD = 6,29$ ) покажува висок ефект на влијание ( $F = 76,42$ ;  $p < 0,001$ ;  $\eta^2 = 0,50$ ), а просечната динамометрија останува исклучително дискриминативна по пол (момчиња:  $M = 46,01$  kg;  $SD = 7,55$ ; девојчиња:  $M = 28,80$  kg;  $SD = 5,93$ ;  $F = 92,54$ ;  $p < 0,001$ ;  $\eta^2 = 0,55$ ). Овие резултати упатуваат на висока стабилност на половата диференцијација во силата проценета преку динамометрија и во ретест услови.

Функционалната сила и силовата издржливост на трупот, проценета преку подигнување труп за 30 секунди, исто така остана значајно повисока кај момчињата ( $M = 27,72$ ;  $SD = 4,30$ ) во споредба со девојчињата ( $M = 22,76$ ;  $SD = 5,51$ ), со умерен ефект на влијание ( $F = 17,45$ ;  $p < 0,001$ ;  $\eta^2 = 0,19$ ), што укажува дека половите разлики во овој моторички тест се релативно стабилни на краток временски интервал. Кај тестот за агилност  $4 \times 10$  m се забележа статистички значајна разлика во корист на момчињата, кои постигнаа пониско време (подобар резултат) ( $M = 10,50$  s;  $SD = 2,14$ ) во споредба со девојчињата ( $M = 12,20$  s;  $SD = 3,10$ ), при што ефектот на полот е мал до умерен ( $F = 7,32$ ;  $p = 0,008$ ;  $\eta^2 = 0,09$ ). Иако разликата останува значајна, големината на ефектот е пониска во однос на повеќето силиви показатели, што може да сугерира поголема варијабилност на резултатите или поголема чувствителност на тестот на ситуациски фактори во ретест услови.

Заклучно, резултатите од Табела 14 потврдуваат дека и во второто мерење полот претставува силен детерминант на мултиваријантниот профил на антропометриските и моторичките показатели, со висока објаснета варијанса на глобално ниво ( $\eta^2 = 0,74$ ). На униваријатно ниво, половите разлики се најизразени кај телесната висина и мерките на силата на дланката (доминација на момчињата со многу високи  $\eta^2$ ), потоа кај експлозивната сила и делумно кај агилноста и силовата издржливост на трупот, додека

ВМІ и флексибилноста остануваат без статистички значајна полова диференцијација. Овие конзистентни обрасци во ретест интервал од 7 дена дополнително упатуваат дека регистрираните разлики не се случајни флукутации на мерењето, туку релативно стабилни полого специфични карактеристики во рамките на анализираната возрасна и функционална структура на примерокот.

Табела 14. Разлики во антропометриските мерки и моторичките тестови помеѓу момчињата и девојчињата во второто (ретест) мерење

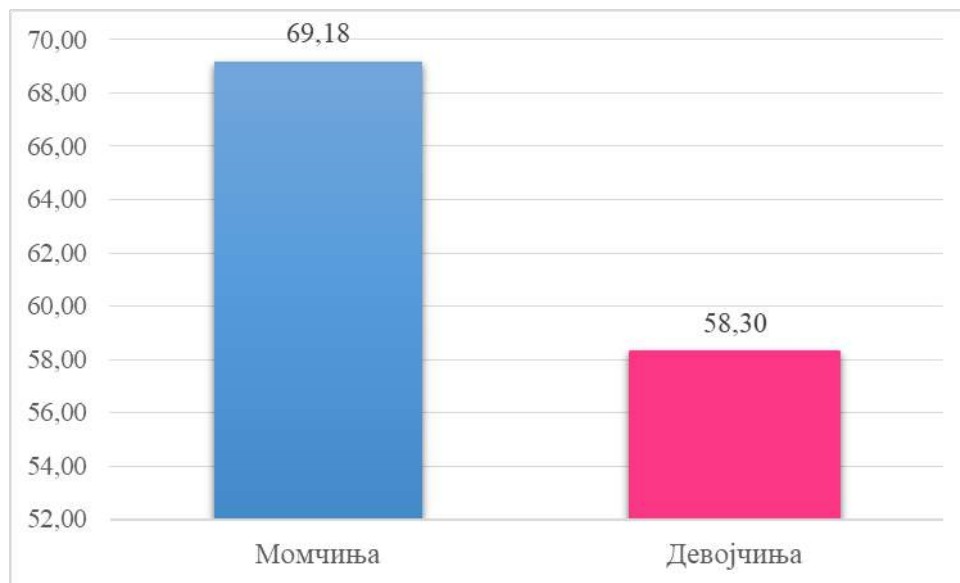
	Value	F	Hypothesis df	Error df	Sig.	$\eta^2$
Wilks' lambda	0,26	19,33	10	67	,000	0,74

Варијабли	Момчиња		Девојчиња		F	p.	$\eta^2$
	Mean	SD	Mean	SD			
Телесна висина (cm)	178,17	6,72	164,26	5,52	70,96	<b>,000</b>	0,48
Телесна тежина (kg)	69,18	10,22	58,30	7,18	20,63	<b>,000</b>	0,21
ВМІ (kg/m <sup>2</sup> )	21,74	2,54	21,61	2,57	0,01	,913	0,00
Обем на половина (cm)	77,60	6,01	71,43	5,61	16,89	<b>,000</b>	0,18
Претклон во сед (cm)	20,70	9,71	21,90	7,67	0,32	,571	0,00
Скок од место (cm)	204,58	32,45	153,62	25,93	41,64	<b>,000</b>	0,35
Динамометрија десна (kg)	47,29	7,92	29,20	6,04	96,34	<b>,000</b>	0,56
Динамометрија лева (kg)	44,72	7,72	28,41	6,29	76,42	<b>,000</b>	0,50
Динамометрија просек (kg)	46,01	7,55	28,80	5,93	92,54	<b>,000</b>	0,55
Подигнување труп 30 s (повт.)	27,72	4,30	22,76	5,51	17,45	<b>,000</b>	0,19
Чунесто 4×10 m (s)	10,50	2,14	12,20	3,10	7,32	<b>,008</b>	0,09

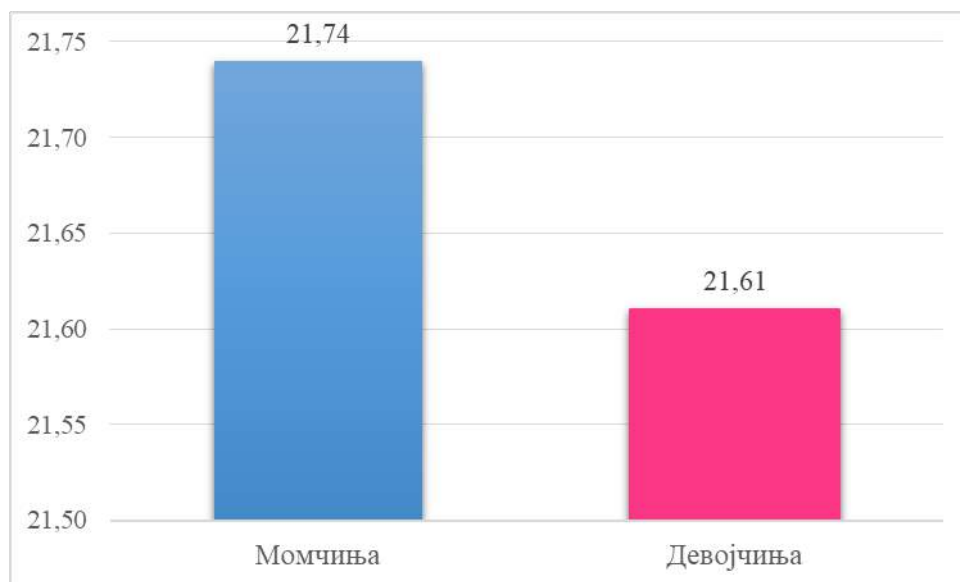
Графикон 34. Преглед на аритметичките средини на антропометриската мерка телесна висина помеѓу момчињата и девојчињата во второто мерење



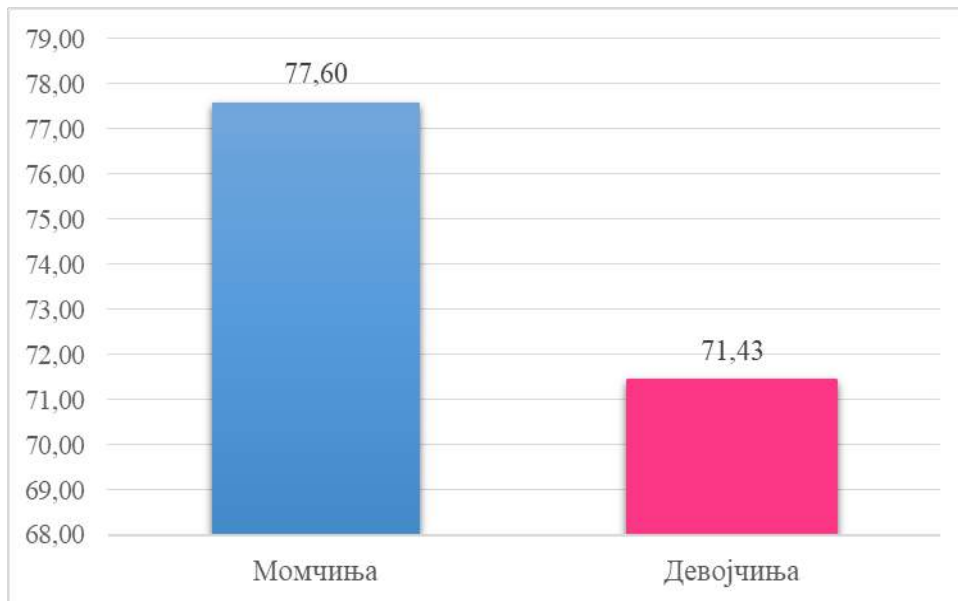
Графикон 35. Преглед на аритметичките средини на антропометриската мерка телесна тежина помеѓу момчињата и девојчињата во второто мерење



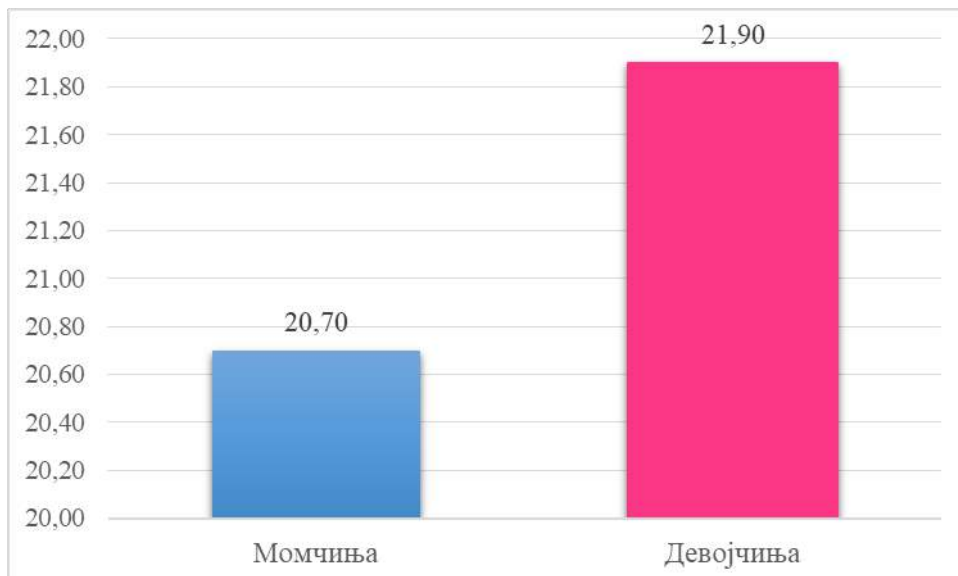
Графикон 36. Преглед на аритметичките средини на антропометриската мерка индекс на телесна маса помеѓу момчињата и девојчињата во второто мерење



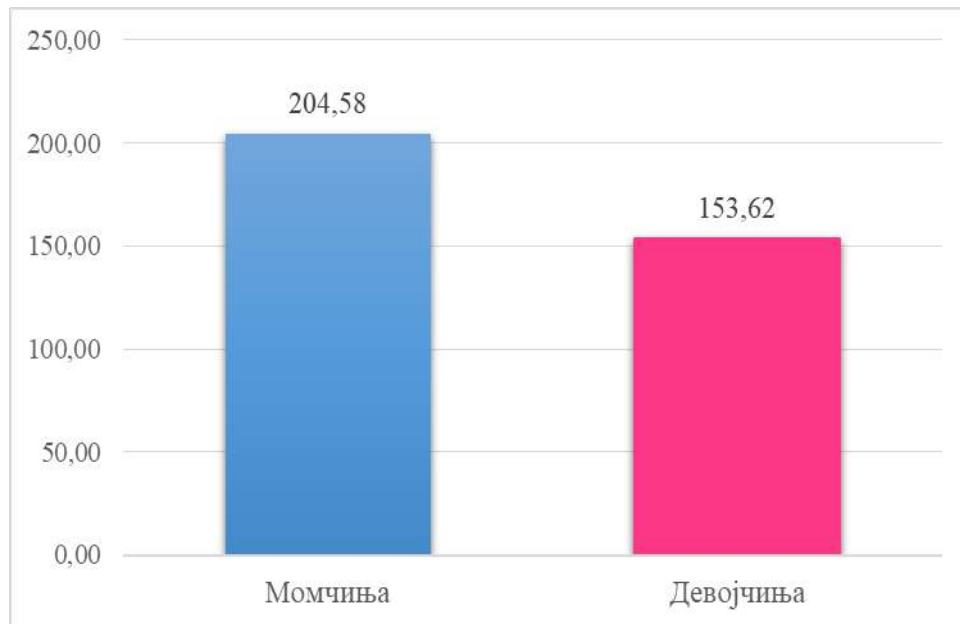
Графикон 37. Преглед на аритметичките средини на антропометриската мерка обем на половината помеѓу момчињата и девојчињата во второто мерење



Графикон 38. Преглед на аритметичките средини на моторичкиот тест претклон во сед помеѓу момчињата и девојчињата во второто мерење



Графикон 39. Преглед на аритметичките средини на моторичкиот тест скок во далечина помеѓу момчињата и девојчињата во второто мерење



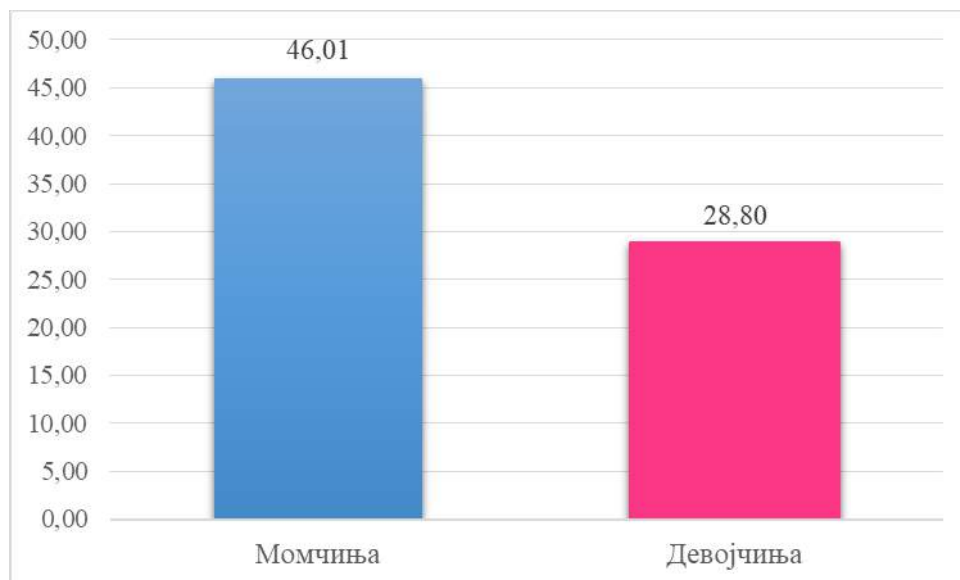
Графикон 40. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на десна рака помеѓу момчињата и девојчињата во второто мерење



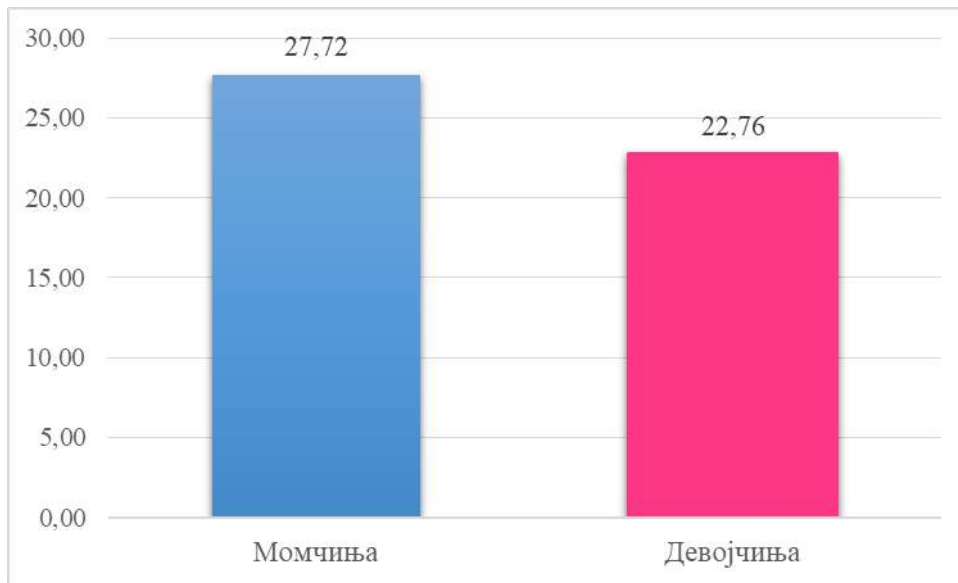
Графикон 41. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка на лева рака помеѓу момчињата и девојчињата во второто мерење



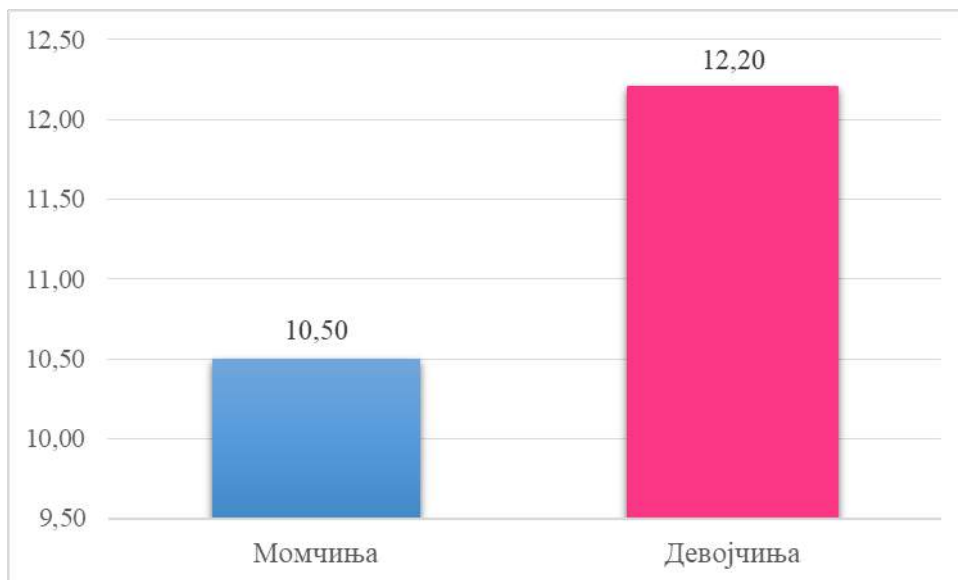
Графикон 42. Преглед на аритметичките средини на моторичкиот тест динамометрија на дланка просечна вредност од десна и лева рака помеѓу момчињата и девојчињата во второто мерење



Графикон 43. Преглед на аритметичките средини на моторичкиот тест подигнување труп 30 s помеѓу момчињата и девојчињата во второто мерење



Графикон 44. Преглед на аритметичките средини на моторичкиот тест чунесто 4×10 m помеѓу момчињата и девојчињата во второто мерење.



## 5.9. ПРИМЕНЛИВОСТА НА ПРЕДЛОЖЕНАТА БАТЕРИЈА МАКФИТ

Применливоста на предложената батерија МАКФИТ беше оценувана преку комбинација од квантитативни индикатори (времетраење на подготовката и реализацијата на тестовите, број на комплетирани мерења) и субјективни проценки на наставниците по физичко и здравствено образование во врска со организациската изводливост, јасноста на протоколите и оптовареноста на учениците. Повеќето показатели за применливост покажаа позитивен одговор во  $\geq 95\%$  од случаите, што укажува дека во голем дел од ситуациите наставниците и учениците не се соочиле со значајни потешкотии при реализирањето на тест-батеријата. Ваквото високо ниво на изводливост е особено важно ако се има предвид дека тестирањето се спроведува во реални училишни услови, во рамки на редовните часови по физичко образование и со ограничени временски и материјални ресурси.

Анализата на времето потребно за подготовка и изведување на тестовите покажа дека максималното време за подготовка на мерењата по ученик изнесувало приближно 3 минути за антропометриската проценка и дополнителни 3 минути за процената на физичката подготвеност, при што во вториот сегмент беа вклучени само тестот за сила на стисокот на дланка и скокот во далечина од место. Просечното време потребно за спроведување на антропометриската проценка беше околу 3 минути по испитаник, од кои приближно 1 минута беше потребна за мерење на телесната тежина и телесната висина, а околу 2 минути за мерење на обемот на половината. Тестот за сила на стисок на дланка траеше приближно 90 секунди по испитаник, додека тестот скок во далечина од место просечно се реализираше за околу 50 секунди. Овие временски рамки покажуваат дека протоколот е компактно структуриран и може да се интегрира во наставниот процес без да го наруши редовниот тек на часот.

Вкупно, просечното време за подготовка и спроведување на целата батерија кај група од 20 ученици изнесуваше околу 1 час и 35 минути, што соодветствува на приближно три наставни часа по физичко образование од околу 45 минути. Овој податок има директна педагошка импликација: целосното тестирање на паралелка во рамки на МАКФИТ може реално да се планира во склоп на краток серијал од два до три последователни часа, без неопходност од дополнителни вонредни термини или специјални организациски интервенции. На тој начин се демонстрира дека батеријата е временски економична и усогласлива со реалната динамика на училиштето, при што се задржува висок степен на стандардизација.

Од аспект на професионалното искуство на наставниците, добиените податоци укажуваат на јасно разграничени области на рутина и области во кои е потребна дополнителна едукација. Претходно искуство во мерење на телесна тежина пријавиле 50% од наставниците (3 наставници), 66,7% (4 наставници) имале искуство во мерење на телесна висина, додека 83,3% (5 наставници) претходно редовно го применувале тестот скок во далечина од место. Притоа, 66,7% (4 наставници) пријавиле искуство со уште еден од стандардните фитнес-тестови вклучени во протоколот. Наспроти тоа, ниту еден наставник не пријавил дека некогаш го проценувал тестот за сила на стисок на дланката, дебелината на кожните набори или обемот на половината. Овие наоди укажуваат дека дел од предложените мерки и тестови се веќе интегрирани во традиционалната наставна практика, додека други претставуваат нови елементи кои бараат насочена обука, стандардизирање и практично увежување.

Ваквата комбинација од висока општа применливост и селективен недостаток на претходно искуство во одредени мерни техники има двоен практичен заклучок. Од една страна, демонстрира дека МАКФИТ батеријата може да се спроведува без поголеми организациски пречки и без прекумерно оптоварување на наставниците и учениците. Од друга страна, ги нагласува приоритетните области за професионален развој, односно потребата од систематска едукација за правилна примена на динамометријата на дланка, антропометриските мерења на обемот на половината и проценката на кожните набори. Со вклучување на овие содржини во програми за континуирана едукација на наставниците и со обезбедување јасни писмени и визуелни протоколи, применливоста на батеријата може дополнително да се подобри и да се намали ризикот од мерни грешки.

Во целина, резултатите на ова истражување упатуваат на тоа дека МАКФИТ батеријата претставува реално изводлива, временски рационална и организациски прифатлива алатка за систематско тестирање на физичкиот и моторичкиот развој кај адолесцентите во училишни услови. Оваа висока применливост е клучен предуслов за нејзина поширока имплементација на национално ниво и за создавање континуиран мониторинг систем кој ќе им овозможи на училиштата, стручните служби и креаторите на политики да располагаат со доверливи податоци за состојбата и трендовите на физичкиот фитнес кај младата популација.

## 5.10. БЕЗБЕДНОСТА НА ПРИМЕНЕТИТЕ АНТРОПОМЕТРИСКИ МЕРЕЊА И МОТОРИЧКИ ТЕСТОВИ

Безбедноста на применетите антропометриски мерења и моторички тестови беше систематски и континуирано следена, со цел да се утврди дали батеријата МАКФИТ може безбедно да се спроведува во редовни училишни услови. Во текот на целокупниот процес на тестирање наставниците по физичко и здравствено образование, заедно со истражувачкиот тим, внимателно ги набљудуваа учениците, бележејќи ги сите евентуални неповолни реакции, како и потенцијални несакани настани поврзани со мерната опрема или со зголемениот физички напор. Особено внимание беше посветено на појави како акутна болка, вртоглавица, нагло влошување на општата состојба, мускулни грчеви, падови, трауматски повреди, како и на можни алергиски реакции при контакт со инструментите за мерење.

Повеќето показатели за безбедност покажаа успешен исход во повеќе од 99% од случаите, што укажува дека тестовите беа генерално добро толерирани од страна на учениците и дека нивната примена не беше поврзана со сериозни ризици по здравјето. Ниту кај еден испитаник не беа регистрирани алергиски реакции на материјалот од мерните инструменти, ниту пак беа евидентирани акутни повреди на мускулно-скелетниот систем (истегнување, дисторзии, падови или удари кои би барале медицинска интервенција). Исто така, не беше забележано ни појавување на синкопа, изразена диспнеја или други алармантни симптоми кои би наложиле прекин на тестирањето или итен медицински третман.

За дополнителна процена на одложените ефекти од физичкиот напор, безбедноста беше проценета и преку прашалник за мускулна болка со одложен почеток (DOMS), кој учениците го пополнуваа два до три дена по завршување на тестирањето. Стапката на одговор на овој прашалник изнесуваше три ученици (2,94%) кои пријавија појава на умерена мускулна болка. Испитаниците претпоставија дека болката е поврзана пред сè со тестот подигнување на трупот за 30 секунди и тестот чуњесто трчање 4×10 метри, а локализацијата на болката беше претежно во пределот на абдоминалната мускулатура и мускулите на долните екстремитети. Овие симптоми беа краткотрајни, не бараа медицинска интервенција и не доведоа до ограничување на секојдневните активности, поради што можат да се сметаат за очекувана и физиолошки вообичаена реакција на зголемен физички напор кај невообичаено оптоварени ученици.

Сумирано, во рамките на спроведеното истражување батеријата МАКФИТ се покажа како високо безбеден инструмент за проценка на физичкиот и моторичкиот развој кај адолесцентите. Отсуството на сериозни несакани настани, минималната зачестеност на лесна и транзиторна мускулна болка и целосното избегнување на алергиски или трауматски реакции, укажуваат дека тестовите можат безбедно да се интегрираат во редовната наставна практика по физичко и здравствено образование, под услов да се почитуваат стандардизираните протоколи, да се обезбеди адекватно загревање и да се следи индивидуалната толеранција на оптоварување кај секој ученик.

## 6. ДИСКУСИЈА

Спроведувањето на две последователни мерења во интервал од една седмица овозможи поцврста и појасна проценка на стабилноста, прецизноста и практичната применливост на употребената батерија тестови, како и на степенот до кој половите разлики се манифестираат во антропометриските и моторичките профили кај адолесцентите, при што се создадоа услови за поуверливо разграничување меѓу вистинските биолошки и функционални карактеристики на испитаниците и варијациите што произлегуваат од мерната грешка, моментната мотивација или ефектот на запознавање со тестовите.

Резултатите покажуваат висока тест-ретест релијабилност на антропометриските показатели и на повеќето моторички тестови: просечните вредности на речиси сите варијабли останаа статистички непроменети помеѓу првото и второто мерење, а корелациите беа високи. Овој наод е конзистентен со досегашните истражувања кои утврдуваат дека теренските фитнес-тестови кај млади (на пр., тестот 4×10 м за агилност, скокот во далечина од место, динамометријата на дланка и сл.) покажуваат висока релијабилност доколку се спроведат во стандардизирани услови (España-Romero et al., 2010). Всушност, тестовите употребени во ова истражување во голема мера се поклопуваат со сетот инструменти препорачани во европската ALPHA-студија како доверливи теренски мерки за проценување на физичкиот фитнес кај млади (España-Romero et al., 2010). Истовремено, во нашето испитување се потврдија и значајни разлики меѓу момчињата и девојчињата во повеќето антропометриски и моторички параметри, што укажува на изразен полов диморфизам во физичките способности во средниот дел од адолесценцијата. Момчињата во просек беа повисоки и потешки од девојчињата, имаа поголем обем на половината и постигнаа значајно подобри резултати во тестовите на мускулна сила и моќ скок во далечина од место, рачен стисок (динамометрија на дланка), силова издржливост на трупот и агилност (4×10 м шатл-тест) во споредба со нивните врстички. Од друга страна, не беа евидентирани статистички значајни полови разлики во флексибилноста (претклон во сед) ниту во односот на телесната маса со висината (BMI), што значи дека индексот на телесна маса беше сличен кај двата пола.

Наративната синтеза на клучните наоди упатува дека применетите мерни инструменти обезбедуваат стабилни и повторливи процени и дека половите разлики во повеќе показатели се во рамки на очекуваните развојни закономерности. Сепак, покрај констатацијата дека разликите постојат, суштински е да се објасни и нивната генеза, односно кои процеси стојат зад ваквиот образец на резултати. Во тој контекст, интерпретацијата на наодите треба да се темели на заемното дејство на биолошките и хормоналните промени во пубертетот, функционалните адаптации на организмот и бихејвиоралните фактори како нивото и типот на физичка активност, мотивацијата и навиките на движење, кои заедно ја обликуваат физичкиот фитнес во адолесценцијата. Пред сè, високата тест-ретест стабилност на повеќето мерења беше очекувана: за една недела не доаѓа до реални биолошки промени кај испитаниците, па така конзистентните вредности главно ја рефлектираат репродуцибилноста на тестовите.

Во нашето истражување повеќето теренски фитнес-тестови покажаа висока тест-ретест релијабилност, што е во согласност со наодите од меѓународните студии на адолесцентска популација. Антропометриските мерки (телесна висина, тежина, ВМІ, обем на половината) беа особено конзистентни при повтореното мерење на пример, висината и тежината постигнаа речиси совршени ICC вредности ( $\approx 0.99$ ), без значајни систематски разлики меѓу тестот и ретестот. Ова се очекувани резултати бидејќи ваквите мерки традиционално покажуваат минимална варијабилност при стандардизирани протоколи. Слично на нашите наоди, Ramírez-Vélez et al. (2015) известуваат дека техничката грешка при антропометриски мерења кај деца и адолесценти е многу мала и конзистентно резултира со  $>95\%$  совпадливост меѓу мерењата. Иако кај обемот на половината забележавме малку понизок ICC ( $\sim 0.95$ ), тоа се вклопува во литературата некои студии бележат ситни, но статистички значајни отстапувања од редоследот, најчесто во рамки од неколку милиметри до половина сантиметар, како резултат на варијации во позиционирањето на мерната лента или респираторниот циклус. На пример, España-Romero et al. (2010) нотираат просечно намалување на обемот на половината од околу 0,3–0,8 cm при второто мерење и покрај тоа заклучуваат дека мерката останува доверлива со соодветна стандардизација. Вкупно земено, високата релијабилност на антропометриските индикатори во нашата студија е во линија со

воспоставените докази дека антропометриските мерки (висина, тежина, ВМІ, обем) се применливи и сигурни за следење на телесниот состав.

Релијабилноста на тестовите за мускулна сила и издржливост исто така беше на високо ниво, но со интересни разлики помеѓу видовите на тестови. Тестот за максимална сила на стисок (динамометрија на дланка) покажа висока стабилност, ICC беше околу 0.93–0.96 без систематска разлика меѓу мерењата што се совпаѓа со универзалните трендови. Истражувањата конзистентно потврдуваат дека тестот за снага на стисок е многу релијабилен кај младите (обично  $ICC > 0.90$ ), со просечни разлики кои статистички не се значајни. Така, Artero et al. (2011) го препорачуваат динамометарскиот тест како стандарден показател за мускулна сила токму поради високата повторливост. Слично на тоа, тестот за експлозивна сила на долните екстремитети скок во далечина од место во нашата студија имаше висок релативен индекс на стабилност ( $ICC \approx 0.91$ ). Меѓутоа, кај овој тест забележавме значајно подобрување на просечниот резултат при ретестот (приближно +6.8 cm во далечина), што укажува на потенцијален ефект на запознавање со задачата. Овој исход е умерено повисок од оној во некои други истражувања каде не се најдени значајни групни разлики. На пример, Ortega et al. (2008) во рамки на HELENA студијата не забележуваат систематски поместувања кај скокот од место просечните разлики меѓу два тестирања изнесувале само неколку милиметри (незначајни -0,3 cm кај машки и +0,3 cm кај женски). Отсуството на просечен напредок во таа европска кохорта сугерира дека, кога протоколите и мотивацијата се оптимални, учесниците даваат максимален напор уште при првиот обид. Во нашиот примерок, пак, подобриот резултат на второто тестирање може да е последица на поголема отпуштеност и техники на изведба или зголемена мотивираност откако учениците стекнале искуство со тестот. Во литературата се документирани вакви 'ефекти на учење' кај теренските фитнес-тестови; Artero et al. (2011) опишуваат студии во кои, и покрај високите коефициенти на корелација меѓу мерењата ( $ICC 0.95–0.97$ ), испитаниците покажале статистички значајни подобрувања при второто тестирање како зголемен број изведени склекови или стомачни подигнувања што укажува на влијанието на претходното искуство со задачата. Овој феномен подразбира дека тестот останува рангирачки доверлив (испитаниците кои биле најдобри остануваат најдобри), но апсолутната

промена кај некои може повеќе да се должи на привикнување отколку на реален краткорочен напредок.

Интересно, најголеми ретест-подобрување во нашата студија беа евидентирани кај тестовите за мускулна издржливост и флексибилност. Кај издржливоста на трупот (подигнување на трупот за 30 s) просечниот скор се подобри значајно за околу 1.7 повторувања, со среден до голем ефект ( $dz \approx 0.8$ ). Ова резултираше со нешто понизок ICC ( $\approx 0.82$ ) и поширок интервал на доверба, што индицира поголема варијабилност меѓу поединците при повторено тестирање. Сепак, ваквото наоѓање не е невообичаено: други истражувања за слични тестови на силова издржливост исто така нотираат подобрен učinok на втората проба. На пример, една студија со адолесценти забележала дека учениците направиле значајно повеќе stomачни стегања на вториот тест, иако вкупната релијабилност била висока ( $ICC \approx 0.95$ ) Artero et al. (2011). Можно објаснување е дека испитаниците при првото мерење не го достигнуваат својот реален максимум поради неоптимална техника или недоволна запознаеност со темпото на тестот, па вториот пат се во состојба да реализираат поголем број повторувања. Слична шема се забележува и кај тестот на флексибилност (претклон во сед). Ние регистриравме просечно подобрување од околу 2.5 cm при повторното мерење, што претставува статистички значајна разлика. Иако релативната релијабилност остана висока ( $ICC \approx 0.89$ ), јасно е дека испитаниците во просек успеале да се наведнат подлабоко на вториот обид веројатно благодарение на подобра претходна подготовка или елиминација на почетната вкочанетост и несигурност. Ваквите умерени подобрувања се во склад со сознанијата од други студии. Ramírez-Vélez et al. (2015) утврдуваат дека „sit-and-reach“ тестот кај колумбиските ученици има мал систематски ефект: приближно 1% поголема амплитуда на второто тестирање. Наспроти тоа, голема европска студија (HELENA) не пронашла значаен просечен поместување кај овој тест, што укажува дека соодветното загревање и контролирани услови може да го намалат ефектот на запознавање. Во контекст на нашиот примерок, веројатно е дека дел од учениците при првата сесија не ја постигнале својата оптимална флексибилност (можеби поради недоволна загреаност или трема), додека при ретестот биле поопуштени и поподготвени да го дадат максималниот дострел. Ова го потенцира значењето на стандардизираното загревање и пробно

покажување на задачата пред првото мерење, со цел да се минимизираат ваквите артефакти при проценка на вистинските промени во флексибилноста.

Тестот за агилност и брзина на промена на насока (чунско трчање 4×10 метри) се истакна како еден од најстабилните во нашето истражување. Вредностите на повторливост беа екстремно високи ( $ICC \approx 0.99$ ), а разликата помеѓу просечните времиња на двата теста беше статистички незначајна (практично нулта пристрасност). Овој резултат е во целосна согласност со извештаите од литературата, каде 4×10 m шатл-тестот важи за многу релијабилна мерка на моторичките способности кај децата и младите. Така, Ortega et al. (2008) и Artero et al. (2011) посочуваат дека овој тест покажува силна меѓусесиона стабилност, без трендовски промени меѓу обидите. Нашите податоци го потврдуваат тоа агилноста на учениците остана практично непроменета на групно ниво, што имплицира дека ефектот на учење или замор е минимален во ваков краток, анаеробен тест. Високата сензитивност и истовремена сигурност на 4×10 m трчањето го прави погоден и практичен за следење на брзинско-агилните способности во училишна средина, бидејќи дури и мали подобрувања над половина секунда би биле над прагот на нормалната варијабилност.

Иако во нашата студија немаше посебен теренски тест за максимална кардиореспираторна издржливост (со оглед на ограничувањата да се спроведе тешкиот 20-метарски шатл-тест повеќепати во краток рок), вредно е да се стави нашиот труд во контекст на наодите од други истражувања за аеробната фитнес-компонента. Постојните докази сугерираат дека теренските тестови за кардиореспираторна издржливост се репродуктивни доколку се применат соодветни протоколи. На пример, 20-метарскиот шатл-тест (т.н. „beer“ тест) генерално покажува висока повторливост кај адолесцентите – неколку студии известуваат за непостоечки или минимални систематски разлики меѓу два последователни обиди, со вредности на корелација што најчесто спаѓаат во опсегот 0.89–0.94. Во големата европска HELENA студија, ниту еден ефект на учење или замор не бил детектиран при ретестирање на шатл-тестот по две недели, што значи дека просечниот максимален број на постигнати циклуси останал статистички ист. Слично на тоа, мета-анализите укажуваат дека тестот на трчање на 1 km или тестот на трчање на една милја (1609 m) исто така може да биде доверлив во училишна популација, иако варијабилноста кај овие тестови е нешто поголема отколку кај шатл-тестот.

Овие сознанија ја поткрепуваат претпоставката дека, дури и без директно повторување на аеробниот тест во нашето истражување, може да се очекува висок степен на релијабилност на кардиореспираторните показатели доколку се спроведат во контролирани услови со соодветно време за опоравување на испитаниците. (Mayorga-Vega et al., 2016)

Накратко, споредбата на нашите резултати со достапните меѓународни истражувања покажува повеќе сличности отколку разлики во однос на релијабилноста на теренските фитнес-тестови. Кај сите основни компоненти телесен состав, кардиореспираторна издржливост, мускулна сила/издржливост, флексибилност и агилност, брзината и координацијата забележана е висока конзистентност на измерените вредности, што сугерира дека избраните тестови се соодветни за доверливо следење на физичката подготвеност кај адолесцентите. Разликите што ги воочивме (на пример, систематско подобрување во одредени перформанси при ретест) најверојатно се должат на ефекти на запознавање со тестот и мотивациски фактори, а не на непредвидлива грешка или нестабилност на самите инструменти. Овие наоди се конгруентни со извештаите на други автори кои нотираат дека некои теренски-тестови бараат претходна обука или демонстрација за да се минимизираат варијациите при повторено мерење. Дополнително, контекстот на мерењето може да игра улога кога тестирањето се изведува во училишни услови од страна на наставници, најчесто се добиваат слично сигурни резултати, но строгата контрола на протоколот (време на ден, стандардизирано загревање, поттикнување на максимален напор) може да ги намали суптилните отстапувања. Нашиот примерок се состоеше од постари адолесценти (~16–17 години) со релативно хомогени карактеристики, што веројатно допринесе кон високата релијабилност, бидејќи познато е дека во овој возрасен период варијабилноста поради раст и развој е помала, а способноста за репродуцирање на максимален напор е поголема. Во студиите кои опфаќаат помлади испитаници понекогаш се забележуваат нешто поголеми варијации, но сепак повеќето истражувања сугерираат дека правилно избраните теренски тестови можат со доверба да се користат за проценка на физичката способност кај децата и младите.

Во целина, нашата споредбена анализа ја потврдува валидноста на пристапот за примена на едноставни, евтини теренски мерења, но и укажува на потребата резултатите секогаш да се интерпретираат со оглед на можните мали ефекти од повторувањето особено кај тестовите каде забележавме систематски подобрувања. Тоа значи дека при следење на индивидуалниот напредок, покрај релативните показатели (на пример, ранг-скор според врсници), полезно е да се разгледа и апсолутната промена во однос на границите на доверливата разлика (на пример, SEM, MDC) за да се разликува вистинскиот напредок од варијацијата на мерката. Соодветно, комбинацијата на нашите наоди со постоечките докази дава цврста основа дека применетата батерија „МакФит“ е релијабилна алатка за проценка на здравствено-релевантната физички фитнес кај адолесцентите, при што добиените вредности се споредливи со оние во меѓународни рамки и можат да послужат за понатамошни истражувања и практични интервенции.

Добиените полови разлики во антропометријата и моторичките способности се во согласност со биолошките сознанија за пубертетскиот развој и половата диференцијација. Познато е дека уште од самиот почеток на пубертетот телата на момчињата и девојчињата почнуваат да се развиваат во различни насоки под влијание на хормоните: машкиот пубертет (воден од тестостеронот) стимулира значително зголемување на безмастата телесна маса мускулите и коските додека женскиот пубертет (под влијание на естрогените) резултира со поголемо таложење поткожни масти и поинаква телесна дистрибуција на масата (Lomba-Albrecht & Styne, 2009). Како последица на тоа, до средината на адолесценцијата момчињата во просек стануваат повисоки и посилни од девојчињата, при што разликите се особено изразени во способностите поврзани со мускулна сила, моќ и брзина (мошне помали или никакви разлики се забележуваат во аеробниот капацитет пред доцниот пубертет и во флексибилноста) (Joynes, Hunter & Senefeld, 2025). Нашите резултати ја рефлектираат токму таа слика. Момчињата беа во просек ~14 cm повисоки од девојчињата и имаа ~11 kg поголема телесна тежина, што е очекувано со оглед на подолгиот период на раст кај машкиот пол (машките испитаници беа на ивица на 17 години, возраст кога повеќето сè уште се во фаза на интензивен линеарен раст, додека девојчињата вообичаено завршуваат со растот неколку години порано). Иако момчињата имаа поголем обем на половината, кај нас се покажа дека индексот на телесна маса (BMI) практично не се

разликува меѓу двата пола просечно изнесува  $\sim 21,7 \text{ kg/m}^2$  во двете групи што сугерира дека разликата во тежината е пропорционална на разликата во висината. Сепак, тоа не значи дека телесниот состав е ист: момчињата веројатно имаат поголем процент на мускулна маса, а девојчињата нешто повисок процент масти при ист BMI, што е типичен исход на пубертетските хормонални влијанија (Loomba-Albrecht & Styne, 2009). Навистина, во научната литература се опишува дека до доцна адолесценција разликите во мускулната сила се огромни кај возрасните мажи мускулната сила во горните екстремитети е  $\sim 50\text{--}60\%$  поголема отколку кај жените, а во долните  $\sim 30\text{--}40\%$  (Nuzzo & Pinto, 2025). Кај адолесцентите на возраст 14–17 години, мета-анализите покажуваат дека момчињата се во просек околу 50% посилни во горниот дел на телото и  $\sim 30\%$  во долниот дел во споредба со девојчињата исти години (Nuzzo & Pinto, 2025). Овој диспарат се должи на две компоненти: (1) биолошката - момчињата имаат поголема мускулна маса, поголеми и посилни мускулни влакна (особено брзи мускулни влакна стимулирани од тестостеронот), поголема срцева и белодробна капацитетност, како и повисок хемоглобин (што им дава предност во издржливите активности); и (2) бихејвиоралната – статистички гледано, машките адолесценти почесто се занимаваат со спортови и активности кои ја развиваат силата, експлозивноста и брзината, додека девојчињата, особено во подоцниот пубертет, покажуваат тренд на намалување на физичката активност или пак, се насочени кон активности кои не секогаш ја одржуваат мускулната маса (на пример, повеќе се фокусираат на естетски спортови, танц, рекреативни активности). Нашите податоци го потврдуваат тоа: во тестовите за максимална сила на горните екстремитети (сила на стисок измерена со динамометар) момчињата постигнаа над 60% повисоки вредности од девојчињата (46 kg спроти 28–29 kg просечно), што е разлика со многу голема големина на ефект ( $\eta^2 > 0,5$ ) и имплицира дека полот објаснува повеќе од половина од варијансата на рачната сила во нашата хомогена старосна група. Слично високи и статистички сигурни разлики се забележани и кај експлозивната сила на долните екстремитети: момчињата во просек достигнаа близу 200 cm во скок во далечина од место, додека девојчињата  $\sim 151$  cm, што значи дека машките испитаници имаат за околу една третина подобар резултат во овој тест податок кој одговара на глобално објавените нормативни вредности за оваа возраст. Треба да се истакне дека иако девојчињата вообичаено ги надминуваат момчињата во

флексибилност (особено во пред и рана пубертетска возраст), во нашата студија таа разлика не беше статистички значајна: просекот на претклонот во сед беше многу сличен (19,2 cm кај девојчињата и 18,4 cm кај момчињата во првото мерење, а ~21 cm спрема ~18,8 cm во второто мерење). Ваквиот исход може да се должи на големата варијабилност на флексибилноста внатре во секој пол ( $CV \approx 40\text{--}50\%$  кај двата пола), што значи дека има широк опсег, односно некои поединци (и момчиња и девојчиња) се екстремно нефлексибилни, додека други се многу флексибилни, па просеците не се драстично различни. Воедно, можно е девојчињата од овој возрасен опсег (16–17 години) да немаат толкава предност во флексибилноста како што би имале на помала возраст делумно затоа што дел од нив веќе не се физички активни во дисциплини што ја одржуваат флексибилноста, а дел затоа што момчињата кои се спортски активни можеби ја подобриле својата подвижност со тренинзи (на пример, преку спортови кои бараат истегнување и подвижност). Независно од причините, фактот што флексибилноста е изедначена сугерира дека тука полот нема систематски ефект, што го потврдуваат и статистичките анализи (незначајни F-тестови,  $\eta^2$  близу 0).

Оваа комбинација на наоди, висока релијабилност на тестовите и значајни полови разлики има голема практична и теоретска важност. Од практичен аспект, резултатите ја нагласуваат употребливоста на применетиот батериски тестови во образовен и спортски контекст. Ако наставниците по физичко образование или спортските тренери ги спроведуваат овие мерења со стандарден протокол, можат да бидат уверени дека добиените вредности се доверливи и повторливи, па со тоа малите промени низ време претставуваат реална повратна информација, а не артефакт на мерењето. Ова е особено значајно за мониторирање на физичката форма кај младите. Многу земји веќе воведуваат стандардизирани фитнес-батерии во училиштата, бидејќи физичката подготвеност во детството и адолесценцијата е силен показател за здравјето и поврзана е со ризик-факторите за идни заболувања. Нашето истражување го потврдува овој пристап, бидејќи покажува дека применетите теренски тестови за сила, издржливост, брзинско-агилни способности и телесни мерки даваат стабилни и повторливи резултати, односно при повторно тестирање по една седмица не се појавуваат значајни отстапувања што би укажале на голема мерна грешка. Всушност, најдената релијабилност кај нас (ICC најчесто 0,90–0,99) е споредлива со онаа пријавена во контролирани студии на

меѓународно ниво (на пример, Vanhelst et al., 2016; Ramírez-Vélez et al., 2015), што имплицира дека и во наши услови може да се спроведува систематски мониторинг на физичката подготвеност. Притоа, треба да се води сметка за неколку практични детали: (1) Кај одредени тестови со изразен ретест-ефект (на пример, скок од место, претклон во сед, подигнување на трупот за 30 секунди), пожелно е да се обезбеди претходно запознавање на учениците со методологијата на пример, да се даде еден пробен обид за да се намали влијанието на факторот на учење. Доколку тоа не е можно, тогаш при споредба на резултати (особено во краток период) треба да се има предвид дека второто тестирање речиси сигурно ќе даде подобар резултат само поради запознаеност, па наставниците/истражувачите треба да разликуваат вистински напредок од статистички артефакт. (2) Половите разлики воочени кај нас имаат импликации за пристапот во наставата и тренингот: девојчињата во адолесценција заостануваат зад момчињата главно во силовите и брзинските способности, па препорачливо е во програмите за физичко воспитување да се вклучат целни содржини за развој на мускулната сила и експлозивност кај женските ученици (на пример, соодветни вежби за сила со сопствена тежина или со справи, плиометриски вежби, спринтови и сл.). Ова не само што би ја намалило разликата во перформансите, туку би имало и позитивен ефект врз здравјето бидејќи истражувањата сугерираат дека подобрата мускулна фитнес кај девојките е поврзана со понизок кардиометаболен ризик, подобра густина на коските и помал ризик од повреди (Ortega et al., 2008; Smith et al., 2014). Од друга страна, момчињата покажаа послаби резултати во флексибилноста, па наставниците можат да посветат внимание на вежби за истегнување и мобилност кај машките ученици – нешто што можеби природно е повеќе застапено во активностите популарни кај девојчињата (танц, гимнастика, јога). На тој начин, и двата пола би добиле посеопфатен развој на физичката форма. Теоретската значајност на наодите лежи во потврдувањето на неколку фундаментални концепти во развојната кинезиологија. Прво, добиваме локална емпириска потврда за универзалноста на пубертетскиот полов диморфизам: дури и во контролирана старосна група (16–17 години), полот е фактор што објаснува голем дел од варијабилноста во силата, моќта и брзината. Ова оди во прилог на глобалните статистики и ја нагласува улогата на ендокриниот статус во трансформацијата на телото во адолесценцијата. Нашите резултати придонесуваат и кон дискусијата за развојот на физичките

перформанси: гледаме дека во кус временски период (една недела) способностите се многу стабилни, што значи дека кај оваа возрасна група ненадејни скокови или падови во перформансот би укажувале на некаква интервенција или промена (тренинг, болест, повреда и сл.), а не на природни осцилации. Од друга страна, пак, претпоставуваме дека на подолг рок (неколку години) истите овие млади ќе покажат тренд на подобрување, особено момчињата, како резултат на доцниот пубертетски развој факт кој го потврдуваат надолжни студии кои бележат континуиран пораст на аеробните капацитети и мускулната сила кај момчињата до крајот на тинејџерскиот период, додека кај девојчињата некои параметри стагнираат или дури опаѓаат по средниот пубертет (на пример, аеробната издржливост) поради хормонални и телесносоставни промени (Armstrong & McManus, 2011; García-Artero et al., 2008). Во тој контекст, нашата студија претставува основа за идни споредби и внесува релевантни локални податоци во пошироката научна база, со што може да се помогне во изработка на соодветни норми на физичка подготвеност за македонската популација на адолесценти. Ова е во согласност со препораките во литературата дека е потребна стандардизација и консензус при проценката на физичката форма кај младите во школски услови, за резултатите да бидат споредливи и употребливи.

Истражувањето се одликува со повеќе методолошки предности. Прво, примерокот беше релативно хомоген во однос на возраста и беше опфатен доволен број испитаници од машки и женски пол, што овозможи статистички моќни споредби. Второ, мерењата беа спроведени со строго стандардизирани протоколи од страна на стручен тим, при што контролирани беа надворешните услови (време на ден, инструменти, инструкции за изведба), а тоа придонесе мерната грешка да се сведе на минимум. Трето, анализата на податоците опфати повеќе аспекти од класични тестови на разлика, до метрики за релијабилност (ICC, SEM, MDC) и графички приказ (Bland-Altman) со што добиените заклучоци се поткрепени од различни статистички агли. Конечно, една од значајните предности е и примената на тест-ретест дизајн во краток рок: ова овозможи разграничување на трајните карактеристики од случајните варијации и издвојување на ефектот на учење, што не би било возможно со еднократно мерење.

Сепак, постојат и одредени ограничувања кои треба да се имаат предвид при толкување и генерализирање на резултатите. Прво, испитаниците во студијата беа

средношколци на возраст ~16–17 години, што значи дека добиените резултати важат претежно за доцна адолесцентна популација. Не можеме непосредно да ги проектираме овие наоди на помлади адолесценти или деца, бидејќи е познато дека во раните пубертетски фази величината на половите разлики е помала, а исто така и способноста за повторување на тестовите може да биде различна (на пример, помалите деца понекогаш покажуваат поголема варијабилност поради непредвидливиот степен на ангажираност и разбирање на инструкциите). Идни истражувања би требало да опфатат поширок возрасен дијапазон пред-пубертетски деца, рани тинејџери за да се мапира континуумот на промените. Второ, иако во нашата студија хронолошката возраст беше тесно контролирана, не беа директно измерени индикаторите на биолошката зрелост (на пример, стадиумите на пубертет според Танер или концентрацијата на половите хормони). Можно е дел од варијабилноста внатре во секој пол да се должи токму на различното темпо на созревање на пример, некои од 16-годишните момчиња може штотуку влегуваат во касен пубертет и нагло добиваат сила и маса, додека други веќе го минале врвот на раст и влегуваат во постпубертетска фаза. Истото важи и за девојчињата кај кои индивидуалните разлики во менархе и пубертетски статус би можеле да влијаат на перформансите. Отсуството на такви податоци претставува ограничување, бидејќи не можевме да испитаме дали дел од половите разлики или варијансата се објаснуваат со тој фактор. Трето, ретест интервалот од 7 дена, иако идеален за проценка на краткорочна релијабилност, не ни дава информации за долгорочната стабилност на рангирањето. Можно е некои испитаници да покажат значителни промени по неколку месеци, било спонтано (со растот) или поради надворешни влијанија (тренинг, промена на навики), па за проценка на трајната стабилност би биле потребни подолги надолжни следења. Дополнително, нашата проценка на релијабилноста се однесува на интра-испитувачка конзистентност (исти мерители ги спроведоа двата теста). Во практична ситуација во училиштата, можно е различни наставници да администрираат тестирања и прашање е дали тоа би внело поголема варијабилност. Сепак, охрабрувачки се наодите од студии како FUPRECOL (Колумбија), каде што тестирањата ги спроведувале наставници по физичко воспитување и била потврдена висока усогласеност техничката грешка на мерење за антропометријата била минимална, а останатите тестови немале значајни систематски отстапувања меѓу различни испитувачи. Оттука, може да се заклучи дека

со добро стандардизиран прирачник и обука и повеќе лица-тестери би се постигнувале споредливи резултати. Четврто ограничување е можноста за selection bias во примерокот, иако учениците беа избрани случајно од неколку училишта, сепак во студијата учествуваа оние кои дадоа согласност и се појавија на двете мерења. Можно е физички поактивните и мотивирани адолесценти почесто да се согласат на вакви тестирања, што би значело дека општото ниво на фитнес во популацијата би можело да биде малку пониско од измереното. Сепак, овој ефект тешко дека би влијаел на релијабилноста или трендовите на разлика, туку повеќе на апсолутните просечни вредности. И конечно, ограничување претставува и фактот што батеријата на тестови, иако релативно широка, не ги опфати сите димензии на физичката способност. На пример, кардиореспираторната издржливост (аеробниот фитнес) не беше тестиран со повторување популарниот 20 м шатл-тест (beep-test) беше спроведен само еднаш поради напорноста на тестот. Затоа немаме податок за неговата непосредна релијабилност во овој примерок, а од литературата е познато дека издржливостните тестови во зависност од мотивацијата можат да покажат нешто поголема варијабилност отколку, да речеме, спринтерските тестови. Истражувањето на Vanhelst et al. (2016) во Франција, на пример, покажа дека кај училишните деца тестот трчање на 1/2 милја има умерена повторливост (ICC ~0,66) за разлика од останатите тестови кои беа >0,90, веројатно поради потешкотијата учениците двапати да одржат идентично темпоо на трчање. Ваквите сознанија не укажуваат на невалидност, туку на потреба од внимателност при спроведување и интерпретација на одредени тестови кои подразбираат максимален напор.

Врз основа на сето досега дискутирано, произлегуваат неколку насоки за понатамошни истражувања и применети активности. Прво, би било корисно да се прошират ваквите тест-ретест студии и на други возрасни групи – особено на претпубертетска и рано-пубертетска возраст за подобро да се разбере како релијабилноста и половите разлики еволуираат со созревањето. Надолжни истражувања кои би ги следеле истите испитаници од, да речеме, 12 до 18 години, би дале непроценлив увид во времето на појава и проширување на половиот јаз во физичките способности, како и во тоа колкава е природната варијација кај секој поединец со текот на растот. Особено значајно би било во тие студии да се вклучат и објективни мерила за

пубертетски стадиум (хормонални анализи, индекси на зрелост или проценка на скелетна возраст), за да се изолира ефектот на календарската возраст од ефектот на биолошката зрелост. Второ, идни истражувања би можеле да тестираат интервенции насочени кон намалување на воочените разлики: на пример, програми за јакнење на мускулната сила кај девојчиња или програми за зголемување на флексибилноста кај момчиња, проследени со мерења на ефектот. Така би се утврдило во колкава мерка ваквите тренинзи можат да го подобрат учинокот и дали постојат можни „плафони“ на напредок различни за двата пола. Трето, препорачливо е во рамки на секое идно тестирање да се имплементира протокол за запознавање со тестовите – на пример, кратка обука или демонстрација и пробни обиди и тоа да се документира, за да може количествено да се оцени неговото влијание врз намалување на ретест-пристрасноста. Нашите наоди јасно посочуваат дека без таква фамилијаризација, одредени тестови носат систематска грешка при повторување, па идните студии би требало да испитаат начини за корекција на резултатите (на пример, употреба на коригирачки фактори за второто мерење) или пак за калибрирање на тестот преку повеќе повторувања. Четврто, треба да се испита и долгорочната одржливост на мерните инструменти не само во период од една седмица, туку и во подолг период без значителен тренинг. Ваквите податоци би ни кажале колку мерењата се чувствителни на реални промени во статусот на физичката форма, наспроти краткорочни флукуации. Конечно, препорачуваме идните истражувања да вклучат и повеќе показатели на здравствено-ориентирана физичка спремност, како што е кардиореспираторната издржливост. Иако во оваа студија фокусот беше ставен на антропометријата, мускулната сила/моќ и флексибилноста, за заокружување на сликата би било полезно да се анализира и тестот како 20 m шатл-рун (beep-test) во контекст на пол, пубертет и релијабилност. Спојувајќи ги сите овие сегменти, би се добил сеопфатен увид во физичката подготвеност на младата популација. Тоа би придонело не само научно – во разбирање на човечкиот развој туку и практично, преку градење бази на податоци за норми и референтни вредности. Меѓународните напори за воспоставување стандарди (како референтните табели за фитнес на европско ниво на пример) ќе имаат поголема вредност доколку се поткрепени со локални истражувања. Нашето истражување може да послужи како една алка во тој синџир на знаење, бидејќи ги потврди валидноста и применливоста на

ALPHA-фитнес тестовите во македонски услови и даде конкретни бројки за мерната грешка и варијабилноста што треба да се земат предвид. Во иднина, интеграцијата на ваквите податоци во образовниот систем (преку редовни тестирања на учениците) би овозможила креирање политики и програми насочени кон унапредување на физичкото здравје на младите со почитување на половите специфики и потреби.

Во целина, дискутираните наоди ја истакнуваат сигурноста и значењето на спроведените антропометриски и моторички мерења. Тест-ретест анализата потврди дека применетата батерија тестови е високо релијабилна и може да открие дури и мали промени надвор од границите на мерната грешка. Притоа, евидентните полови разлики во повеќето показатели го нагласуваат влијанието на биолошкиот пол и пубертетските процеси врз физичката подготвеност. Овие разлики треба да бидат признати и адресирани во програмите за физичко образование, да се работи на слабостите на секој пол, без притоа да се запостави индивидуалниот пристап (бидејќи преклопувања помеѓу половите постојат). Истражувањето дава основа за понатамошни студии и интервенции кои би се надоврзале на поставените прашања, со крајна цел подобро разбирање на развојот на младите и унапредување на нивното здравје преку физичка активност.

## ПРИМЕНЛИВОСТ НА ПРЕДЛОЖЕНАТА БАТЕРИЈА МАКФИТ

Применливоста на предложената батерија МАКФИТ беше оценувана преку комбинација од квантитативни индикатори (времетраење на подготовката и реализацијата на тестовите, број на комплетирани мерења) и субјективни проценки на наставниците по физичко и здравствено образование во врска со организациската изводливост, јасноста на протоколите и оптовареноста на учениците. Повеќето показатели за применливост покажаа позитивен одговор во  $\geq 95\%$  од случаите, што укажува дека во голем дел од ситуациите наставниците и учениците не се соочиле со значајни потешкотии при реализирањето на тест-батеријата. Ваквото високо ниво на изводливост е особено важно ако се има предвид дека тестирањето се спроведува во реални училишни услови, во рамки на редовните часови по физичко образование и со ограничени временски и материјални ресурси.

Анализата на времето потребно за подготовка и изведување на тестовите покажа дека максималното време за подготовка на мерењата по ученик изнесувало приближно

3 минути за антропометриска процена и дополнителни 3 минути за процената на физичката подготвеност, при што во вториот сегмент беа вклучени само тестот за сила на стисокот на дланка и скокот во далечина од место. Просечното време потребно за спроведување на антропометриската процена беше околу 3 минути по испитаник, од кои приближно 1 минута беше потребна за мерење на телесната тежина и телесната висина, а околу 2 минути за мерење на обемот на половината. Тестот за сила на стисок на дланка траеше приближно 90 секунди по испитаник, додека тестот скок во далечина од место просечно се реализираше за околу 50 секунди. Овие временски рамки покажуваат дека протоколот е компактно структуриран и може да се интегрира во наставниот процес без да го наруши редовниот тек на часот.

Вкупно, просечното време за подготовка и спроведување на целата батерија кај група од 20 ученици изнесуваше околу 1 час и 35 минути, што соодветствува на приближно три наставни часа по физичко образование од околу 45 минути. Овој податок има директна педагошка импликација: целосното тестирање на паралелка во рамки на МАКФИТ може реално да се планира во склоп на краток серијал од два до три последователни часа, без неопходност од дополнителни вонредни термини или специјални организациски интервенции. На тој начин се демонстрира дека батеријата е временски економична и усогласлива со реалната динамика на училиштето, при што се задржува висок степен на стандардизација.

Од аспект на професионалното искуство на наставниците, добиените податоци укажуваат на јасно разграничени области на рутина и области во кои е потребна дополнителна едукација. Претходно искуство во мерење на телесна тежина пријавиле 50% од наставниците (3 наставници), 66,7% (4 наставници) имале искуство во мерење на телесна висина, додека 83,3% (5 наставници) претходно редовно го применувале тестот скок во далечина од место. Притоа, 66,7% (4 наставници) пријавиле искуство со уште еден од стандардните фитнес-тестови вклучени во протоколот. Наспроти тоа, ниту еден наставник не пријавил дека некогаш го проценувал тестот за сила на стисок на дланката, дебелината на кожните набори или обемот на половината. Овие наоди укажуваат дека дел од предложените мерки и тестови се веќе интегрирани во традиционалната наставна практика, додека други претставуваат нови елементи кои бараат насочена обука, стандардизирање и практично унежбување.

Ваквата комбинација од висока општа применливост и селективен недостаток на претходно искуство во одредени мерни техники има двоен практичен заклучок. Од една страна, демонстрира дека МАКФИТ батеријата може да се спроведува без поголеми

организациски пречки и без прекумерно оптоварување на наставниците и учениците. Од друга страна, ги нагласува приоритетните области за професионален развој, односно потребата од систематска едукација за правилна примена на динамометријата на дланка, антропометриските мерења на обемот на половината и проценката на кожните набори. Со вклучување на овие содржини во програми за континуирана едукација на наставниците и со обезбедување јасни писмени и визуелни протоколи, применливоста на батеријата може дополнително да се подобри и да се намали ризикот од мерни грешки.

Во целина, резултатите на ова истражување упатуваат на тоа дека МАКФИТ батеријата претставува реално изводлива, временски рационална и организациски прифатлива алатка за систематско тестирање на физичкиот и моторичкиот развој кај адолесцентите во училишни услови. Оваа висока применливост е клучен предуслов за нејзина поширока имплементација на национално ниво и за создавање континуиран мониторинг систем кој ќе им овозможи на училиштата, стручните служби и креаторите на политики да располагаат со доверливи податоци за состојбата и трендовите на физичкиот фитнес кај младата популација.

## 7. ЗАКЛУЧОЦИ

Заклучоците од ова истражување претставуваат синтеза на најважните емпириски наоди и нивна интерпретација во контекст на поставените цели, формулираните хипотези и современите научни сознанија во областите на кинезиологијата, педијатриската физиологија и јавното здравје. Истражувањето беше насочено кон систематска процена на тест–ретест релијабилноста на антропометриските мерки и моторичките (фитнес) тестови вклучени во батеријата МАКФИТ, како и кон разгледување на половите разлики во физичкиот и моторичкиот профил кај адолесцентите во доцна средношколска возраст. Со тоа се создаде основа не само за проверка на методолошката издржаност на применетите инструменти, туку и за нивно позиционирање како алатка за рутински мониторинг на физичката подготвеност во образовниот систем.

Во рамките на докторандската студија беше применет тест–ретест дизајн со интервал од една седмица, што овозможи да се разграничат трајните биолошки и функционални карактеристики од краткорочните варијации поврзани со мерната грешка, мотивацијата или ефектот на запознавање со тестот. Притоа беа користени комплементарни индикатори на релативна и апсолутна релијабилност, како и анализа на согласност на индивидуално ниво, што овозможи заклучоците да се темелат на повеќеслојна статистичка аргументација. Ваквиот пристап ја зајакнува валидноста на добиените резултати и обезбедува цврста основа за нивна примена во практиката.

Во продолжение, заклучоците се формулирани така што на концизен и кохерентен начин ги обединуваат резултатите од тест–ретест анализата, оценката на мерната грешка и минимално детектибилната промена, проценката на практичната заменливост на мерењата и утврдените полови разлики во антропометрискиот и моторичкиот профил. Притоа, акцентот е ставен на тоа колку и на кој начин добиените наоди можат да послужат како научно заснована подлога за дијагностика, следење и унапредување на физичкиот фитнес кај младите, како и за понатамошни истражувачки и применети интервенции во школски и спортски контекст.

### **Заклучок за X1**

Добиените резултати јасно покажуваат дека антропометриските мерки и поголемиот дел од моторичките (фитнес) тестови во батеријата МАКФИТ поседуваат добра до одлична тест–ретест релијабилност и кај момчињата и кај девојчињата. Интракласните корелациски коефициенти ICC(2,1) кај сите антропометриски показатели и кај повеќето моторички тестови се во зоната на висока релијабилност, со многу тесни интервали на доверба и минимални отстапувања меѓу T1 и T2. Ова упатува на стабилно рангирање на испитаниците во сите основни домени на физичкиот фитнес. Оттука, хипотезата X1 во целост се потврдува.

### **Заклучок за X2**

Индикаторите на апсолутна релијабилност (SEM, SEM%, MDC и CV%) упатуваат дека кај антропометриските мерки и тестовите за сила и агилност мерната грешка е релативно мала, а минимално детектибилната промена (MDC) е доволно ниска за да овозможи детекција на реални индивидуални промени во разумен временски интервал. Кај тестовите со изразен ретест-ефект (претклон во сед, скок од место, подигнување на труп за 30 s) SEM% и CV% се нешто повисоки, но и понатаму во прифатливи граници, што овозможува практична употреба доколку промените се интерпретираат во однос на дефинираните прагови за MDC. Генерално, апсолутната релијабилност е задоволителна и овозможува смислено следење на поединечни ученици, па хипотезата X2 е во најголем дел потврдена, со напомена за внимателна интерпретација кај тестовите со поголем ефект на учење.

### **Заклучок за X3**

Анализата на разликите меѓу T1 и T2 покажува дека кај основните антропометриски показатели не постојат статистички значајни систематски разлики, што е во согласност со очекувањето за стабилност на овие мерки во интервал од една седмица. Наспроти тоа, кај дел од моторичките тестови (флексибилност, експлозивна сила на долните екстремитети, силова издржливост на трупот) се регистрирани статистички значајни подобрувања при ретестот, со мала до умерена/голема големина на ефект, што јасно упатува на ретест-ефект (ефект на запознавање со тестот). Оттука, хипотезата X3 е делумно потврдена: се потврдува отсуство на систематски разлики кај

антропофијата, но се потврдува и постоењето на значаен ретест-ефект кај повеќе моторички тестови.

#### **Заклучок за X4**

Bland–Altman анализата покажа дека кај најголем дел од антропометриските и моторичките показатели просечниот bias меѓу T1 и T2 е мал и статистички незначаен, а индивидуалните разлики претежно се наоѓаат во тесни и клинички прифатливи 95% граници на согласност, без изразена хетероскедастичност. Ова значи дека, за повеќето тестови, резултатот од ретестот може практично да го замени резултатот од првото мерење без да се промени интерпретацијата за реалниот статус на поединецот, под услов при интерпретацијата да се имаат предвид праговите на MDC и познатиот ретест-ефект кај неколку тестови. Врз основа на тоа, хипотезата X4 суштински се потврдува.

#### **Заклучок за X5**

Мултиваријантната анализа потврди статистички значаен ефект на полот врз збирната варијација на антропометриските карактеристики и моторичките способности во првото (и ретест) мерење, при што кај повеќе поединечни варијабли се евидентираат значајни полови разлики со умерена до многу голема големина на ефект на влијание. Момчињата имаат значително повисоки вредности на телесна висина и тежина и постигнуваат јасно супериорни резултати во динамометријата, скокот од место и тестот на агилност, додека кај ВМІ и флексибилноста половите разлики не се статистички значајни или се минимални. Овој образец укажува дека полот е силен детерминант на силовите и брзинските компоненти на фитнесот во доцна адолесценција. Следствено, хипотезата X5 е во целост потврдена, со очекувани варијации по поединечни тестови.

Општите наоди од истражувањето упатуваат на тоа дека применетата батерија од антропометриски мерки и моторички (фитнес) тестови во рамки на МАКФИТ-протоколот обезбедува високо стабилни и репродуктивни резултати кај адолесцентите. Во интервал од една седмица, основните телесни размери, како и тестовите за мускулна сила, експлозивност и агилност, покажаа конзистентни средни вредности и високи интракласни корелациски коефициенти, што јасно укажува на добра до одлична тест–ретест релијабилност. Овој образец потврдува дека, во услови на стандарден протокол и контролирана процедура, мерењата можат да се повторуваат без значајно статистичко

„шумеење“ и дека евентуалните промени во подолг временски период ќе одразуваат реални биолошки и функционални трансформации, а не артефакти на мерниот процес.

Попрецизната анализа на апсолутната релијабилност покажа дека стандардната грешка на мерењето и минимално детектибилната промена се во прифатливи граници за повеќето тестови, што овозможува сигурно следење на индивидуалните промени кај учениците. При тоа, кај тестовите за телесна висина, телесна тежина и обем на половина, праговите за реална промена се релативно ниски, што дава можност за чувствителна детекција на надолжни трендови. Кај дел од моторичките тестови, особено претклонот во сед, скокот од место и подигнувањето на трупот за 30 секунди, се евидентира нешто поизразен ретест-ефект, со статистички значајно подобрување на просечните вредности при второто мерење. Тоа укажува дека овие тестови, покрај реалната физичка способност, ја рефлектираат и запознаеноста со задачата и усовршувањето на техниката, поради што при кратки интервали помеѓу мерењата е неопходно резултатите да се толкуваат во контекст на дефинираните прагови за минимално детектибилна промена.

Анализата на согласноста на индивидуално ниво дополнително ја потврди употребливоста на батеријата. Bland–Altman дијаграмите покажаа дека просечниот bias меѓу првото и второто мерење е занемарлив кај најголемиот број варијабли и дека доминантен дел од индивидуалните разлики се наоѓаат во тесни, клинички и педагошки прифатливи граници на согласност. Отсуството на изразена хетероскедастичност кај повеќето тестови значи дека точноста на мерењето не зависи значајно од нивото на постигнувањето, односно дека протоколот е подеднакво стабилен и кај послабите и кај подобрите ученици. Во практична смисла, ова овозможува резултатите од ретестот да се користат како валидна замена за првичното мерење при следење на истото дете, без ризик од погрешна интерпретација на неговиот вистински статус, доколку притоа се почитуваат познатите прагови на мерната грешка и ретест-ефектот кај чувствителните тестови.

Половите разлики добиени во анализата се конзистентни со очекуваните биолошки и функционални закономерности во доцна адолесценција. Момчињата се значајно повисоки и потешки, со поголем обем на половината и супериорни резултати во тестовите за мускулна сила, експлозивна моќ и брзинско-агилни способности, додека индексот на телесна маса е сличен кај двата пола, а флексибилноста не покажува јасна

полова предност. Овој образец укажува дека полот е силен детерминант на силовите и брзинските компоненти на физичкиот фитнес, додека одредени аспекти, како флексибилноста и глобалниот индекс на телесна маса, се помалку зависни од половиот статус во разгледуваната старосна група. Ваквите наоди не само што го потврдуваат присуството на очекувани полови диморфизми, туку обезбедуваат и конкретни, локално релевантни податоци за профилот на физичката подготвеност кај адолесцентите.

Сумирано, резултатите од истражувањето јасно демонстрираат дека батеријата МАКФИТ претставува методолошки валиден и релијабилен инструмент за проценка и следење на антропометрискиот и моторичкиот развој кај учениците од средношколска возраст. Високата тест–ретест стабилност, прифатливата апсолутна релијабилност и задоволителната согласност меѓу мерењата на индивидуално ниво создаваат цврста основа за употреба на овие тестови во образовен, спортски и здравствен контекст. Истовремено, јасно изразените полови разлики во одредени компоненти на физичкиот фитнес го нагласуваат значењето на индивидуализиран и полого-сензитивен пристап при планирањето на наставните содржини и тренажните програми. На тој начин, добиените емпириски сознанија не остануваат само во рамките на теоретската анализа, туку директно се преведуваат во практични насоки за систематско мониторирање, рана детекција на отстапувања и целно унапредување на физичката подготвеност и здравјето на младата популација.

## 8. ТЕОРЕТСКО И ПРАКТИЧНО ЗНАЧЕЊЕ НА ИСТРАЖУВАЊЕТО

Теоретското значење на ова истражување произлегува од фактот дека физичкиот и моторичкиот развој на децата и адолесцентите претставува сложен, мултидимензионален конструкт кој може валидно да се опфати само доколку се користат добро концептуализирани и психометриски проверени мерни инструменти. Во рамките на кинезиологијата, перформансите на кардиореспираторниот фитнес, силата, издржливоста, агилноста, координацијата, рамнотежата, флексибилноста и телесниот состав се гледаат како меѓусебно поврзани компоненти на здравјето и функционалниот развој кај детската популација. Сепак, нивната интерпретација зависи од тоа дали тестовите што ги користиме се доволно релијабилни, објективни, применливи и безбедни. Без стабилни и повторливи мерења, тешко може да се утврди дали промените во резултатите се последица на вистински развоен напредок, ефект од тренинг или образовна интервенција, или пак се должат на случајна мерна грешка. Во таа смисла, ова истражување ја продлабочува теоретската рамка на кинезиолошката метриска практика, бидејќи во национален контекст го проблематизира и систематски го оценува квалитетот на мерните инструменти што се употребуваат за процена на физичкиот и моторичкиот развој кај младите.

Особен теоретски придонес на истражувањето е тоа што релијабилноста, објективноста, безбедноста и применливоста на антропометриските мерки, мерките на телесниот состав и моторичките тестови од батеријата МАКФИТ се разгледуваат не само описно, туку и преку современи психометриски показатели. Воведувањето и систематската употреба на интра класа корелации, стандардна грешка на мерење, најмала забележлива разлика и поврзани индикатори овозможува понијансиран увид во тоа до кој степен добиените резултати се стабилни, односно до кој степен забележаните варијации се статистички и практично значајни. На тој начин, МАКФИТ студијата не само што обезбедува податоци за нивото на физичкиот фитнес и моторичкиот развој, туку истовремено го проширува знаењето за мерната прецизност на тестовите во реални училишни услови. Ова е теоретски важно бидејќи во меѓународната литература постои јасен тренд за премин од едноставни корелациски оценки на релијабилност кон комбинирани модели кои ги вклучуваат и релативните и апсолутните показатели на мерната стабилност, а национални податоци за ваков пристап досега изостануваа.

Истражувањето има теоретска вредност и во доменот на развојната кинезиологија и педагогијата, бидејќи создава основа за понатамошни лонгитудинални и компаративни студии за физичкиот и моторичкиот развој кај македонските ученици. Со утврдување на релијабилни и објективни мерки се овозможува со поголема сигурност да се креираат возрасни и полого-специфични референтни вредности и перцентилни криви, кои во иднина може да се користат за следење на трендовите во физичката кондиција на генерациско и популациско ниво. Истовремено, потврдата дека мерните инструменти се применливи и безбедни во училишни услови го олеснува нивното интегрирање во редовната наставна практика, што пак, ја зајакнува врската помеѓу теоријата и практиката. Со други зборови, ова истражување поставува концептуална и методолошка рамка за создавање национален систем за мониторинг на физичкиот фитнес и моторичкиот развој, кој може да се надградува и проширува во иднина и да се споредува со меѓународни бази на податоци.

Практичното значење на истражувањето е директно поврзано со неговиот фокус на училишното опкружување и на реалните услови во кои наставниците по физичко и здравствено образование ги спроведуваат тестирањата. Применливоста на батеријата МАКФИТ во стандардни училишни услови значи дека наставниците можат, без неоправдано дополнително материјално и кадровско оптоварување, да ги мерат антропометриските карактеристики, телесниот состав и моторичките способности кај учениците и да ги добијат потребните информации за нивниот актуелен статус. Од посебно значење е фактот што во рамките на истражувањето се проценува и временската оптовареност, организациските барања, разбирливоста на инструкциите за учениците, како и субјективната перцепција на наставниците за леснотијата на примена на тестовите. Овие информации се клучни за да се утврди дали предложената батерија е навистина практична и изводлива во секојдневната наставна практика, а не само во контролирани истражувачки услови.

Резултатите од истражувањето имаат значајни импликации за наставната и вон наставната работа со децата и адолесцентите. Годишната проценка на физичкиот и моторичкиот развој преку стандардизирани МАКФИТ тестови им овозможува на учениците да го следат сопствениот напредок, да развиваат свест за сопственото тело и да ги разберат врските помеѓу редовната физичка активност, квалитетот на живот и

здравствените исходи. Наставниците по физичко образование, со помош на објективни и релијабилни податоци, можат да ја индивидуализираат наставата, да креираат диференцирани задачи и вежби во зависност од почетното ниво на фитнес и да препознаат ученици со потенцијално зголемен ризик за неповолни здравствени исходи. Притоа, тестовите што се лесни за администрирање, безбедни и добро прифатени од учениците, придонесуваат за позитивен однос кон проценувањето и го намалуваат стресот и отпорот што често го предизвикуваат предизвикувачките тестирања.

Практичните импликации се значајни и за родителите, здравствените работници и спортските организации. Родителите, преку јасни и споредливи показатели, може да го следат физичкиот и моторичкиот развој на своите деца и да го споредуваат со врсниците на национално или меѓународно ниво. Ова создава реална основа за информирани одлуки во врска со вклучување во спортски активности, промена на дневните навики, намалување на седентарното однесување и подобрување на животниот стил воопшто. Здравствените организации, пак, може да ги користат податоците од МАКФИТ за рано препознавање на ризични состојби поврзани со дебелина, недоволен фитнес или нарушен моторички развој и да планираат превентивни или корективни програми ориентирани кон детската популација. Спортските клубови и тренерите можат да ги искористат резултатите од тестирањето за пообјективна селекција на деца со специфичен потенцијал, но и за следење на ефектите од тренинг програмите и навремена корекција на оптоварувањето.

Од перспектива на образовните и здравствените политики, ова истражување има потенцијал да послужи како основа за дизајнирање и имплементирање национален систем за мониторинг на физичкиот фитнес и моторичкиот развој. Кога тестовите се потврдени како релијабилни, безбедни и применливи, резултатите можат да се соберат, обработат и групираат на ниво на училиште, општина, регион и држава. На тој начин се создава база на податоци преку која може да се следат долгорочни трендови, да се идентификуваат разлики помеѓу различни средини и да се процени влијанието на одредени образовни или јавно здравствени интервенции. Креаторите на политики добиваат можност објективно да ги оценат ефектите од новите наставни програми, проектите за промоција на физичка активност или инфраструктурните вложувања во спортски објекти. Воедно, податоците добиени преку МАКФИТ тестирањето можат да

бидат интегрирани во меѓународни иницијативи и компаративни истражувања, со што Северна Македонија ќе се позиционира рамноправно во европскиот и светскиот простор на следење на фитнесот кај младите.

Крајната практична поента на ова истражување е во тоа што создава јасна, оперативна патека за користење на добиените податоци во секојдневната практика. На индивидуално ниво, наставникот и ученикот добиваат алатка за дијагностика, поставување реални цели и следење на напредокот во физичкото образование и вон наставните активности. На ниво на одделение и училиште, податоците овозможуваат планирање и адаптација на наставните програми, организирање таргетирани програми за ученици со пониско ниво на фитнес или со одредени здравствени ризици и објективна евалуација на реализираните активности. На популациско и национално ниво, резултатите од истражувањето служат како основа за креирање, спроведување и ревидирање на политики во доменот на физичкото образование, спортот и јавното здравје, со цел да се поддржи здравиот раст и развој на генерациите деца и адолесценти во Република Северна Македонија. Со тоа, теоретските сознанија и емпириските податоци што ги обезбедува оваа студија се трансформираат во конкретни, применливи алатки за унапредување на наставната и здравствената практика

## ЛИТЕРАТУРА

- Alricsson, M., Harms-Ringdahl, K., & Werner, S. (2001). Reliability of sports related functional tests with emphasis on speed and agility in young athletes. *Scandinavian journal of medicine & science in sports*, 11(4), 229-232.
- Artero, E. G., España-Romero, V., Castro-Piñero, J., Ortega, F. B., Suni, J., Castillo-Garzon, M. J., & Ruiz, J. R. (2011). Reliability of field-based fitness tests in youth: A systematic review. *International Journal of Sports Medicine*, 32(3), 159–169. <https://doi.org/10.1055/s-0030-1268488>.
- Australian Council for Health Physical Education and Recreation (ACHPER)* (1996). Handbook for the Australian Fitness Education Award manual (1996) South Australia: ACHPER Publications.
- Baumgartner, T.A. Norm-referenced measurement: Reliability (1989). In *Measurement Concepts in Physical Education and Exercise Science*; Safrit, M.J., Woods, T.M., Eds.; Human Kinetics: Champaign, IL, USA; pp. 45–72.
- Beets, M. W., & Pitetti, K. H. (2006). Criterion-referenced reliability and equivalency between the PACER and 1-mile run/walk for high school students. *Journal of Physical Activity and Health*, 3(s2), S21-S33.
- Bianco A, Jemni M, Thomas E, Patti A, Paoli A, Ramos Roque J, et al (2015) A systematic review to determine reliability and usefulness of the field-based test batteries for the assessment of physical fitness in adolescents—The ASSO Project. *Int J Occup Med Environ Health* 28(3), 445–478.
- Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476), 307-310.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical methods in medical research*, 8(2), 135-160.
- Buchan, D. S., Ollis, S., Young, J. D., Cooper, S. M., Shield, J. P., & Baker, J. S. (2013). High intensity interval running enhances measures of physical fitness but not metabolic measures of cardiovascular disease risk in healthy adolescents. *BMC public health*, 13, 1-12.

- Canadian Society for Exercise Physiology (CSEP), (2003). The Canadian Physical Activity, Fitness & Lifestyle Approach (CPAFLA): CSEPH Health & Fitness Program's Health-Related Appraisal and Counselling Strategy*Ottawa: CSEP.
- Caprio, S., Daniels, S. R., Drewnowski, A., Kaufman, F. R., Palinkas, L. A., Rosenbloom, A. L., & Schwimmer, J. B. (2008). Influence of race, ethnicity, and culture on childhood obesity: implications for prevention and treatment: a consensus statement of Shaping America's Health and the Obesity Society. *Diabetes care*, *31*(11), 2211.
- Castro-Piñero, J., Ortega, F. B., Artero, E. G., Girela-Rejón, M. J., Mora, J., Sjöström, M., & Ruiz, J. R. (2010). Assessing muscular strength in youth: usefulness of standing long jump as a general index of muscular fitness. *The Journal of Strength & Conditioning Research*, *24*(7), 1810-1817.
- Castro-Piñero, J., Ortega, F. B., Mora, J., Sjöström, M., & Ruiz, J. R. (2009). Criterion related validity of 1/2 mile run-walk test for estimating VO<sub>2</sub>peak in children aged 6–17 years. *International journal of sports medicine*, 366-371.
- Cole, T. J., Flegal, K. M., Nicholls, D., & Jackson, A. A. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. *Bmj*, *335*(7612), 194.
- Cooper Institute for Aerobics Research. *The Prudential Fitnessgram: Test Administration Manual* (2004). Champaign, IL: Human Kinetics.
- Council of Europe Committee for the Development of Sport. (1993) EUROFIT: Handbook for the EUROFIT Tests of Physical Fitness* Strasbourg: Council of Europe.
- De Miguel-Etayo, P., Mesana, M. I., Cardon, G., De Bourdeaudhuij, I., Gózdź, M., Socha, P., ... & ToyBox-study group. (2014). Reliability of anthropometric measurements in European preschool children: The ToyBox-study. *Obesity reviews*, *15*, 67-73.
- Departamento Administrativo Nacional de Estadística (DANE) (2007) Los grupos étnicos de Colombia en el censo de 2005. Análisis de Resultados*, Bogotá, Colombia.
- Dumith, S. C., Van Dusen, D., & Kohl, H. W. (2012). Physical fitness measures among children and adolescents: are they all necessary?. *The Journal of sports medicine and physical fitness*, *52*(2), 181-189.
- Eisenmann, J. C. (2007). Aerobic fitness, fatness and the metabolic syndrome in children and adolescents. *Acta Paediatrica*, *96*(12), 1723-1729.

- Ekelund, U., Anderssen, S. A., Froberg, K., Sardinha, L. B., Andersen, L. B., Brage, S., & European Youth Heart Study Group. (2007). Independent associations of physical activity and cardiorespiratory fitness with metabolic risk factors in children: the European youth heart study. *Diabetologia*, *50*, 1832-1840.
- España-Romero, V., Artero, E. G., Jiménez-Pavón, D., Cuenca-García, M., Ortega, F. B., Castro-Piñero, J., Sjöström, M., Castillo, M. J., & Ruiz, J. R. (2010). Assessing health-related fitness tests in the school setting: Reliability, feasibility and safety; the ALPHA study. *International Journal of Sports Medicine*, *31*(7), 490–497. <https://doi.org/10.1055/s-0030-1251990>.
- España-Romero, V., Artero, E. G., Santaliestra-Pasias, A. M., Gutierrez, A., Castillo, M. J., & Ruiz, J. R. (2008). Hand span influences optimal grip span in boys and girls aged 6 to 12 years. *The Journal of hand surgery*, *33*(3), 378-384.
- Garcia, S., Sarmiento, O. L., Forde, I., & Velasco, T. (2013). Socio-economic inequalities in malnutrition among children and adolescents in Colombia: the role of individual-, household-and community-level characteristics. *Public health nutrition*, *16*(9), 1703-1718.
- Gómez-Cabello, A., Vicente-Rodríguez, G., Albers, U., Mata, E., Rodríguez-Marroyo, J. A., Olivares, P. R., ... & Ara, I. (2012). Harmonization process and reliability assessment of anthropometric measurements in the elderly EXERNET multi-centre study.
- González, S. A., Sarmiento, O. L., Cohen, D. D., Camargo, D. M., Correa, J. E., Páez, D. C., & Ramírez-Vélez, R. (2014). Results from Colombia's 2014 report card on physical activity for children and youth. *Journal of Physical Activity and Health*, *11*(s1), S33-S44.
- Grøntved, A., Ried-Larsen, M., Møller, N. C., Kristensen, P. L., Froberg, K., Brage, S., & Andersen, L. B. (2015). Muscle strength in youth and cardiovascular risk in young adulthood (the European Youth Heart Study). *British journal of sports medicine*, *49*(2), 90-94.
- Gutin, B., Litaker, M., Islam, S., Manos, T., Smith, C., & Treiber, F. (1996). Body-composition measurement in 9–11-y-old children by dual-energy X-ray absorptiometry, skinfold-thickness measurements, and bioimpedance analysis. *The American journal of clinical nutrition*, *63*(3), 287-292.

- Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., & Ekelund, U. (2012). Global physical activity levels: surveillance progress, pitfalls, and prospects. *The lancet*, 380(9838), 247-257.
- Hannibal, N. S., Plowman, S. A., Looney, M. A., & Brandenburg, J. (2006). Reliability and validity of low back strength/muscular endurance field tests in adolescents. *Journal of Physical Activity and Health*, 3(s2), S78-S89.
- Hartman, J. G., & Looney, M. (2003). Norm-referenced and criterion-referenced reliability and validity of the back-saver sit-and-reach. *Measurement in Physical Education and Exercise Science*, 7(2), 71-87.
- Hurtig-Wennlöf, A., Ruiz, J. R., Harro, M., & Sjöström, M. (2007). Cardiorespiratory fitness relates more strongly than physical activity to cardiovascular disease risk factors in healthy children and adolescents: the European Youth Heart Study. *European Journal of Cardiovascular Prevention & Rehabilitation*, 14(4), 575-581.
- Joyner, M. J., Hunter, S. K., & Senefeld, J. W. (2025). Evidence on sex differences in sports performance. *Journal of Applied Physiology*, 138(1), 274-281.
- Kaminsky, L. A., Arena, R., Beckie, T. M., Brubaker, P. H., Church, T. S., Forman, D. E., ... & Williams, M. A. (2013). The importance of cardiorespiratory fitness in the United States: the need for a national registry: a policy statement from the American Heart Association. *Circulation*, 127(5), 652-662.
- Kelishadi, R., Gheiratmand, R., Ardalan, G., Adeli, K., Gouya, M. M., Razaghi, E. M., ... & CASPIAN Study Group. (2007). Association of anthropometric indices with cardiovascular disease risk factors among children and adolescents: CASPIAN Study. *International journal of cardiology*, 117(3), 340-348.
- Kostovski, Z. i Georgiev, G. (2009). Relijabilnost primenjenih testova za procenu specifične karate koordinacije kod vrhunskih karatista sa različite hronološke uzrasti. Bo: Zbornik naučnih i stručnih radova “Sport i zdravlje”, sa II Medjunarodni simpozijum sport i zdravlje , 127-130. Tuzla: Fakultet za tjelesni odgoj i sport.
- Kostovski, Z. i Georgiev, G. (2009). Validnost testova za procenu specifične karate koordinacije kod 12 i 14 godišnjih karatista. Bo: Zbornik naučnih i stručnih radova “Sport i zdravlje”, sa II Medjunarodni simpozijum sport i zdravlje ”, 120-126. Tuzla: Fakultet za tjelesni odgoj i sport.

- Kostovski, Z., & Georgiev, G. (2009). Measure Characteristics of Motor Tests for Assessing Rhythmic Structure and Explosive Strength with Karate Athletes and Non-Athletes at the Age of 12. [Мерни карактеристики на моторни тестови за проценување на ритмичка структура и експлозивна сила кај карате спортисти и неспортисти на возраст од 12 години]. *Sport scientific and practical aspects* , VI (2), 35-40.
- Kostovski, Z., Determination of the measurement characteristics of some standard and specific motor tests applied to karate athletes of different chronological ages. Doctoral dissertation. FFK Skopje, 2004.
- Kostovski, Z., Masić, Z., & Đukanović, N., (2011). Measuring performance of the strokes with leg (mavashi geri and ushiro geri) in karate sport [Merne karakteristike udarca nogom (mavashi geri i ushiro geri) kod karate sporsrtista]. IV Medjunarodni simpozijum Sport i zdravlje , 245-248. Tuzla: Fakultet za tjelesni odgoj i sport.
- Kostovski, Z., Ћeljković, M., Ibri, L., Soklevska, E., & Zaborski, B. (2011). Validity, reliability and sensitivity of the test stroke with leg mae geri. *Megjunarodni naucni kongres antropoloshki aspekti sporta, fizickog vaspitanja i rekreacija* , 37-38.
- Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., & Katzmarzyk, P. T. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet*, 380(9838), 219-229.
- Leger LA, Mercier D, Gadoury C, Lambert J (1988) The multistage 20 metre shuttle run test for aerobic fitness. *J Sports Sci* 6: 93–101. PMID: 3184250.
- Liu, N. Y. S., Plowman, S. A., & Looney, M. A. (1992). The reliability and validity of the 20-meter shuttle test in American students 12 to 15 years old. *Research quarterly for exercise and sport*, 63(4), 360-365.
- Loomba-Albrecht, L. A., & Styne, D. M. (2009). Effect of puberty on body composition. *Current Opinion in Endocrinology, Diabetes & Obesity*, 16(1), 10–15. <https://doi.org/10.1097/MED.0b013e328320d54c>.
- Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ editors (2006) Global burden of disease and risk factors. Washington, DC: World Bank.
- Lubans, D. R., Morgan, P., Callister, R., Plotnikoff, R. C., Eather, N., Riley, N., & Smith, C. J. (2011). Test–retest reliability of a battery of field-based health-related fitness measures for adolescents. *Journal of sports sciences*, 29(7), 685-693.

- Marques, A., Henriques-Neto, D., Peralta, M., Martins, J., Gomes, F., Popovic, S., Masanovic, B., Demetriou, Y., Schlund, A., & Ihle, A. (2021). Field-based health-related physical fitness tests in children and adolescents: A systematic review. *Frontiers in Pediatrics*, 9, 640028. <https://doi.org/10.3389/fped.2021.640028>.
- Mayorga-Vega, D., Bocanegra-Parrilla, R., Ornelas, M., & Viciano, J. (2016). *Criterion-Related Validity of the Distance- and Time-Based Walk/Run Field Tests for Estimating Cardiorespiratory Fitness: A Systematic Review and Meta-Analysis*. PLOS ONE, 11(3), e0151671. <https://doi.org/10.1371/journal.pone.0151671> PLOS.
- Moir, G., Shastri, P., & Connaboy, C. (2008). Intersession reliability of vertical jump height in women and men. *The Journal of Strength & Conditioning Research*, 22(6), 1779-1784.
- Molenaar, H. M., Zuidam, J. M., Selles, R. W., Stam, H. J., & Hovius, S. E. (2008). Age-specific reliability of two grip-strength dynamometers when used by children. *The Journal of bone and joint surgery. American volume*, 90(5), 1053–1059.
- Moreno, L. A., Joyanes, M., Mesana, M. I., González-Gross, M., Gil, C. M., Sarría, A., ... & AVENA Study Group. (2003). Harmonization of anthropometric measurements for a multicenter nutrition survey in Spanish adolescents. *Nutrition*, 19(6), 481-486.
- Nagy, E., Vicente-Rodriguez, G., Manios, Y., Béghin, L., Iliescu, C., Censi, L., Dietrich, S., Ortega, F. B., De Vriendt, T., Plada, M., Moreno, L. A., Molnar, D., & HELENA Study Group (2008). Harmonization process and reliability assessment of anthropometric measurements in a multicenter study in adolescents. *International journal of obesity (2005)*, 32 Suppl 5, S58–S65.
- Norman, K., Stobäus, N., Gonzalez, M. C., Schulzke, J. D., & Pirlich, M. (2011). Hand grip strength: outcome predictor and marker of nutritional status. *Clinical nutrition*, 30(2), 135-142.
- Nuzzo, J. L., & Pinto, M. D. (2025). Sex Differences in Upper-and Lower-Limb Muscle Strength in Children and Adolescents: A Meta-Analysis. *European Journal of Sport Science*, 25(5), e12282.
- Nuzzo, J. L., Anning, J. H., & Scharfenberg, J. M. (2011). The reliability of three devices used for measuring vertical jump height. *Journal of strength and conditioning research*, 25(9), 2580–2590.

- Ortega, F. B., Artero, E. G., Ruiz, J. R., Vicente-Rodriguez, G., Bergman, P., Hagströmer, M., ... & Castillo, M. J. (2008). Reliability of health-related physical fitness tests in European adolescents. The HELENA Study. *International journal of obesity*, 32(5), S49-S57.
- Ortega, F. B., Labayen, I., Ruiz, J. R., Kurvinen, E., Loit, H. M., Harro, J., Veidebaum, T., & Sjöström, M. (2011). Improvements in fitness reduce the risk of becoming overweight across puberty. *Medicine and science in sports and exercise*, 43(10), 1891–1897.
- Ortega, F. B., Ruiz, J. R., Castillo, M. J., & Sjöström, M. (2008). Physical fitness in childhood and adolescence: A powerful marker of health. *International Journal of Obesity*, 32(1), 1–11. <https://doi.org/10.1038/sj.ijo.0803774>.
- Parra, D. C., Iannotti, L., Gomez, L. F., Pachón, H., Haire-Joshu, D., Sarmiento, O. L., Kuhlmann, A. S., & Brownson, R. C. (2015). The nutrition transition in Colombia over a decade: a novel household classification system of anthropometric measures. *Archives of public health = Archives belges de sante publique*, 73(1), 12.
- Patterson, P., Bennington, J., & La-Rosa, T. D. (2001). Psychometric properties of child-and teacher-reported curl-up scores in children ages 10–12 years. *Research Quarterly for Exercise and Sport*, 72(2), 117-124.
- Pitetti, K. H., Fernhall, B., & Figoni, S. (2002). Comparing two regression formulas that predict VO<sub>2</sub>peak using the 20-m shuttle run for children and adolescents. *Pediatric Exercise Science*, 14(2), 125-134.
- Pratt, M., Norris, J., Lobelo, F., Roux, L., & Wang, G. (2014). The cost of physical inactivity: moving into the 21st century. *British journal of sports medicine*, 48(3), 171-173.
- Pritchard, T., O'Bryant, H., & Everhart, B. (2001). An alternative to the full sit-up testing for middle school students. *The Physical Educator*, 58(1), 42-42.
- Ramírez-Vélez, R., Rodrigues-Bezerra, D., Correa-Bautista, J. E., Izquierdo, M., & Lobelo, F. (2015). Reliability of health-related physical fitness tests among Colombian children and adolescents: the FUPRECOL study. *PloS one*, 10(10), e0140875.
- Rikli, R. E., Petray, C., & Baumgartner, T. A. (1992). The reliability of distance run tests for children in grades K—4. *Research quarterly for exercise and sport*, 63(3), 270-276.
- Robertson, D. G., & Fleming, D. (1987). Kinetics of standing broad and vertical jumping. *Can J Sport Sci*, 12(1), 19-23.

- Ruiz, J. R., Castro-Piñero, J., Artero, E. G., Ortega, F. B., Sjöström, M., Suni, J., & Castillo, M. J. (2009). Predictive validity of health-related fitness in youth: a systematic review. *British journal of sports medicine*, 43(12), 909-923.
- Ruiz, J. R., Castro-Piñero, J., España-Romero, V., Artero, E. G., Ortega, F. B., Cuenca, M. M., Jimenez-Pavón, D., Chillón, P., Girela-Rejón, M. J., Mora, J., Gutiérrez, A., Suni, J., Sjöström, M., & Castillo, M. J. (2011). Field-based fitness assessment in young people: the ALPHA health-related fitness test battery for children and adolescents. *British journal of sports medicine*, 45(6), 518–524.
- Saint Romain, B., & Mahar, M. T. (2001). Norm-referenced and criterion-referenced reliability of the push-up and modified pull-up. *Measurement in physical education and exercise science*, 5(2), 67-80.
- Santos, E. J., & Janeira, M. A. (2012). The effects of resistance training on explosive strength indicators in adolescent basketball players. *The journal of strength & conditioning research*, 26(10), 2641-2647.
- Schoonjans, F. R. A. N. K., Zalata, A., Depuydt, C. E., & Comhaire, F. H. (1995). MedCalc: a new computer program for medical statistics. *Computer methods and programs in biomedicine*, 48(3), 257-262.
- Schreuders, T. A., Roebroek, M. E., Goumans, J., van Nieuwenhuijzen, J. F., Stijnen, T. H., & Stam, H. J. (2003). Measurement error in grip and pinch force measurements in patients with hand injuries. *Physical therapy*, 83(9), 806-815.
- Steene-Johannessen, J., Anderssen, S. A., Kolle, E., & Andersen, L. B. (2009). Low muscle fitness is associated with metabolic risk in youth. *Medicine & Science in Sports & Exercise*, 41(7), 1361-1367.
- Stomfai, S., Ahrens, W., Bammann, K., Kovacs, E., Mårild, S., Michels, N., ... & Molnár, D. (2011). Intra-and inter-observer reliability in anthropometric measurements in children. *International journal of obesity*, 35(1), S45-S51.
- The President's Council on Physical Fitness and Sports. (2012). The President's Challenge: The Health Fit-ness Test. Available at: Accessed October 4.

- Tong, T. K., Fu, F. H., & Chow, B. C. (2001). Reliability of a 5-min running field test and its accuracy in VO<sub>2</sub> (max) evaluation. *Journal of sports medicine and physical fitness*, 41(3), 318.
- Ulijaszek, S. J., & Kerr, D. A. (1999). Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition*, 82(3), 165-177.
- Vanhelst, J., Béghin, L., Fardy, P. S., Ulmer, Z., & Czaplicki, G. (2016). Reliability of health-related physical fitness tests in adolescents: the MOVE P rogram. *Clinical Physiology and Functional Imaging*, 36(2), 106-111.
- Waninge, A., Van Der Weide, W., Evenhuis, I. J., Van Wijck, R., & Van Der Schans, C. P. (2009). Feasibility and reliability of body composition measurements in adults with severe intellectual and sensory disabilities. *Journal of Intellectual Disability Research*, 53(4), 377-388.
- Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research*, 19(1), 231-240.

# ПРИЛОГ

## ЗАКЛУЧНА СПОРЕДБА И ПРЕПОРАКИ

Табеларна спордба на релијабилноста: Студија vs. Литература

Тест/Варијабла	ИСС (момчиња)	ИСС (девојчиња)	ИСС (литература)	Забелешки (литература)
Телесна висина (cm)	0.999	0.998	>0.99	$\Delta_{12} \sim 0$ cm; R% $\approx 99.7\%$
Телесна тежина (kg)	0.996	0.991	0.95–0.99	$\Delta_{12} \sim 0$ kg; R% $\approx 96\%$
ВМИ (kg/m <sup>2</sup> )	0.992	0.988	0.95–0.99	Повисока рел. од кожни набори
Обем на половина (cm)	0.948	0.918	$\sim 0.95$ –0.98	$\Delta_{12}$ незначајна, ТЕМ $\sim 0.86$ cm
Претклон во сед (cm)	0.894	0.886	0.85–0.96	Хетеросцедастичност кај помалку флексибилни; ИСС пресколци $\sim 0.63$ –0.77
Скок од место (cm)	0.913	0.884	0.90–0.97	Без знач. $\Delta_{12}$ во многу студии; подобрување до 12 cm можно
Рачен стисок (просек)	0.961	0.941	0.84–0.99	Без $\Delta_{12}$ , без учење; ИСС дом. рака пресколци 0.67–0.81
Стомачни преси 30 s	0.822	0.819	0.90–0.97	Чест ефект на учење: +1–2 повтор. на ретест
Чуност 4×10 m (s)	0.994	0.995	0.80–0.94	Без знач. $\Delta_{12}$ ; тип. LoA $\pm 1$ –2 s

Сумирајќи ја релијабилноста на применетите тестови во ова истражување и споредувајќи ја со податоците од релевантната меѓународна литература, се добива конзистентна слика која дополнително ја валидира методолошката цврстина на протоколот. Највисока релијабилност беше забележана кај основните антропометриски мерења – телесната висина и тежина – каде што и кај момчињата и кај девојчињата интеркластерските корелациски коефициенти (ИСС) беа во опсег од 0.991 до 0.999, што е во согласност со податоците од Artero et al. (2011) и King-Dowling et al. (2023), каде тие мерки редовно демонстрираат ИСС над 0.95 и многу низок (SEM). Слична стабилност се забележува и кај индексот на телесна маса (ВМИ), каде што вредностите во нашето истражување (0.988–0.992) се целосно компатибилни со оние во ALPHA студиите, каде ВМИ често се препорачува како попрактична и понепристрасна алтернатива на кожните набори (Ortega et al. 2008).

Обемот на половината, како индикатор на централизирана адипозност, исто така демонстрира висока релијабилност, со ИСС од 0.948 кај момчињата и 0.918 кај девојчињата, што е благо пониско од референтните вредности ( $\sim 0.95$ –0.98) пријавени од Ruiz et al. (2011), но сè уште во опсегот на „одлична релијабилност“. Кај тестовите на флексибилност (претклон во сед) беа забележани малку пониски коефициенти на релијабилност (0.886–0.894), што е очекувано со оглед на хетеросцедастичниот карактер

на овие тестови и документираната зависност од почетната флексибилност на испитаникот – феномен кој често доведува до варијабилност во ретест резултатите, особено кај пофлексибилни деца (Artero et al. 2011).

Меѓу моторичките тестови, скокот во далечина од место покажува висока стабилност (ICC 0.884–0.913), што е во согласност со претходни трудови кои укажуваат на тоа дека овој тест има ниска чувствителност на „ефекти на учење“ и добра конзистентност при повторени мерења (Ortega et al. 2008). Тестовите за максимален стисок на дланката, пресметан како просек од десна и лева рака, имаат извонредна релијабилност (ICC 0.941–0.961), што исто така е компарабилно со опсегот 0.84–0.99 пријавен од King-Dowling et al. (2023), со напомена дека кај помлади популации (предшколци) се пријавуваат нешто пониски вредности, особено кај недоминантната рака.

Кај тестот за подигнување на трупот од легната положба во 30 секунди, ICC вредностите беа околу 0.82, што е класифицирано како „добра релијабилност“, но нешто пониско од литературата (0.90–0.97), каде се забележува и редовен ефект на учење од 1 до 2 повторувања на ретест (Artero et al., 2011). Конечно, тестот за агилност 4×10 m, кој во повеќе студии се смета за тест со највисока релативна релијабилност, кај нашата популација прикажа извонредни вредности од 0.994–0.995, со што дури ги надминува просечните вредности од 0.80–0.94 пријавени во систематски прегледи (Ruiz et al., 2011).

Овие споредби потврдуваат дека добиените резултати од нашето истражување не само што се методолошки конзистентни и стабилни, туку и се целосно усогласени со меѓународните стандарди за релијабилност кај полеви тестови за адолесценти. Ова дополнително ја зајакнува довербата во применетата батерија од тестови како валиден и сигурен инструментариум за следење на фитнес-профилот и физичкиот развој кај младата популација.

**Методолошки и популациски влијанија:** Некои разлики меѓу нашите резултати и поединечни студии може да се објаснат со:

*Возраст на испитаниците:* Помладите деца (под ~8 години) генерално покажуваат пониска релијабилност во фитнес тестови поради недоволна способност за повторување на задачата исто (концентрација, техника). Ова е евидентно кај претклонот и агилноста, каде што предучилишните деца имаат само умерена до добра релијабилност. Во адолесцентската група (15–18 год.), варијабилноста е помала и релијабилноста повисока – што нашите наоди го потврдуваат.

*Протокол на тестирање:* Бројот на дозволени обиди и инструкциите пред првиот тест се круцијални. Ако на испитаниците не им се даде шанса да се запознаат со тестот, вториот пат речиси сигурно ќе бидат подобри (ефект на учење). Нашата студија покажа мал систематски напредок во скокот и значаен во стомачните преси, што наведува дека првиот пат можеби немале оптимална техника/напор. Литературата сугерира едноставно решение дозволување 1–2 пробни обиди или тренинг-сесија, особено за тестови каде техниката игра улога (скок, агилност). Исто така, одморот меѓу обидите и деновите треба да биде доволен за да нема заморен ефект. Во нашата процедура интервалот од една недела беше соодветен и споредлив со другите студии (кои често користат 1 недела или 7–10 дена).

*Мотивираност и инструкции:* Тестовите како стомачни преси зависат од желбата на ученикот да истрае до крај. Доколку мотивацијата не е конзистентна, варијабилноста се зголемува. Обезбедување стандардни вербални поттикнувања и јасно кажување на целта („колку повеќе, толку подобро“) при секое тестирање помага резултатите да бидат повторливи. Во студиите каде наставниците беа вклучени, се нагласува дека добра комуникација и охрабрување се дел од причините за високата релијабилност во школски услови.

*Опрема и мерење:* Како што беше дискутирано, употребата на технички помагала (фото-келиии, дигитални уреди) ја намалува грешката. Во рамките на нашиот компаративен преглед, студиите со подобра опрема (на пример, фотокелиии за време, дигитална платформа за скок) пријавуваат исклучително мали грешки. Сепак, дури и со рачно мерење, ако е конзистентно, се добиваат доволно доверливи резултати што е важно за практичната применливост во училишта.

*Варијабилност на испитаниците:* Некои истражувања нотираат дека релијабилноста може да е малку пониска кај екстремни групи на пр. многу слабо или многу фитнес деца. Пример е тестот на издржливост (шатл-ран 20 m) каде консензусот е висок ИСС за просечни групи, но варијабилност кај најмалите деца или тие со ниска мотивација. Во нашиов труд се фокусиравме на здрави адолесценти, па немавме таков екстреман спектар. Но, вреди да се спомене дека при примена на тестовите во различни популации (пример, деца со посебни потреби, или спортисти со многу висок перформанс) може да се појават различни предизвици за релијабилност кои овде не се изразени.