

# Мета-аналітичне дослідження критеріальної точності валідних поліграфологічних методик

Звіт, підготовлений для Ради директорів Американської Поліграфологічної Асоціації

*[Meta-Analytic Survey of Criterion Accuracy of Validated Polygraph Techniques. Report Prepared For The American Polygraph Association Board of Directors Nate Gordon, President (2010-2011) by The Ad-Hoc Committee on Validated Techniques Mike Gougler, Committee Chair Raymond Nelson, Principal Investigator Mark Handler, Donald Krapohl, Pam Shaw, Leonard Bierman. Polygraph, 2011. 40(4). P. 193–305].*

Мета-аналіз опублікований в журналі *Polygraph*, 2011. 40(4). P.193–305. Всі права на матеріали належать Американській Поліграфологічній Асоціації (АПА). Переклад та публікація зроблені з дозволу АПА. Подальші публікація, копіювання та розповсюдження мають проводитись з дозволу АПА.

## Нейт Гордон

президент (2010-2011)  
Спеціального комітету з валідних методик

## Майк Гуглер

голова комітету

## Реймонд Нельсон

провідний дослідник

## Марк Хендлер

## Дональд Крепол

## Пем Шоу

## Леонард Бірман

США

переклад:

## Діана Алексєєва-Процюк

## Дмитро Кушнір

## Максим Рудницький

## Олег Котловський

м. Київ, Україна

## Вступ

**Памела Шоу** (Pamela Shaw), президент Американської Поліграфологічної Асоціації

Протягом останніх кількох років АПА старанно впроваджувала низку важливих ініціатив, які стосуються її членів. Під впливом доповіді Національної академії наук за 2003 рік і про-стимульовані поточним оглядом експертно-криміналістичної (forensic science) спільноти, керівники та члени АПА визнали,

що попереду нас чекають реальні небезпеки, якщо ми продовжуватимемо йти тим шляхом, яким завжди йшли.

Для стратегічного планування та забезпечення нашого виживання в наступні роки, АПА робить ініціативи, щоб впровадити необхідні вдосконалення. Зокрема, АПА прагне підвищити науковий рівень у нашій практиці, стандартизувати нашу методологію, зосередитися на постійному вдосконаленні, покращити нашу освіту та розширити наше бачення, щоб охопити не лише інтереси членів АПА, а також і захист громадськості. Іншими словами, ми активно прагнемо підвищення професіоналізму в поліграфології.

Для досягнення цієї мети в Стратегічному плані АПА визначено кілька ключових орієнтирів на цьому шляху. Вони включають такі заходи, як розробка найкращої моделі політики щодо практики, обов'язкове безперервне навчання та розміщення статей про поліграфологічні дослідження на вебсайті АПА для легкого доступу до них членами АПА. Одним із ключових кроків є вимога до членів АПА починаючи з 2012 року використовувати методи, які можна захистити опублікованими дослідженнями. Саме цей останній крок є темою цього спеціального випуску *Polygraph*.

Вимога використання перевірених (валідних) методик тестування, звичайно, не є новою ідеєю. Інші галузі, такі як медицина та психологія, зрештою прийшли до такого ж висновку, хоча й через багато років після початку практичного застосування. Для них це виявилось великою справою. Спробуйте уявити, якими були б галузі медицини та психології, якби не було вимоги щодо валідності їхніх методів. Валідація (підтвердження) виконує низку важливих функцій, не останньою з яких є захист громадськості від зловживання, некомпетентності та шарлатанства. Кінцевий результат полягає не лише в допомозі громадськості, а й у підвищенні статусу професії; це найкраще рішення для всіх. Саме з цієї точки зору «просвітленого егоїзму» виникають ініціативи в сфері поліграфа в останні роки.

Коли ми наблизилися до 2012 року та до вимог встановлення валідності наших методик, стало зрозуміло, що деякі члени АПА, можливо, не отримали інформації від своїх поліграфологічних шкіл щодо методів, які мають наукове обґрунтування. Здавалося розумним, що АПА слід провести огляд літератури для встановлення наявних доказів і представити їх членам АПА. На початку цього року тодішній президент Нейт Гордон створив спеціальний комітет для виконання цього завдання. Я пишаюсь тим, що працювала у цьому комітеті. Протягом наступних місяців комітет старанно працював над завершенням звіту, який міг бути доступний членам АПА до кінця року. На наступних сторінках ви побачите їхні висновки.

Коли ви читаєте звіт комітету, вам слід знати кілька застережень. По-перше, як зазначено в самому звіті, він не є політикою АПА, а є узагальненням літератури. Є інші огляди літератури, опубліковані АПА за останні 20 років, які також не були політикою АПА. Цей звіт лише для інформування та може бути корисним для членів, які вже використовують або мають намір використовувати перелічені методики.

По-друге, Раді стало відомо про незавершені дослідження щодо скринінгових методик. Рада розгляне нові положення внутрішніх правил про скринінгову методологію, щоб заохотити членів використовувати найкращі доступні методики. Ці рішення будуть прийняті на початку 2012 року.

І, нарешті, АПА докладе чимало зусиль на своїх семінарах, щоб допомогти членам ознайомитися та навчитися методикам, які відповідають вимогам валідності.

Багато з вас висловлювали підтримку, підбадьорювали та захищали поточні зусилля щодо валідації та підвищення стандартів: за це я вам дякую. Я також хотіла би висловити вдячність членам комітету за невтомні та віддані зусилля, за незліченну кількість годин, які вони присвятили цьому проєкту. Ваші зусилля є ключовими для наших майбутніх починань.

Ми переживаємо великий етап в історії поліграфа, і ми можемо пишатися кроками, які робимо для просування нашої професії. Щоб забезпечити наш майбутній успіх, ми всі маємо зростати та відповідати вимогам у цій галузі. Цей звіт є важливим кроком у цьому напрямку, і я з гордістю представляю цей документ у спеціальному випуску *Polygraph*.

## Резюме

У 2007 році Американська Поліграфологічна Асоціація (АПА) прийняла Стандарт практики, який набув чинності 1 січня 2012 року, який вимагає від членів АПА використання валідних поліграфологічних методик, які відповідають певним рівням критеріальної точності<sup>1</sup>. Ці вимоги передбачають наступне: діагностичне дослідження щодо конкретної події, яке використовується для доказових цілей, має проводитися з використанням методик, які дають середній рівень критеріальної точності .90 або вище, з невизначеним показником .20 або нижче. Діагностичні дослідження, проведені з використанням протоколу для парного тестування, повинні мати середній рівень критеріальної точності .86 або вище, з невизначеним показниками .20 або нижче. Тестування, що проводяться з дослідницькою метою, повинні проводитися з використанням методик, які мають середній рівень критеріальної точності .80 або вище, з невизначеними показниками .20 або нижче<sup>2</sup>. **Метою є припинення використання нестандартизованих, невалідних або експериментальних методик в польових умовах, де рішення можуть вплинути на життя окремих людей, громадську безпеку, професійну доброчесність і національну безпеку.**

Сьогодні існує незрозумілий набір форматів тестових запитань, які водночас схожі та несхожі, і для яких також існують альтернативи у виборі методу аналізу тестових даних. Є велика кількість опублікованих наукових досліджень щодо методик, заплутаних за своїм значенням та застосуванням, що використовуються в польових умовах.

Рада директорів АПА взяла на себе відповідальність та систематизувала метааналіз опублікованої наукової літератури, що описує критерії валідності доступних на цей час поліграфологічних методик. У ході цієї роботи часом необхідно було визначити те, що здається очевидною концепцією. Однією з таких концепцій є валідація (валідність), яка стосується поліграфологічних досліджень, і передбачена Стандартами практики АПА (розділ 3.2.10) для позначення комбінації: 1) формату тестових запитань, який відповідає чинним принципам для вибору цілі, побудові

<sup>1</sup> Критеріальна точність, як правило, означає ступінь, в якому результат тесту відповідає тому, для чого він призначений. У поліграфології критеріальна точність означає здатність комбінації методик тестування та оцінювання розрізняти правдивих і оманливих досліджуваних і коливається від 0.00 при відсутності валідності до 1.00 для повної валідності. Критеріальна точність є однією з форм валідності, і в деяких дослідницьких звітах вона може згадуватися як точність рішення або просто точність.

<sup>2</sup> Незабаром після завершення цього дослідження Рада директорів АПА прийняла зміни в стандарти, схвалюючи використання скринінгових методик, для яких дослідження вказують на рівень точності, який значно перевищує випадковість.

запитання та пред'явлення тестових стимулів у тесті, а також 2) валідного методу аналізу тестових даних, що стосується конкретного формату тестових запитань. Незважаючи на те, що багато факторів можуть вплинути на загальну ефективність поліграфологічних досліджень, ці дві частини визнаються фундаментальними для критеріальної точності поліграфологічних досліджень. Точність усіх тестів залежить від цих двох видів діяльності: отримання достатньої кількості діагностичної інформації та правильна інтерпретація інформації. Подвійна мета цього мета-аналізу полягала в тому, щоб повідомити АПА та її членів про те, які поліграфологічні методики відповідають вимогам стандартної практики, станом на 1 січня 2012 року, і відповісти на запитання щодо нашої поточної бази знань щодо критеріальної валідності поліграфологічних методик, які зараз застосовуються.

Під час засідання Ради директорів у березні 2011 року президентом АПА Нейтом Гордоном був створений спеціальний комітет для вивчення доказів щодо критеріальної точності поліграфологічних методик. Комітет складався з колишнього президента та директора Ради Майка Гуглера (Mike Gougler) (голова комітету), колишнього президента та головного редактора Дона Крепола (Don Krapohl), обраного президента Пем Шоу (Pat Shaw), директора Ради Реймонда Нельсона (Raymond Nelson) та членів АПА Марка Хендлера (Mark Handler) та Леонарда Бірманна (Leonard Bierman).

Також, комітетом враховано під час формування списку методик наявність фінансових та майнових інтересів зацікавлених сторін, які представляють різноманітні групи поліграфологів та інтереси. Ефективність АПА та довіра до професії поліграфолога вимагали від комітету віддати перевагу точності та надійності огляду дослідження над фінансовими та особистими інтересами будь-якого окремого розробника методик поліграфологічного тестування. Стандартним підходом комітету був комплексний процес розгляду, у межах якого будь-яка зацікавлена сторона могла подати допоміжні дані та інформацію для розгляду. Це не означало, що все, що представляється, буде автоматично включено або схвалено як валідне, але це означало, що всі рекомендації будуть розглянуті.

Комітет розпочав свій процес з обговорення переваг і сильних сторін лабораторних і польових досліджень. Польові дослідження є важливими для поліграфологічних досліджень, оскільки вони мають перевагу екологічної валідності<sup>3</sup>, і тому передбачається, що вони мають підвищену можливість узагальнення. Проте узагальненість польових досліджень певною невідомою мірою порушується в результаті процесу відбору, який обов'язково залежить від наявності часто неповних даних підтвердження. Реальні дані підтвердження є вибірковими, не є ні випадковими, ні репрезентативними для всіх даних, і підтверджені випадки частіше можуть мати правильні поліграфологічні результати, ніж непідтверджені випадки. У результаті польові дослідження можуть певною мірою переоцінювати точність поліграфологічних рішень. Хоча польові дослідження дуже корисні для вивчення кореляцій, вони дають недосконалі показники критеріальної валідності.

Лабораторні дослідження також важливі для поліграфологічних досліджень, оскільки ці дослідження легше контролювати та зменшувати похибки досліджен-

---

<sup>3</sup> Екологічна валідність стосується того, наскільки добре експериментальні умови, процеси, суб'єкти та матеріали відповідають умовам реального життя. Хоча це не те ж саме, що зовнішня валідність, більша екологічна валідність може забезпечити більшу впевненість в тому, що результати дослідження зможуть бути узагальненими для інших умов.

ня та вибірки. Використання експериментальних та квазіекспериментальних дослідницьких проєктів разом із випадковою вибіркою та випадковим призначенням критеріїв може підвищити можливість узагальнення та повторюваності результатів дослідження. Завдяки своїй здатності контролювати більшу кількість змінних, лабораторні дослідження є основоположними для нашої здатності вивчати питання причинно-наслідкового зв'язку та конструювати валідність. Однак можливість узагальнення лабораторних досліджень ускладнюється тим фактом, що вони можуть не відображати широкий діапазон змінних, які, як вважають, впливають на результати польових досліджень. Тому передбачається, що вони мають екологічну валідність, яка до певної невідомої міри слабша, ніж польові дослідження.

Комітет вважає, що як польові, так і лабораторні дослідження мають свої переваги та недоліки, і що жоден тип сам по собі не буде достатнім для вивчення всіх проблем, які хвилюють поліграфологів. Обидва типи досліджень мають життєво важливе значення для вивчення та розвитку знань для професії поліграфолога. Відмінності між критеріальною точністю польових і лабораторних досліджень були визнані статистично невеликими і незначущими в звіті про поліграф, підготовлений Національною дослідницькою радою США в 2002 році. З метою перегляду поточного стану валідності існуючих поліграфологічних методик, комітет приділяв рівну увагу польовим і лабораторним дослідженням.

Комітет склав перелік досліджень, які відповідали якісним та кількісним вимогам для включення в огляд, і для яких було дві або більше належні публікації, які описують узагальнені докази критеріальної валідності. Комітет сформулював рекомендації щодо того, які методики задовольняють вимоги положень АПА 2012 щодо очікуваної критеріальної точності. Потім комітет завершив мета-аналіз, в якому містились висновки щодо співставлення зі стандартами АПА 2012 року відносно доказових, парних та дослідницьких процедур. Крім того, були проведені статистичні тести для перевірки об'єктивності дослідження та для пошуку невідповідностей і результатів-викидів.

Однією з умов включення наукового дослідження до огляду (вивчення) мета-аналізу мала бути його публікація в журналі *Polygraph* або інших рецензованих наукових публікаціях<sup>4</sup>. Дослідження розглядалися для включення в мета-аналіз, якщо вони були опубліковані установою, що присуджує наукові ступені, акредитованою визнаною Міністерством освіти США або аналогічним іноземним закладом акредитаційною агенцією. Крім того, для відбору також розглядалися дослідницькі наукові публікації, які фінансувались державними установами. Розглянуто відредаговані наукові тексти, у тому числі окремі розділи. Однак наукові дослідження лише у власноруч опублікованих виданнях не включались в огляд. Додаткові якісні вимоги до вибірки наукових досліджень передбачали використання в них поліграфологічної методики, до якої є опублікований опис структури тесту, визначена послідовність тестових запитань та адміністрування. У дослідженнях також повинен використовуватися відомий метод аналізу тестових даних (TDA) і включений опис інформативних ознак, числові перетворення (присвоєння балів), правила прийняття рішень і нормативні дані або порогові значення. Крім того, у дослідженнях повинні вико-

---

<sup>4</sup> Журнал *Polygraph* започаткував експертне рецензування у 2003 році. Статті, опубліковані до того часу, підлягали лише редакційній рецензії. Оскільки *Polygraph* є важливим науковим та історичним ресурсом, дослідження, опубліковані до 2003 року і які не мали належного рецензування, були включені в цей мета-аналіз, якщо вони задовольняли всім іншим якісним і кількісним вимогам для відбору.

ристовуватися прилади та компонентні давачі, які відображають загальну польову практику. У вибраних дослідженнях повинні були використовуватися різноманітні критерії підтвердження, включаючи зізнання досліджуваних, високоякісні криміналістичні докази або обґрунтовані докази того, що злочин не було вчинено<sup>5</sup>. Результати дослідження, засновані на вибірках, які були предметом експериментальних маніпуляцій (наприклад, втома, інтоксикація, заплановане застосування контрзаходів), не включались<sup>6</sup>.

Вимоги до кількісної інформації включали певну форму статистики надійності для кожної методики як доказ узагальненості результатів. В опублікованих джерелах йшлося про кілька типів статистичних вимірювань надійності поліграфологічних тестів, і всі вони були прийняті для включення в розділ<sup>7</sup>. Комітет також оцінив надійність, можливість, узагальненість та репрезентативність вибірових розподілів за допомогою багатовимірного ANOVA з використанням оманливих і правдивих оцінок. Очікувалося, що множинні вибірки, відібрані з тієї самої основної популяції, до яких застосовуються ті самі поліграфологічні методики і які оцінені за допомогою того самого методу оцінювання (TDA), повторюватимуться серед вибірових розподілів балів. Також очікувалось, що агрегування результатів повторюваних розподілів вибірки буде більш репрезентативним і придатним для узагальнення, ніж результати будь-якого одиничного розподілу вибірки. На додаток до інформації про розмір вибірки, для мета-аналізу були потрібні щонайменше чотири статистичні значення: чутливість і специфічність тесту, а також невизначені показники випадків щодо обману і правди.

Для достовірності висновків комітету було важливо, щоб наукові дослідження були прийняті або схвалені на основі переваг включених у них досліджень. З цієї причини пріоритет був наданий у випадку наявності опублікованих доказів та вихідних даних, а список підтверджених методик було складено відповідно до систематичного огляду опублікованих досліджень. На додаток до повторного аналізу даних були включені дослідження, на основі яких можна було розрахувати повний розмірний профіль критеріальної точності.

Після завершення дослідження літератури список усіх ідентифікованих методик було надіслано директорам шкіл або представникам усіх акредитованих АПА поліграфологічних навчальних закладів. Директорам шкіл або їхнім представникам було запропоновано надати будь-які опубліковані дослідження або посилання на опубліковані дослідження щодо методик, які ще не були визначені. У той час було запропоновано включити лише одну методику<sup>8</sup>. У подальшому всім директорам шкіл, представникам та іншим дослідникам, які беруть участь у розробці або перевірці поліграфологічних методик, було розіслано коротший список. Цей список включав лише ті методики, дослідження щодо яких було опубліковано та до яких опубліковано повторене дослідження, разом із іншим запитом надати будь-які пу-

<sup>5</sup> Одне дослідження не відповідало цій вимозі та містило лише випадки, підтвержені зізнаннями. У відповідності до відомого занепокоєння щодо завищеної точності, що є результатом надмірної залежності від підтвердження зізнання для вибірки випадків, це дослідження показало майже ідеальний рівень точності рішень.

<sup>6</sup> Існуючі стандарти для досліджень та публікацій АПА передбачають, що головні дослідники не повинні також виступати в якості учасників дослідження (тобто досліджуваних, поліграфологів або оцінювачів). Однак у минулому цієї вимоги не було, і дослідження не були виключені з мета-аналізу на основі цього критерію.

<sup>7</sup> Одна методика, яка була включена, не має опублікованих доказів надійності або узгодженості.

<sup>8</sup> Було запропоновано включити методику Магсу. Однак опублікованих досліджень щодо цієї методики знайти не вдалося.

блікації чи посилання на методики, які ще не були включені в огляд. У той час було запропоновано включити ще два додаткові дослідження<sup>9</sup>. Після подання початкових результатів Раді директорів АПА та представлення попередньої версії цього огляду членам АПА, було подано ще один додатковий, нещодавно опублікований звіт про дослідження на підтримку IZCT<sup>10</sup>.

Комітет зв'язався з розробниками поліграфологічних методик, щодо яких було недостатньо опублікованих досліджень для включення в мета-аналіз, і запропонували допомогу всім, хто просив про це. Два дослідження щодо методики Backster You-Phase були завершені, прийняті до публікації та очікували на друк. Ця методика була включена в мета-аналіз. Крім того були проведені додаткові дослідження для Air Force Modified General Question Test (AFMGQT), Federal You-Phase Technique та Directed Lie Screening Test (DLST), які також були прийняті до публікації. Результатом цих додаткових зусиль стало те, що до мета-аналізу було включено повний набір методик, які на сьогодні широко застосовуються.

Тридцять вісім досліджень відповідали якісним та кількісним вимогам для включення в мета-аналіз. Ці дослідження включали 32 різні вибірки та описували результати 45 різних експериментів та оглядів. Ці дослідження включали 295 оцінювачів, які надали 11737 результатів 3723 досліджень, включаючи 6109 оцінювань з 2015 підтвердженими результатами обману, 5628 оцінювань з 1708 підтвердженими результатами правди. Деякі з випадків були оцінені кількома оцінювачами та з використанням кількох методів аналізу тестових даних (TDA).

В Таблиці 1 наприкінці цього огляду представлені результати для тих методів, для яких є опубліковані та повторені докази критеріальної валідності, що відповідають вимогам Стандартів практики АПА. П'ять комбінацій поліграфологічного тестування та аналізу відповідають вимогам АПА 2012 для доказового тестування, п'ять – для парного тестування<sup>11</sup> і чотири – для дослідницьких процедур<sup>12</sup>.

Дві поліграфологічні методики продемонстрували показники точності, які відрізнялися від розподілу результатів усіх інших методик і не узгоджувалися з ним. Це були Integrated Zone Comparison Technique (IZCT) і Matte Quadri-Track Zone Comparison Technique (MQTZCT). Хоча ймовірно, що ці дві методики перевершують інші методики, дослідження на їх підтримку виявили більше невирішених методологічних тем, ніж інші, включені в цей мета-аналіз. На додаток до виявлення комітетом аномальних розподілів вибірки, обидві ці методики підтверджуються дослідження-

<sup>9</sup> Дослідження Integrated Zone Comparison Technique (IZCT), проведене Гордоном та іншими дослідниками (2000), не вдалося включити через відсутність належної статистичної інформації. Головний автор повідомив комітет, що доповідь було завершено, але він не мав доступу до даних, які належать розвідувальній службі іноземного уряду. Дані для дослідження Мангана, Армітіджа та Адамса (Mangan, Armitage та Adams, 2008) щодо Matte Quadri-Track Zone Comparison Technique були надані комітету, і це дослідження було включено.

<sup>10</sup> Частина даних для дослідження IZCT Шурані (Shurani) (2011) була надана комітету, і це дослідження було включено.

<sup>11</sup> **Усі поліграфологічні методики, які відповідають вимогам щодо критеріальної точності для парного тестування, також відповідають стандартним вимогам для дослідницького тестування, а ті методики, які відповідають стандартним вимогам для доказового тестування, також відповідають вимогам для парного та дослідницького тестування.**

<sup>12</sup> Усі методики, які використовували 3-позиційні методи аналізу тестових даних (TDA), постійно перевищували граничні вимоги 2012 року щодо невизначених показників (20%). Оскільки рівень критеріальної точності для 3-позиційних методів аналізу тестових даних не відрізнявся суттєво від критеріальної точності 7-позиційної системи оцінювання, польові практики, які передбачають початковий аналіз за допомогою 3-позиційної системи оцінювання, можуть вважатися прийнятними, якщо невизначені результати будуть вирішені шляхом наступного аналізу за допомогою методу TDA, який забезпечує як точність, так і невизначені показники, які відповідають вимогам стандартів АПА 2012.

ми, авторами яких є розробники та власники, у яких розробник/власник виступав як головним дослідником, так і учасником дослідження. З наукової точки зору, навіть добре сплановане наукове дослідження, проведене прихильниками методу, які мають особисту зацікавленість у результаті та які виступають як учасники та автори звіту про дослідження, не має переконливої сили, як ті наукові дослідження, які не настільки обтяжені цими факторами. Ці методики були належним чином включені сюди, оскільки вони відповідали більш загальним вимогам, викладеним у Стандартах практики АПА. Комітет радить, що через потенційний вплив на ефективність дослідників, який може бути обумовлено використанням аномальних результатів, було б розумно, щоб поліграфологи проявляли обережність, перш ніж приймати дані наукових досліджень, які демонструють надзвичайні ефекти, перш ніж вони будуть підлягати незалежному підтвердженню та поглибленому аналізу.

Профіль критеріальної точності був розрахований для кожної поліграфологічної методики, включаючи показники незважених середніх значень правильних рішень для випадків обману і правди, за винятком невизначених результатів, а також показники незважених середніх невизначених результатів<sup>13</sup>. Результати були агреговані для методик, які відповідають вимогам АПА 2012 щодо доказового тестування, парного тестування, дослідницького тестування та для всіх поліграфологічних методик, включених до мета-аналізу. За винятком аномальних результатів (викидів), методики із запитаннями порівняння, призначені для діагностичного одностороннього одноаспектного тестування конкретної події (*event-specific (single issue) diagnostic testing*), в якому дисперсія критерію кількох релевантних запитань вважається не-незалежною (*non-independence*)<sup>14</sup>, дали загальний показник точності рішення .890 (.829 – .951), із сукупним невизначеним показником .110 (.047 – .173). Поліграфологічні методики із запитаннями порівняння, розроблені для інтерпретації з припущенням незалежності дисперсії критерію кількох релевантних запитань, дали загальний показник точності рішень .850 (.773 – .926) із сукупним невизначеним показником .125 (.068 – .183). Комбінація всіх валідних поліграфологічних методик, за винятком аномальних результатів (викидів), дала рівень точності рішення .869 (.798 – .940) з невизначеним показником .128 (.068 – .187). Даних на даний момент достатньо, щоб підтвердити, що поліграф має високу точність, але недостатньо, щоб підтвердити твердження, що поліграфологічне тестування може забезпечити ідеальну або майже ідеальну точність.

За винятком аномальних результатів (викидів), результати багатомірного аналізу продемонстрували відсутність значних односторонніх відмінностей у точності рішень для будь-якої поліграфологічної методики, яка відповідає вимогам стандартів практики АПА 2012. Також не було жодних значних односторонніх відмінностей в поліграфологічних методиках на різних рівнях встановлення достовірності

<sup>13</sup> Незважене середнє значення вважалося більш консервативним і реалістичним розрахунком загальної точності всіх досліджень щодо поліграфологічних методик. Розрахунок середньозваженої величини або простої частки правильних рішень часто призводить до вищих статистичних результатів, які є менш надійними до різниць у базових ставках і, отже, менш узагальненими.

<sup>14</sup> У науковому тестуванні це означає припущення про те, що зовнішні фактори, які впливають на стан критерію кожного запитання (тобто правдивість щодо минулої поведінки), впливають на стан критерію інших запитань. Під час поліграфологічного тестування результати багатоаспектних (*multi-facet*) і багатотемних (*multi-issue*) досліджень інтерпретуються з правилами прийняття рішень, заснованими на припущенні про незалежність, тоді як результати одностороннього одноаспектного дослідження, пов'язаних з конкретною подією (*event-specific single-issue examinations*), частіше інтерпретуються з правилами прийняття рішень, заснованими на припущенні про відсутність незалежності.



(валідації), зазначених у стандартах практики АПА, або для поліграфологічних методик, інтерпретованих з правилами прийняття рішень на основі припущення про незалежність, у порівнянні з поліграфологічними методиками, інтерпретованими з використанням правил прийняття рішень на основі припущення про відсутність незалежності (*non-independence*). Це показує, що категоричні відмінності АПА є довільними, емпірично не обґрунтованими та науково безглуздими. Цих даних недостатньо, щоб підтвердити думку про те, що будь-яка поліграфологічна методика є кращою за іншу, і натомість припускають, що відмінності у поліграфологічних форматах запитань можуть бути менш важливими, ніж передбачалося раніше. Відмінності в поліграфологічних методиках можуть обмежуватися припущеннями та процедурними відмінностями, що стосуються методології TDA (систем оцінювання), призначених для дослідження незалежних або не-незалежних цільових запитань дослідження. Це має бути предметом подальшого вивчення.

Наявні дані підтверджують аргумент, що поліграфологічне тестування може забезпечити як чутливість тесту до обману, так і специфічність тесту до правди з показниками, які значно перевищують випадковість, якщо проводиться та інтерпретується з припущеннями незалежності критерію, а також не-незалежності (*non-independence*) запитань тесту. Фактичні дані показують, що всі поліграфологічні методи, включені в мета-аналіз, забезпечують чутливість тесту з показниками, які значно перевищують випадковість. Проте наявних доказів недостатньо, щоб підтвердити, що кожна поліграфологічна методика здатна забезпечити специфічність тесту для правдивості з показниками, які значно перевищують випадковість. Подробиці можна знайти в повному звіті.

Якщо вважати поточні результати завищеною точністю в поліграфології, головною причиною такого завищення будуть недоліки в методології вибірки. Одним із таких факторів є те, що занадто покладаються на підтвердження випадку через зізнання досліджуваного, що може представляти потенціал для систематичного виключення хибнопозитивних або хибнонегативних помилок, щодо яких не могло бути отримано зізнання. Ще однією проблемою, пов'язаною з вибіркою, може бути систематична упередженість у відношенні публікації або файлів, коли менш сприятливі результати не надаються для публікації і, отже, недоступні для включення в мета-аналіз або інший систематичний огляд. Іншою потенційною причиною завищення точності може бути відсутність незалежності між розробником методики, головним дослідником та дослідниками, які беруть участь у деяких включених дослідженнях. Якщо нинішні результати є заниженою оцінкою точності в поліграфології, основною причиною можна вважати недоліки в екологічній валідності експериментальної та дослідницької методології включених наукових досліджень. Наведені результати призначені лише для узагальнення доступних на даний момент публікацій, які задовольняють вимогам для включення в мета-аналіз. Обмеження (недоліки) мета-аналізу обговорюються в повному звіті.

На заключення слід додати, що не потрібно намагатися представити результати цього мета-аналітичного огляду як обов'язкову політику чи стандарт. Незважаючи на те, що розповсюдження переліку валідних поліграфологічних методик дехто може розглядати як форму фактичного схвалення цих методик АПА, фактична роль цього мета-аналізу полягає в поглибленому підсумку існуючої літератури в поліграфології. Хоча політика може залишатися незмінною протягом певного періоду часу, наукові докази постійно вдосконалюються. Комітет вважає, що питання та об-

говорення валідності тестів є питанням науки, а не просто політики. Таким чином, на ці запитання найкраще відповідати науковими доказами. Цей мета-аналіз слід розглядати лише як інформаційний ресурс, і не слід намагатися представити цей перелік поліграфологічних методів як остаточну інстанцію щодо валідності поліграфологічних тестів. Хоча це було завершено з метою створення вичерпного та всеосяжного списку валідних методик, залишається ймовірним, що інші дослідження та методики існують, але не були включені в цей мета-аналіз. В опублікованій літературі існують деякі докази валідності поліграфологічних методик, які не вдалося включити в це мета-аналітичне дослідження. Звичайно, мета-аналіз, заснований на різних критеріях відбору або включення в дослідження, може дати різні результати. Нічого в цьому короткому викладі або повному звіті не слід тлумачити як перешкоду використанню будь-якої поліграфологічної методики, для якої рівень критеріальної точності можна аргументувати науковими доказами. Інформація, наведена в цьому документі, надається Раді АПА, щоб поінформувати її професійних членів про доступні на даний момент валідні поліграфологічні методики. Ця інформація призначена лише для того, щоб полегшити навантаження на поліграфологів і допомогти прийняти рішення на основі доказів щодо вибору поліграфологічних методик для застосування в польових умовах. Це також може допомогти адміністраторам програм, політикам і судам приймати рішення на основі доказів щодо інформаційної валідності результатів поліграфологічних тестів загалом.

<p>Таблиця 1. Середнє значення (стандартне відхилення) і [95% довірчі інтервали] для правильних рішень (CD) і невизначених результатів (INC) для валідних поліграфологічних методик. Список літератури можна знайти в кінці повного звіту.</p>	<p><b>Evidentiary Techniques/ TDA Method</b> Доказові методики / метод аналізу тестових даних</p>	<p><b>Paired Testing Techniques/ TDA Method</b> Методики для парного тестування / метод аналізу тестових даних</p>	<p><b>Investigative Techniques/ TDA Method</b> Дослідницькі методики / метод аналізу тестових даних</p>
<p><b>Federal You-Phase / ESS<sup>1</sup></b> CD = .904 (.032) {.841 to .966} INC = .192 (.033) {.127 to .256}</p>	<p><b>AFMGQT<sup>4,8</sup> / ESS<sup>5</sup></b> CD = .875 (.039) {.798 to .953} INC = .170 (.036) {.100 to .241}</p>	<p><b>AFMGQT<sup>6,8</sup> / 7 position</b> CD = .817 (.042) {.734 to .900} INC = .197 (.030) {.138 to .255}</p>	<p><b>CIT<sup>7</sup> / Lykken Scoring</b> CD = .823 (.041) {.744 to .903} INC = NA</p>
<p><b>Event-Specific ZCT / ESS</b> CD = .921 (.028) {.866 to .977} INC = .098 (.030) {.039 to .157}</p>	<p><b>Backster You-Phase / Backster</b> CD = .862 (.037) {.787 to .932} INC = .196 (.040) {.117 to .275}</p>	<p><b>CIT<sup>7</sup> / Lykken Scoring</b> CD = .823 (.041) {.744 to .903} INC = NA</p>	<p><b>DLST (TES)<sup>8</sup> / 7 position</b> CD = .844 (.039) {.768 to .920} INC = .088 (.028) {.034 to .142}</p>
<p><b>IZCT / Horizontal<sup>2</sup></b> CD = .994 (.008) {.978 to .999} INC = .033 (.019) {.001 to .069}</p>	<p><b>Federal You-Phase / 7 position</b> CD = .883 (.035) {.813 to .952} INC = .168 (.037) {.096 to .241}</p>	<p><b>DLST (TES)<sup>8</sup> / 7 position</b> CD = .844 (.039) {.768 to .920} INC = .088 (.028) {.034 to .142}</p>	<p><b>DLST (TES)<sup>8</sup> / ESS</b> CD = .858 (.037) {.786 to .930} INC = .090 (.026) {.039 to .142}</p>
<p><b>MQTZCT / Matte<sup>3</sup></b> CD = .994 (.013) {.968 to .999} INC = .029 (.015) {.001 to .058}</p>	<p><b>Federal ZCT / 7 position</b> CD = .860 (.037) {.801 to .945} INC = .171 (.040) {.113 to .269}</p>	<p><b>DLST (TES)<sup>8</sup> / ESS</b> CD = .858 (.037) {.786 to .930} INC = .090 (.026) {.039 to .142}</p>	<p>–</p>
<p><b>Utah ZCT DLT / Utah</b> CD = .902 (.031) {.841 to .962} INC = .073 (.025) {.023 to .122}</p>	<p><b>Federal ZCT / 7 pos. evidentiary</b> CD = .880 (.034) {.813 to .948} INC = .085 (.029) {.028 to .141}</p>	<p>–</p>	<p>–</p>
<p><b>Utah ZCT PLT / Utah</b> CD = .931 (.026) {.879 to .983} INC = .077 (.028) {.022 to .133}</p>	<p>–</p>	<p>–</p>	<p>–</p>
<p><b>Utah ZCT Combined / Utah</b> CD = .930 (.026) {.875 to .984} INC = .107 (.028) {.048 to .165}</p>	<p>–</p>	<p>–</p>	<p>–</p>

<u>Evidentiary Techniques/ TDA Method</u> <u>Доказові методики / метод аналізу</u> <u>тестових даних</u>	<u>Paired Testing Techniques/ TDA Method</u> <u>Методики для парного тестування /</u> <u>метод аналізу тестових даних</u>	<u>Investigative Techniques/ TDA Method</u> <u>Дослідницькі методики / метод</u> <u>аналізу тестових даних</u>
<p><b>Utah ZCT CPC-RCMP Series A / Utah</b>  CD = .939 (.038) {.864 to .999}  INC = .185 (.041) {.104 to .266}</p> <p><sup>1</sup> Емпірична система оцінювання (Empirical Scoring System).  <sup>2</sup> Можливість узагальнення цього виду результату обмежена тим фактом, що для цієї методики не було опубліковано показників надійності тесту. Крім того, були виявлені значні відмінності у розподілі вибірки включених наукових досліджень, що свідчить про те, що дані вибірки не є репрезентативними один для одного, або що дослідження проводилися та/або оцінювалися по-різному. Одне з наукових досліджень включало невелику вибірку (N = 12), про яку повідомлялося у двох статтях, для яких особа, що брала участь у оцінюванні результатів, також була розробником методики. В одній із публікацій описане дослідження як пілотне без сліпого оцінювання. В обох звітах зазначено, що одного з шести правдивих учасників було виключено з дослідження після того, як він зробив неправдиве зізнання. Повідомлений рівень абсолютної точності не включає неправдиве зізнання. Ані абсолютна точність, ані показник хибних зізнань .167 навіряд чи будуть узагальнюватися у польових умовах.  <sup>3</sup> Узагальнення цього аномального результату обмежена тим фактом, що розробники та дослідники повідомили про необхідність інтенсивного навчання, доступного лише для дослідничих практик методик, і припустили, що складність методики перевищує ту складність, якої інші професіонали можуть навчитися у опублікованих ресурсах. Розробник повідомив про майже ідеальний коефіцієнт кореляції .99 для числового оцінювання, що свідчить про безпрецедентно високий рівень узгодження між оцінювачами, що є неочікуваним, враховуючи передбачувану складність методу. Крім того, дані, спочатку надані комітету для реплікаційних досліджень, включали лише ті випадки, щодо яких оцінювачі прийшли до правильного рішення, за винятком оцінок тих випадків, щодо яких оцінювачі не досягли правильного рішення. Відсутні оцінки пізніше були надані комітету як для досліджень Мангана та ін. (Mangal, 2008), так і для Шурані та Чаваса (Shurani and Chavez 2009). Однак отримані середні значення вибірки відрізнялися від тих, що були зареєстровані для обох реплікаційних досліджень. Через ці розбіжності статистичний аналіз не було повторно обчислено з урахуванням відсутніх оцінок і звіт відображає середні значення розподілу вибірки, які було зазначено. Вибірку для повторних досліджень слід вважати позбавленими помилки або неконтрольованої дисперсії.  <sup>4</sup> Для AFMGQT існують дві версії з незначними структурними відмінностями між ними. Немає доказів того, що продуктивність однієї версії є кращою за іншу. Оскільки для відхилення нульової гіпотези про те, що відмінності є безглуздими, потрібні повторні докази, а також оскільки вибрані дослідження включають поєднання обох версій AFMGQT, ці результати надаються як узагальнені для обох версій. Дослідження з AFMGQT використовуються як у багатомасштабних дослідженнях щодо конкретної події, так і в багатомасштабних скринінгових дослідженнях. <b>І багатомасштабні, і багатомасштабні дослідження були інтерпретовані з правилами прийняття рішень, заснованими на припущенні про незалежність критеріїв серед релевантних запитань RQ.</b></p> <p><sup>5</sup> AFMGQT продемонстрував точність, яка є задовільною для парного тестування, лише якщо оцінювати його за допомогою емпіричної системи оцінювання.  <sup>6</sup> Існує дві методики, щодо яких немає опублікованих досліджень, але які структурно майже ідентичні AFMGQT: LEPEP і Utah MGQT. Валідність AFMGQT можна узагальнити для цих методик, якщо оцінювати за допомогою тих самих методів оцінювання тестових даних (TDA).  <sup>7</sup> Consealed Information Test (CIT), також відомий як Guilty Knowledge Test (GKT) (GKT) і Peak of Tension Test (POT). Дані, використані тут, були надані у звіті про мета-аналіз лабораторних досліджень МакЛареном (MacLaren, 2001).  <sup>8</sup> Дослідження цих поліграфологічних методик проводилися з використанням правил прийняття рішень, заснованих на припущенні про незалежність критеріїв між тестовими цілями. Точність скринінгових методик може бути додатково покращена шляхом систематичного застосування підходу послідовних бар'єрів.</p>	<p>—</p>	<p>—</p>

# Звіт Спеціального комітету з валідних методик

(Report of the Ad Hoc Committee on Validated Techniques)

## Анотація

Мета-аналітичні методи були використані для розрахунку розміру ефекту валідних методик щодо психофізіологічного виявлення обману (PDD), вираженого в термінах критеріальної точності. Для розрахунку статистичних довірчих інтервалів використовували методи Монте-Карло. Результати були підсумовані для 45 різних вибірок для експериментів і опитувань, які включали результати 295 оцінювачів, які надали 11 737 оцінок 3 723 поліграфологічних досліджень, включаючи 6 109 оцінок 2 015 поліграфологічних досліджень з підтвердженими результатами обману, 5 628 оцінок 1 708 поліграфологічних досліджень з підтвердженими результатами правди. Чотирнадцять різних поліграфологічних методик були підтверджені принаймні двома опублікованими науковими дослідженнями, кожне з яких задовольняло якісним та кількісним вимогам для включення в мета-аналіз. Результати для окремих досліджень і для різних поліграфологічних методик порівнювалися за допомогою багатовимірних методів аналізу. Два наукові дослідження дали результати, що виходять за межі, які не враховуються наявними доказами та які не підлягають узагальненню. За винятком викидів (аномалій, *outliers*), не було суттєвих відмінностей у критеріальній точності між будь-якими поліграфологічними методиками, які підтримуються вибраними науковими дослідженнями. За винятком аномальних результатів (викидів), методики із запитаннями порівняння, призначені для діагностичного тестування конкретної події (однотемне одноаспектне, *single issue*), у яких дисперсія критерію кількох релевантних запитань вважається не-незалежною (*non-independent*), дали сукупний агрегований показник точності рішення (*aggregated decision accuracy rate*) .890 (.829 – .951), із сукупним невизначеним показником .110 (.047 – .173). Поліграфологічні методики із запитаннями порівняння, розроблені для інтерпретації з припущенням незалежності дисперсії критерію кількох релевантних запитань (багатоаспектні та багатотемні) (*multiple-issue and -facet*), дали агрегований показник точності рішень .850 (.773 – .926) із сукупним невизначеним показником .125 (.068 – .183). Комбінація всіх валідних поліграфологічних методик, за винятком аномальних результатів (викидів), дала точність рішення .869 (.798 – .940) з невизначеним показником .128 (.068 – .187).

## Вступ

Чи поліграф має наукову достовірність? Наскільки точний поліграф? Просто-та цих поширених запитань означає, що точність поліграфологічних тестів можна описати простою відповіддю або єдиним числовим індексом. Нинішній підхід до відповідей на ці та інші питання про критеріальну валідність<sup>1</sup> – це мета-аналітичний

<sup>1</sup> Терміни «точність», «точність тесту», «критеріальна точність» та «критеріальна валідність» використовуються в цьому документі як взаємозамінні та синоніми. Термін «точність рішення» також використовується для опису критеріальної валідності, але в більш обмеженому сенсі, посилаючись лише на точність рішень, виключаючи невізначені результати. У більш повному розумінні термін «критеріальна точність» відноситься до сукупності проблем розмірності, що стосується всіх аспектів точності тесту.

огляд усіх доступних досліджень критеріальної точності для всіх поліграфологічних методик. Оскільки тест із запитаннями порівняння (CQT) є найбільш часто використовуваним і досліджуваним з усіх поліграфологічних методик, цей аналіз буде, в першу чергу, спрямований на CQT. Буде проведене обмежене обговорення досліджень щодо застосування використання Concealed Information Test (CIT).

Витоки всіх сучасних форматів CQT можна простежити до Ріда (Reid, 1947), який показав, що певна форма запитання порівняння (CQ), спрямована на те, щоб викликати реакцію правдивого досліджуваного, може підвищити точність тесту та зменшити кількість хибнопозитивних помилок. Формати CQT часто належать до однієї з двох основних сімей методик: методики, які виникли як модифікації методики, описаної Рідом (Reid) (1947), і методики, які виникли як модифікація методики, описаної Бакстером (Backster) (1963). Ці методики проводяться з використанням відмінних, але часто подібних процедур, які ґрунтуються на різних припущеннях. Ці різні припущення та процедури можуть призвести до відмінностей у продуктивності (ефективності) або точності тесту. Деякі методики є високотеоретичними щодо точної природи та причини емоційної чи когнітивної активності та відповідних психофізіологічних змін. Інші методики наголошують на науково-обґрунтованому підході, який виключає недоведені гіпотези і складні психологічні припущення щодо точних думок і емоцій досліджуваного.

Загалом, сімейство форматів Zone Comparison Test<sup>2</sup> (ZCT), що виникло в результаті роботи Бакстера (Backster) (1963), найбільш ефективно використовувалося для діагностичного тестування конкретних подій (*event-specific diagnostic testing*). Запитання ZCT сформульовані для опису участі досліджуваного в одній відомій або передбачуваній поведінковій проблемі, що викликає занепокоєння, і інтерпретуються за допомогою правил прийняття рішень, заснованих на припущенні про відсутність незалежності<sup>3</sup> від дисперсії критерію тестових запитань. На відміну від цього, сімейство форматів Modified General Question Test<sup>4</sup> (MGQT), яке виникло в результаті роботи Ріда (Reid, 1947), призначене для опису та оцінки участі досліджуваного в різних поведінкових ролях або різних рівнях участі у відомому чи можливому (*повідомленому*) інциденті. Хоча дослідження підтвердили, що CQT здатний забезпечити точність на рівні, який значно перевищує випадковість, попередні дослідження (Barland, Honts & Barger 1989; Podlesny & Truslow, 1993; Research Division Staff, 1995a, 1995b) не підтвердили ефективність поліграфологічного запитання під час дослідження щодо точного визначення поведінкової ролі або рівня участі, пов'язаного з конкретною подією. На додаток до їх використання в багатоаспектних розслідуваннях відомих або можливих інцидентів, методики MGQT легко адаптуються для використання в багатотемних скринінгових контекстах, у яких тестові запитання формулюються для опису можливої участі досліджуваного в кількох різних поведінках (діях, аспектах), для яких немає підтвердженого (відомого) інциденту або

<sup>2</sup> Іноді його називають історично правильним виразом Zone Comparison Technique, а також Zone **of** Comparison Technique.

<sup>3</sup> Незалежність у науковому тестуванні стосується припущень щодо того, чи припускається, що зовнішні фактори, які впливають на критеріальний стан кожного запитання (тобто правдивість щодо минулої поведінки), впливають на критеріальний стан інших запитань. Під час поліграфологічного тестування результати багатоаспектних і багатотемних досліджень інтерпретуються з правилами прийняття рішень, заснованими на припущенні про незалежність, тоді як результати одностороннього одноаспектного дослідження, пов'язаного з конкретною подією, частіше інтерпретуються з правилами прийняття рішень, заснованими на припущенні про відсутність незалежності.

<sup>4</sup> Також називається модифікованою методикою загальних запитань (Modified General Question Technique).

повідомлення (твердження). Як багатоаспектні, так і багатотемні дослідження з MGQT зазвичай інтерпретуються з правилами прийняття рішень, заснованими на припущенні, що критеріальна дисперсія стимулів релевантного запитання (RQ) є незалежною. У польовій практиці обидва сімейства методик інколи використовували припущення як незалежності, так і відсутності незалежності між релевантними запитаннями.

### **Інші огляди**

Попередні систематичні огляди були завершені, щоб дати об'єктивні відповіді на питання щодо поліграфологічної точності і узгодити це з заявами про досконалість. Абрамс (Abrams, 1973) проаналізував дослідження щодо валідності поліграфа, що відносяться до початку XX століття, і повідомив про рівень точності .980. Пізніше Абрамс (Abrams, 1977) повідомив, що середня точність досліджень щодо поліграфологічної валідності становить .910. Ще пізніше Абрамс (Abrams, 1989) узагальнив точність поліграфологічних тестів як .880.

Анслі (Ansley, 1983) повідомив про результати 1964 лабораторних випадків і 1113 польових випадків і описав рівень точності рішення .968, за винятком невизначених результатів. Однак показники точності не повідомлялися окремо для випадків обману та правди, а також не повідомлялась частка невизначених результатів. У той час методика Relevant-Irrelevant вважалася більш точною (.960), ніж методи CQT (.952). Анслі (Ansley, 1983) повідомив, що точність форматів СІТ становить .912. Пізніше Анслі (Ansley, 1990) узагальнив результати 10 наукових досліджень з польовими тестуваннями, що стосувались 2042 кримінальних справ з 1980 року, повідомляючи про загальний рівень точності .980 для випадків обману і .970 для випадків правди, використовуючи рішення поліграфологів, які безпосередньо проводили ці тестування. Анслі (Ansley, 1990) також описав результати 11 наукових досліджень щодо сліпого оцінювання 922 кримінальних поліграфологічних досліджень, повідомляючи про рівень точності .900, із заявленим показником точності .940 для випадків обману і .890 для випадків правди.

Хонтс і Петерсон (Honts & Peterson, 1997), Раскін і Хонтс (Raskin & Honts, 2002) повідомили, що точність поліграфа перевищує .900. Це узгоджується з оцінкою точності .900 Раскіна та Подлесного (Raskin & Podlesny, 1979). Навпаки, систематичний огляд, проведений Управлінням оцінки технологій (Office of Technology Assessment, ОТА, 1983), показав, що лабораторні дослідження мали середню незважену точність .832 з невизначеним показником .269, тоді як польові дослідження мали середній незважений рівень точності .847 з невизначеним показником .042. Крюсон (Crewson, 2001) проаналізував наукові дослідження щодо діагностичних та скринінгових тестувань<sup>5</sup> у порівнянні з медичними та психологічними тестами та повідомив, що поліграфологічні діагностичні тести мають середній рівень точності .880, а середня точність скринінгових поліграфологічних тестувань .740. Інформацію Крюсона (Crewson) також надав Блекстоун (Blackstone, 2011), який стверджував, що плутанина між діагностичним і скринінговим поліграфологічними тесту-

<sup>5</sup> Діагностичні тести – це будь-які тести, які проводяться у відповідь на відомі проблеми, відомі симптоми, відомі інциденти або відомі твердження. Скринінгові тести – це будь-які тести, які проводяться за відсутності відомої проблеми та призначені для пошуку можливих проблем. На практиці діагностичні тести зазвичай формулюються навколо однієї проблеми, яка викликає занепокоєння. Скринінгові тести, через відсутність будь-яких відомих проблем і через інтерес до кількох типів можливих проблем, часто будуються навколо кількох тем. Терміни багатотемний (multi-issue) і змішанотемний (mixed-issue) використовуються як синоніми. **Різницю між діагностичними та скринінговими тестами визначає не кількість тем, а наявність чи відсутність відомої проблеми.**

ваннями є причиною того, що поліграф не користується більшою підтримкою з боку закону<sup>6</sup>.

Останній науковий огляд був завершений Національною дослідницькою радою (National Research Council, NRC, 2003), яка повідомила про показники точності з точки зору AUC за допомогою аналізу ROC, і дійшла висновку, що лабораторні дослідження мали середню AUC .860, тоді як польові дослідження мали середню AUC .890.

Нещодавно Кокіш, Левенсон і Блєсінгем (Kokish, Levenson, Blasingame, 2005) повідомили про результати опитування сексуальних злочинців, які підлягали моніторингу на поліграфі після звільнення з місць позбавлення волі за умовами нагляду та лікування. Науковці зазначили, що правопорушники погодилися з високим рівнем згоди щодо результатів на поліграфі, понад .900, але зазначили, що правопорушники також стверджували, що рівень помилкового визнання був приблизно .050.

Незважаючи на певну цінність, жодне з цих попередніх опитувань не може забезпечити задовільний рівень вимог стандарту Американської Поліграфологічної Асоціації (АПА) 2012 року щодо використання валідних методик. Ці попередні дослідження не включають інформацію, що описує нещодавні досягнення щодо поліграфологічної критеріальної точності, і жодне з цих досліджень не виконує належної роботи, щоб забезпечити повний розмірний профіль критеріїв валідності окремих методик для поліграфологічних досліджень. Що ще важливіше, жодне з цих попередніх опитувань не задовольняє потребу в зведеній інформації щодо публікації дослідження, рівня надійності та можливості узагальнення результатів дослідження для окремих поліграфологічних методик, які використовуються в польовій практиці. Усі попередні огляди точності поліграфологічного тестування є незадовільними щодо їх спроможності відповісти на поточне запитання щодо валідності, критеріальної точності та надійності поліграфологічних методик, які використовуються сьогодні.

По-перше, попередні огляди не розглядають точність тесту з адекватним описом процедурної комбінації послідовності тестових запитань і методу аналізу тестових даних (TDA), застосованого в дослідженні, обидва з яких, як вважається, мають важливий вплив на ефективність тесту. По-друге, це пов'язане з першою проблемою, і полягає в тому, що жоден із попередніх оглядів не робив зусиль, щоб виключити поліграфологічні методики, які більше не викладаються в межах акредитованих навчальних програм або не застосовуються в практиці. У результаті низка попередніх оглядів практично не використовується поліграфологами в практиці, дослідниками в поліграфології, адміністраторами програм, керівниками та споживачами, що стосується переваг різних поліграфологічних методик.

### **Теперішні цілі**

Основний інтерес для цього огляду становлять ті поліграфологічні методики, для яких існують докази на підтримку критеріальної валідності на рівнях, що вста-

---

<sup>6</sup> Блекстоун (Blackstone, 2011) також плутає різницю між діагностичним і скринінговим поліграфом; спочатку використовуючи менш поширені терміни «криміналістика» та «корисність» замість більш широко зрозумілих термінів «діагностика» та «скринінг», а потім намагаючись зобразити однотемні скринінгові дослідження як діагностичні дослідження. Далі Блекстоун стверджує, що багатоаспектні дослідження є скринінговими, а пізніше багатоаспектні дослідження відрізняються від багатотемних тим, що багатотемні дослідження проводяться за відсутності відомої проблеми, що вказує на те, що багатоаспектні дослідження є типом діагностичного дослідження, проведеного у відповідь на відому проблему. **На практиці вважається, що дисперсія критерію RQ як багатотемного, так і багатоаспектного поліграфологічного дослідження є незалежною, і обидва типи досліджень інтерпретуються за допомогою правил прийняття рішень, які відображають це припущення.**



новлюють стандарти практики АПА, які з 1 січня 2012 року вимагають використувати перевірені методики. Ці вимоги констатують, що діагностичні дослідження щодо конкретної події, які проводяться з доказовою метою і результати яких можуть бути використані як докази в судовому розгляді, повинні проводитися з використанням методик, які забезпечують рівень критеріальної точності .900 або вище, не враховуючи невизначені показники або з невизначеним показником на рівні .200 або нижче.

Діагностичні дослідження, проведені з використанням протоколу парного тестування, можуть досягти дуже високого рівня точності завдяки поєднанню результатів досліджень, проведених за допомогою методик, які забезпечують середній рівень критеріальної точності .860 або вище, не враховуючи невизначені результати, і з невизначеними показниками .200 або нижче. Дослідження, що проводяться з дослідницькою метою, повинні проводитися з використанням методик, які дають середній рівень критеріальної точності .800 або вище, не враховуючи невизначені показники або з невизначеним показником на рівні .200 або нижче<sup>7</sup>. Валідні методики також повинні бути підтверджені опублікованими та повтореними (відтвореними) науковими дослідженнями (реплікаціями). Щоб бути узагальненими, наукові дослідження повинні базуватися на вибірках, репрезентативних для загальної популяції.

На додаток до визначення вимог щодо критеріальної точності поліграфологічних методик, АПА прийняла стандарт, який визначає, що валідна методика складається з комбінації послідовності тестових запитань або формату, який відповідає валідним поліграфологічним принципам тестування, у поєднанні з валідним методом аналізу тестових даних. Поєднання цих двох основних компонентів є визнанням того, що валідний тест повинен спочатку отримати відповідну кількість інтерпретованої та значущої (тобто діагностичної) інформації, після чого інформація має бути ефективно інтерпретована. Нехтування будь-чим із цього призведе до незадовільної продуктивності тесту. Крім того, оскільки результати єдиного неповтореного наукового дослідження вважаються невизначеними в сфері науки, валідні методики повинні підтверджуватися щонайменше двома публікаціями.

Вибір методик, що ґрунтуються на доказах, мінімізує вплив, який може виникнути в результаті використання нестандартизованих, не валідних (не перевірених), субоптимальних або експериментальних методик. Хоча ця спроба була зроблена для того, щоб надати корисну інформацію з цього приводу, нагадуємо читачам, що звіт є оглядом літературних публікацій, доступних на момент підготовки звіту. Нові прилади, нові дослідження щодо валідності та нові методи аналізу стануть доступними в майбутньому. У світлі постійного прогресу в цій галузі висновки в цьому звіті слід вважати довідковими, а не політикою АПА.

Етика проведення тестів не розглядалася в цьому мета-аналізі. Дискусії про точність поліграфологічних методик іноді можуть переходити до дискусій про етику, пов'язану з процедурами проведення тестів у форматах CQT із ймовірним обманом,

<sup>7</sup> Незабаром після завершення цього звіту рада директорів АПА запропонувала змінити Стандарти практики, що стосуються скринінгових методик, через брак доступних досліджень у цій галузі, незважаючи на важливість цього застосування для правоохоронних органів і національної безпеки. Пропозиція дозволить використовувати методики скринінгу, якщо дослідження вкаже на точність, значно більшу за випадковість, і рекомендує використовувати підхід послідовних бар'єрів для мінімізації помилок. Оскільки пропозиція не була відголосована до завершення цього звіту, додатковий аналіз скринінгових методик з використанням запропонованих стандартів тут не включено.

для яких вважається необхідним психологічно маніпулювати досліджуванним, щоб досягти задовільного рівня специфічності тесту до правдивості та обмежити хибно-позитивні помилки до мінімального рівня. Це обговорення також може призвести до непродуктивних і навіть уникаючих відхилень щодо збільшення додаткової валідності (тобто «тестової корисності») рішень, які приймаються професійними користувачами поліграфологічних методик<sup>8</sup>.

Обговорення точності поліграфологічного тесту може також охоплювати етичні складнощі проведення тесту, для якого цілі поліграфолога та досліджуваного можуть бути різними. Наприклад: бажання досліджуваного полягає в тому, щоб отримати дані та результати тестування, щоб звільнити себе від подальшої перевірки, тоді як метою поліграфолога часто є психологічне сприяння зізнанню провини або розкриттю інформації від оманливих досліджуваних. У своєму обмеженому та ненауковому застосуванні поліграф може стати не більше, ніж опорою для підвищення ефективності інтерв'ю чи допиту, причому мало або зовсім не переймаючись за результат тесту. Важливо, повідомляючи про критеріальну точність поліграфологічного дослідження, зосередити увагу лише на тих питаннях, які стосуються рівня точності поліграфологічних методик, а не просто рівня зізнань. Іншими словами, наскільки ефективні сучасні поліграфологічні методики для правильної класифікації досліджуваних як оманливих або правдивих?

Точність тесту в науковому розумінні означає кілька речей і може передавати інформацію про багато важливих проблем. Основоположним серед цих проблем є питання конструктивної валідності, яке, у найбільш спрощеному вигляді відноситься до правильності базових побудов, принципів або ідей, на яких будується тест. Простіше кажучи, конструктивна валідність означає, чи виконує поліграфологічний тест те, для чого він призначений. На практиці поліграфологічні тести часто називають тестами на виявлення обману, отже, широке формулювання питання конструктивної валідності включатиме, чи поліграфологічний тест дійсно перевіряє чи вимірює обман. Обман аморфний і тому не може бути вимірний сам по собі. Обман є тимчасовим актом, і поліграфологічні тести, як і багато інших наукових тестів, оцінюються чисельно шляхом вимірювання або спостереження за реакцією досліджуваних на тестові стимули<sup>9</sup>. Оманливість або правдивість визначаються статистично, шляхом спостереження або вимірювання реакцій на кілька ітерацій тестових стимулів, агрегування реакцій, а потім використовуються структуровані правила прийняття рішень для інтерпретації результату шляхом порівняння з нормативними даними.

Конструктивна валідність також може стосуватися правильності припущень щодо функції структурних компонентів поліграфологічного тесту: чи функціонують запитання належним чином. Кілька досліджень досліджували конструктивну ва-

---

<sup>8</sup> Термін «додаткова (інкрементальна) валідність» є кращим перед старішим терміном «корисність», оскільки він передбачає очікування емпіричних доказів підвищеної точності рішень або ефективності рішень з боку користувачів поліграфологічними результатами тестування, в результаті інформації, отриманої з поліграфа, а не просте припущення, що вся інформація виявиться корисною або потрібною.

<sup>9</sup> Обман під час поліграфологічних тестів виводиться емпіричним шляхом, оскільки наукові дослідження в поліграфології обґрунтували, що оманливість і правдивість можуть бути визначені за допомогою CQT з показниками, які значно перевищують випадковість як функція диференціальної величини реакції на релевантні стимули та стимули порівняння. Вважається, що відмінності у величині реакції залежать від значимості стимулу. Очікується, що люди, які обманюють щодо релевантних стимулів, зазвичай демонструватимуть більші реакції на релевантні стимули, ніж стимули порівняння, тоді як люди, які правдиві щодо релевантних стимулів, як очікується, демонструватимуть більші реакції на порівняльні стимули.

лідність різних типів тестових запитань. Наприклад: виявилось, що запитання про загальну правдивість не функціонують належним чином, Хілліард (Hilliard 1979) і Абрамс (Abrams 1984) обидва надають розуміння загальних ускладнень і проблем, пов'язаних із запитаннями щодо намірів. Подібним чином було показано, що симптоматичні запитання, призначені для перевірки або виправлення зовнішніх проблем, не функціонують належним чином або не відповідають вимогам (Honts, Amato & Gordon, 2004; Krapohl & Ryan, 2001), незважаючи на деякі слабкі докази обґрунтування в ранньому дослідженні (Capps, Knill & Evans, 1993). Технічні запитання, розроблені для перевірки або використання езотеричних феноменів, таких як комплекс провини, не функціонують ефективно (Podlesny, Raskin & Barland, 1976). Подібним чином було показано, що жертвні запитання щодо наміру досліджуваного дати правдиву відповідь щодо релевантних запитань не функціонують належним чином (Capps, 1991; Horvath, 1994).

Незважаючи на велику кількість гіпотез, наукові дослідження не змогли продемонструвати докази конструктивної валідності для низки технічних запитань, за винятком одного. CQ, як правило, здатні викликати більшу реакцію з боку правдивих осіб, ніж RQ. Хоча конструктивна валідність є важливою проблемою, цей мета-аналіз розглядав критеріальну валідність, здатність відрізнити обман від правди на практичному рівні, з наголосом на розпізнавання поліграфологічних методик, для яких існують опубліковані та відтворені докази (реплікації) на підтримку точності тесту. Питання, що стосуються конструктивної валідності, не розглядались.

Критеріальна точність методів CQT, що використовуються в поліграфологічних тестах, не може бути адекватно описана одним числовим значенням. Натомість критеріальна валідність для поліграфологічних CQT тестів є результатом взаємодії кількох вимірів, що викликають занепокоєння, включаючи правильні та хибні збіги для оманливих рішень, а також правильні та хибні збіги для правдивих рішень. У польовій практиці результати тесту інтерпретуються або виражаються в термінах щодо наявності або відсутності значущих реакцій, які вказують на обман. Категоріальні результати тестування для окремих випадків є або позитивними, або негативними<sup>10</sup>, а критеріальна валідність тесту оцінюється шляхом поділу результатів вибіркового випадку на істиннопозитивні, хибнопозитивні, істиннонегативні та хибнонегативні.

Спроби описати критеріальну точність ще більше ускладнюються невизначеними результатами, для яких результати вибірки поділяються на невизначені результати для оманливих і правдивих груп. На додаток до проблем вимірювання та опису оцінок точності поліграфологічного тесту, різні поліграфологічні методики досягають різних розмірних профілів критеріальної точності. Деякі методики можуть бути призначені для забезпечення високої чутливості тесту за рахунок інших параметрів точності тесту, тоді як інші методики можуть бути розроблені для забезпечення балансу чутливості тесту та його специфічності.

Це мета-аналітичне дослідження має на меті узагальнити поточний стан існуючих опублікованих наукових доказів критеріальної валідності поліграфологічних методик, а також надати вказівки щодо тих методик, які, як очікується, надійно

---

<sup>10</sup> Незважаючи на те, що в польовій практиці результати поліграфологічних тестів інтерпретуються на предмет наявності чи відсутності значущих ознак обману, результати наукових досліджень часто обговорюються ціннісно-нейтральною мовою. **Позитивні результати тесту вказують на наявність занепокоєння або проблеми, яка перевіряється. Негативні результати тесту означають відсутність занепокоєння чи проблеми.**

забезпечать критеріальну точність, що задовольняє вимогам АПА щодо точності доказових, парних та дослідницьких процедур. Таким чином, це дослідження обмежується тими методиками, для яких наявні докази підтверджують їх критеріальну валідність, і не включає поліграфологічні методики з неповтореними дослідженнями або методики, для яких є повторені докази на рівнях, які не задовольняють вимоги стандартів АПА<sup>11</sup>.

## Метод

Було проведено огляд літератури, щоб визначити опубліковані наукові дослідження, які надали придатну для використання інформацію щодо критеріальної точності ідентифікованих поліграфологічних методик. Результати неповторених наукових досліджень не є корисними для мета-аналітичних оглядів, тому їх не було включено. Як практичне рішення це означало, що принаймні два опублікованих наукових дослідження були потрібні для включення досліджуваної методики в мета-аналіз. Досліджувані методики були збережені в мета-аналізі, якщо опубліковані та повторені наукові дослідження були ідентифіковані для обґрунтування валідності методики, і якщо сукупні результати включених досліджень вказували на надійний і узагальнений рівень точності, що відповідає вимогам стандартів практики АПА для доказових, парних або дослідницьких тестувань. Однак це не вимагалось, щоб окремі дослідження забезпечували критеріальну точність на рівнях, визначених АПА.

Було визнано важливим максимально залучити зацікавлені сторони (тобто директорів шкіл і розробників поліграфологічних методик) до відбору досліджень і методик для мета-аналізу, оскільки передбачалося, що робота багатьох поліграфологів і тренерів в поліграфології може вплинути на результати та рекомендації цього мета-аналізу.

Щоб досягти цього, наприкінці березня 2011 року було складено довгий список усіх ідентифікованих поліграфологічних методик та розповсюджено серед усіх поліграфологічних шкіл, акредитованих АПА, за допомогою контактної інформації, зазначеної на вебсайті АПА ([www.polygraph.org](http://www.polygraph.org)). Директорам шкіл або їхнім представникам було запропоновано повідомити комітет про будь-які методики, які ще не були визначені, і надати цитати або копії досліджень, які можна було б переглянути на підтримку запропонованих методик<sup>12</sup>. На початку червня 2011 року короткий список було розповсюджено серед усіх акредитованих поліграфологічних шкіл АПА, знову ж таки з використанням контактної інформації на вебсайті АПА, включаючи як методики, так і цитати, для яких були визначені опубліковані та повторені дослідження. Також було надіслано список із описом методик, для яких підстави для публікації були невідповідними для включення в мета-аналіз. Директорів шкіл або їхніх представників знову запросили відповісти та повідомити комітету про будь-які

---

<sup>11</sup> Оскільки наукові дослідження проводяться у всіх галузях науки, а стандарти практики періодично переглядаються та вносяться необхідні зміни, нагадуємо читачам, що інші поліграфологічні методики можуть задовольнити поточним вимогам стандартів АПА. Поліграфологам, адміністраторам програм та спеціалістам із забезпечення якості рекомендується оцінювати цю інформацію, усвідомлюючи нові стандарти та інформацію.

<sup>12</sup> Одну методику, розроблену Лінн Марсі (Lynn Marsy), було запропоновано додати до списку для розгляду. Однак опублікованих досліджень щодо цієї методики знайти не вдалося. Принаймні один представник школи рекомендував додаткові дослідження для кількох поліграфологічних методик.

методики чи будь-які опубліковані дослідження, які слід розглянути для включення в мета-аналіз<sup>13</sup>.

### **Вибір дослідження**

Вимоги до відбору окремих наукових досліджень для мета-аналізу були як якісними, так і кількісними. Деякі наукові дослідження не були розроблені, щоб функціонувати як дослідження щодо критеріальної валідності, а були спрямовані на вивчення конкретних питань, таких як вплив контрзаходів на поліграфологічну точність. Наукові дослідження, призначені для вивчення причинно-наслідкового зв'язку та побудови запитань, можуть бути некорисними для відповіді на питання про критеріальну точність, однак дослідження були включені, якщо вони надавали достатньо інформації для розрахунку критеріальної точності вибірки наукового дослідження або нормальної контрольної групи, яка не була предметом експериментальних маніпуляцій за межами правдивості чи обману. Президент АПА та голова комітету повідомили членам комітету, що слід розглянути питання щодо включення в мета-аналіз додаткових досліджень, якщо буде достатньо часу для завершення процесу перегляду та публікації до завершення мета-аналізу. Кілька наукових досліджень згодом було завершено та подано для рецензування та публікації. Результати цих наукових досліджень були включені після того, як вони були прийняті до публікації, або ще не були надруковані. Усі ці дослідження були розроблені як наукові дослідження критеріальної точності, які відповідають вимогам мета-аналізу.

**Вимоги до якісного відбору.** Вимоги до якісного відбору та включення до огляду полягали в тому, що відібрані наукові дослідження повинні бути опубліковані в журналі *Polygraph* або інших рецензованих наукових виданнях<sup>14</sup>. Дослідження також розглядалися для відбору, якщо воно було опубліковане акредитованою установою, яка присуджує вчені ступені, акредитоване акредитаційним агентством, визнаним Міністерством освіти Сполучених Штатів (США) або його іноземним еквівалентом. Для відбору також розглядалися наукові публікації наукових досліджень, фінансованих державними установами, які пройшли зовнішнє експертне оцінювання. Були також включені відредаговані підручники та їхні розділи. Наукові дослідження, які не підлягали редакційній чи зовнішній рецензії, а також наукові дослідження, описані лише в самовиданих книгах, не розглядалися. Відбір у мета-аналізі вимагав, щоб наукові дослідження проводилися таким чином, щоб забезпечити впевнене узагальнення результатів для польових умов. Ці вимоги включали, що дослідження повинні проводитися з використанням інструментів запису та компонентних давачів, які відображають польову практику (тобто з використанням двох давачів пневмографа,

<sup>13</sup> Одне дослідження було запропоновано включити в мета-аналіз того часу, Gordon et al. (2000) польове дослідження IZCT, розробником якого був президент АПА у 2010-2011 рр., який має майновий інтерес. Однак дослідження Гордона та ін. (Gordon et al. 2000) не включало жодної статистики надійності, жодних статистичних параметрів або опису вибірових розподілів оманливих і правдивих оцінок. Основний автор повідомив комітету (особисте повідомлення, 10 червня 2011), що він ніколи не бачив даних або справ, оскільки вони належать розвідувальній службі іноземного уряду. Було визначено, що це дослідження не можна включити через відсутність опублікованих даних про надійність і неможливість оцінити дані дослідження з даними інших досліджень. Виключення цього дослідження не завадило включенню IZCT у мета-аналіз.

<sup>14</sup> Журнал *Polygraph* започаткував експертне рецензування у 2003 році. Статті, опубліковані до того часу, підлягали лише редакційній рецензії. Оскільки *Polygraph* є важливим учбовим та історичним ресурсом, дослідження, опубліковані до 2003 року і без рецензування, були включені в цей мета-аналіз, якщо вони задовольняли всім іншим якісним і кількісним вимогам для відбору.

давача електродермальної активності і давача кардіографа з манжетою), а також поліграфологічну методику, для якої існує опублікований опис методики, включаючи правила та процедури для вибору цілей, формулювання запитань і тестової презентації тестових стимулів. Крім того, вимагалось, щоб метод аналізу тестових даних відповідав польовій практиці та підтримувався опублікованим описом його характеристик, перетворень (правил оцінювання), правил прийняття рішень і нормативних даних<sup>15</sup> або порогових значень. Результати досліджень, які використовували виключно автоматизовані алгоритмічні моделі аналізу тестових даних не були включені в мета-аналіз.

Основні критерії істинності повинні бути незалежними від рішення поліграфа<sup>16</sup>.

Оскільки на рішення первинного (*original*) поліграфолога могла вплинути екстраполіграфологічна інформація, перевагу надавали сліпим оцінкам, а не початковим рішенням поліграфолога<sup>17</sup>. Необхідна була базова демографічна інформація щодо поліграфологів, а також щодо досвіду та підготовки поліграфологів.

Головні дослідники, які проводять наукові дослідження, відібрані для включення, повинні були не враховувати статус критеріїв випадків та учасників дослідження<sup>18</sup>. Хоча нинішні стандарти АПА щодо наукових досліджень вимагають, щоб головний дослідник не брав участі в зборі даних, ця вимога не існувала до березня 2011 року, тому не використовувалася при виборі досліджень для мета-аналізу.

**Лабораторні та польові дослідження.** Польові дослідження є важливими для наукових дослідів у поліграфології, оскільки ці дослідження мають перевагу відомої екологічної валідності, і тому передбачається, що в цьому відношенні вони мають підвищену можливість узагальнення. Однак репрезентативність і можливість узагальнення польових досліджень певною мірою скомпрометовані не випадковим процесом відбору випадків, який залежить від наявності підтверджуючих даних. У той час як польові дослідження є дуже корисними для вивчення кореляцій, їх корисність підривається через неможливість контролювати достатню кількість змінних для визначення причинно-наслідкового зв'язку та конструктивної валідності.

Лабораторні дослідження також важливі для дослідів щодо поліграфа, оскільки ці дослідження легше проводити за допомогою випадкових методів, які зменшують упередженість дослідів та вибірки, а отже, підвищують можливість узагальнення отриманої інформації. Лабораторні дослідження також забезпечують більшу можливість контролювати ширший спектр змінних і важливі для вивчення причинності та

<sup>15</sup> Більшість поліграфологічних методів аналізу тестових даних не використовують нормативні дані, а використовують традиційні порогові значення, які були визначені гіпотетично або довільно. Незважаючи на те, що традиційні порогові значення підтверджено як ефективні, відсутність нормативних даних означає, що рівень статистичної значущості для традиційних порогових значень залишається невідомим, і ці порогові значення можуть бути неоптимальними.

<sup>16</sup> Підтвердження, засноване лише на зізнанні, виключало б невизначені випадки та випадки помилок і призводило б до завищення точності розрахунків. Судові результати як критерій також не є незалежними, якщо докази поліграфа розглядалися під час судового розгляду, і можуть призвести до завищених оцінок точності. Одне включене дослідження (Mangan, Armitage & Adams, 2008) не відповідало цій вимозі та базувалося лише на вибіркових випадках, підтверджених зізнанням. Не дивно, що результати дослідження показали 100% точність. В статті (Verschuere, Meijer, & Merckelbach, 2008) авторами аргументовано результати цього дослідження методологічним артефактом і тому є ненадійним.

<sup>17</sup> З практичної точки зору включення екстраполіграфологічної (позаполіграфологічної) інформації може бути вигідним, якщо це підвищить точність польових поліграфологів. У цьому аналізі відповіді на питання критеріальної валідності стосуються лише того, чи містять дані поліграфологічного тесту інформацію, яка може бути оцінена та інтерпретована як висновок щодо точності.

<sup>18</sup> Про одне дослідження (Gordon et al., 2005) було повідомлено як про несліпе в іншому звіті (Mohamed et al., 2006), в якому описане таке ж порівняння результатів fMRI з результатами поліграфологічного дослідження.

конструктивної валідності. Однак вважається, що можливість узагальнення лабораторних досліджень зменшується через те, що ці дослідження іноді проводяться за обставин, які неповно відповідають екологічним умовам польових досліджень.

З метою перегляду поточного стану валідності існуючих поліграфологічних методик, спеціальний комітет вирішив однаково розглядати польові та лабораторні дослідження, якщо вони задовольняють якісним та кількісним вимогам для відбору до мета-аналізу. Відмінності між критеріальною точністю польових і лабораторних досліджень історично були статистично незначущими, і було б нерозумно намагатися припустити або висунути гіпотезу щодо значення будь-яких відмінностей, які спостерігаються в одному порівнянні. Дослідження та огляди показали високий рівень узгодженості між польовими та лабораторними дослідженнями, і остаточна причина будь-яких відмінностей повинна бути визначена шляхом вивчення даних.

Андерсон, Лінсдей та Бушмен (Anderson, Lindsay, & Bushman, 1999) нещодавно дослідили широкий спектр лабораторних психологічних досліджень щодо агресивної поведінки та дійшли наступного висновку: «відповідність між лабораторними та польовими розмірами ефекту концептуально подібних незалежних і залежних змінних була значною. Коротше кажучи, психологічна лабораторія, як правило, створює істину, а не банальності» (стор. 3).

У сфері дослідів, які безпосередньо стосуються поліграфа, NRC (2003) не виявив істотних відмінностей між результатами лабораторних і польових досліджень. Подібним чином Поллін та ін. (Pollina et al. 2004) не виявили істотних відмінностей у точності класифікації польових і лабораторних дослідів щодо поліграфа.

**Вимоги до кількісного відбору.** Кількісні вимоги для включення в мета-аналіз полягали в тому, що дослідження надають достатньо інформації для розрахунку надійності та критеріальної валідності використаної методики. Щоб розрахувати повний розмірний профіль критеріальної точності, звітні або доступні дані повинні, як мінімум, включати розмір вибірки для правдивих і оманливих груп, а також правильні рішення, невизначені показники та показники помилок для правдивих і оманливих випадків. З кількома головними дослідниками зв'язалися для отримання додаткової інформації щодо розподілу вибірки. У кількісному відношенні дослідження були виключені лише через відсутність достатньої інформації для розрахунку надійності та можливості узагальнення результатів дослідження та розмірного профілю критеріальної точності.

**Надійність.** Кілька різних статистичних показників використовувалися для опису надійності поліграфологічних методик, включаючи кореляцію моментів продукту Pearson числових балів учасників дослідження, статистику Cohen's Кappa, що описує шанс-скоригований (*chance-corrected*) рівень згоди між двома учасниками дослідження, статистику Fleiss' Кappa, що описує рівень згоди між трьома або більше учасниками та нескореговану частку згоди між учасниками дослідження. Жоден із цих показників надійності не отримав переваги над іншими, і дослідження не були виключені через відсутність даних про надійність, якщо вони включали будь-який статистичний опис інтерратерної надійності (*interrater reliability*) числових оцінок або рішень<sup>19</sup>.

---

<sup>19</sup> Одна включена методика не відповідала цій вимозі. Жодне з опублікованих досліджень щодо IZCT не містило жодних статистичних доказів надійності між оцінювачами.

**Вибіркові розподіли.** Параметри середнього значення та стандартного відхилення були потрібні або, принаймні, могли бути розраховані на основі наявних даних для оманливого та правдивого розподілу балів для всіх вибраних наукових досліджень. Очікувалося, що множинні зразки, відібрані з тієї самої основної популяції та оцінені за допомогою того самого методу аналізу тестових даних, нададуть вибіркові розподіли, які суттєво не відрізнятимуться. Очікувалося також, що реплікація та агрегація результатів вибіркового розподілу дасть результати, які є більш репрезентативними та придатними для узагальнення, ніж будь-який один вибірковий розподіл, отже, це одна з причин необхідності проведення принаймні двох наукових досліджень методики.

Деякі наукові дослідження були опубліковані з неповним описом вибіркового розподілу. Таким чином, було необхідно отримати попередні оцінки від деяких головних дослідників, щоб обчислити цю відсутню статистику<sup>20</sup>. Більшість даних все ще була доступна, і дослідники були готові надати додаткову інформацію за запитом<sup>21</sup>. Очікувалося, що вибіркові розподіли для кожної поліграфологічної методики не відрізнялися б на статистично значущих рівнях, якби вибірки були отримані від досліджуваних, які були репрезентативними для тієї самої основної популяції, поліграфологічна методика застосовувалась подібним чином для кожного дослідження, а дані оцінювалися та інтерпретувалися за подібним застосуванням правил і процедур аналізу тестових даних.

Відсутність суттєвих відмінностей у розподілах вибірки буде інтерпретуватися як показник того, що розподіли вибірки є репрезентативними один для одного. Це підвищить нашу впевненість щодо надійності, з якою вибірки є репрезентативними та придатними для узагальнення для інших популяцій тестування. Значні відмінності будуть інтерпретуватися як ознаки вибірок, відібраних з різних популяцій, або як відмінності в проведенні поліграфологічного тестування або застосуванні правил і процедур аналізу даних тестування. За цих обставин впевненість у надійності та відтворюваності буде знижена.

### **Критеріальна точність**

Інформація про дослідження щодо критеріальної валідності вважалася достатньою для включення в мета-аналіз, якщо дослідження містило достатньо інформації для розрахунку повного розмірного профілю критеріальної точності, включаючи: чутливість (**sensitivity**) і специфічність (**specificity**) тесту (тобто точність для оманливих і правдивих груп, за винятком невизначених результатів), невизначені показники (**inconclusive rates**) для оманливих і правдивих груп, позитивне прогностичне значення (**positive predictive value, PPV**) (тобто частка істинно позитивних результатів до всіх позитивних результатів) і негативне прогностичне значення (**negative predictive value, NPV**) (тобто частка істинно негативних результатів до всіх нега-

---

<sup>20</sup> Статистичні описи розподілу вибірки тепер зазвичай потрібні для публікації в наукових журналах, включаючи журнал *Polygraph*. Редактори та рецензенти наукових публікацій не завжди вимагали цю інформацію в минулому, оскільки важливість майбутніх мета-аналітичних досліджень не завжди передбачалася.

<sup>21</sup> Не вдалося отримати статистичні параметри або необроблені бали для трьох наукових досліджень щодо двох методик, які проводилися Міністерством оборони США (US Department of Defense). Хоча члени комітету знали, що наукові дослідження були піддані ретельному та адекватному перегляду вченими з Міністерства оборони, відсутність даних створює незручності. Статистику надійності було надано у звітах про дослідження Міністерства оборони, і було вирішено зберегти ці дослідження в мета-аналізі. Пізніше ці наукові дослідження були відтворені незалежно.



тивних результатів), а також частка правильних рішень за винятком невизначених результатів серед оманливих і правдивих випадків, позначених як незважена точність (***unweighted accuracy***). Не було необхідності, щоб дослідження містило всі ці розмірні описи, і вони були включені, якщо комітет міг обчислити ці статистичні дані з наявної інформації. **Повний розмірний профіль може бути розрахований як мінімум з п'яти значень: точність прийняття рішень для правдивих і оманливих груп, з або без невизначених результатів, а також невизначені показники для оманливих і правдивих груп і кількість учасників дослідження, віднесених до кожної групи.**

Зведення цих важливих параметрів критерію до одного числа не можна здійснити, не знехтувавши значною частиною важливої інформації. Інша складність полягала в тому, що деякі показники точності не пов'язані з різницею в базових ставках або апіорних ймовірностях. Наприклад: PPV і NPV вразливі до відмінностей у базових ставках і непереконливих показниках, тому ці параметри критеріїв менш корисні для порівняння точності різних методик, для яких дослідження можуть проводитися з використанням вибірок із різними базовими ставками. Було встановлено, що незважене середнє значення правильних рішень і незважене середнє значення невизначених показників для оманливих і правдивих випадків дають найбільш зручну та узагальнену інформацію про критерії під час порівняння досліджень із потенційно різними базовими ставками.

### **Змінні величини**

Результати включених наукових досліджень і поліграфологічних методик були закодовані та згруповані для відповідності стандартам АПА 2012 для доказового, парного та дослідницького тестування. Поліграфологічні методики також були закодовані щодо того, чи призначені вони для інтерпретації з припущенням незалежності або відсутності незалежності критеріїв дисперсії серед стимулів RQ. Жодних інших модераторів чи медіаторів для цього дослідження не було закодовано.

### **Аналіз даних**

Усі вибірки розглядалися як упереджені та недосконалі представлення популяцій, з яких вони взяті. Це означало, що очікувалося б спостерігати деякі відмінності між параметрами розподілу вибірки та параметрами розподілу популяції, якби було можливо отримати дані з усієї популяції. Проте можна очікувати, що репрезентативні вибірки будуть відрізнятися від загальної популяції способами, які не є статистично значущими. Також очікується, що зразки з різних досліджень, якщо всі є репрезентативними для популяції, відрізнятимуться статистично не значущими способами.

Багатофакторні ANOVAs використовувались для порівняння параметрів розподілу вибірки кожного наукового дослідження з розподілом вибірки інших повторюваних наукових досліджень для кожної поліграфологічної методики. Методи Монте Карло (Monte Carlo) використовувались для розрахунку стандартних помилок і статистичних довірчих інтервалів для профілів критеріальної точності кожного з наукових досліджень, включених до мета-аналізу. Дані були агреговані для тих методик, які призначені для інтерпретації з припущенням незалежності серед стимулів RQ, і тих, які інтерпретуються з припущенням відсутності незалежності (*non-independence*.)

Результати мета-аналізу не були оцінені або зважені за якістю наукового дослідження, окрім кількісних критеріїв відбору, які були раніше описані для включення в мета-аналіз. Іншими словами, усі наукові дослідження розглядалися однаково, якщо вони задовольняли якісним вимогам для включення в мета-аналіз. Результати дослідження були зважені за розміром вибірки та кількістю оцінювачів, які брали участь, оскільки наукові дослідження, які базувалися на більших вибірках, і наукові дослідження, які включали більшу кількість оцінювачів, отримували пропорційно більшу вагу в мета-аналізі. Оскільки деякі вибірки були зібрані з кількома оцінювачами, які оцінювали лише підмножину всієї вибірки, значення ваги еквівалентні кількості оцінених результатів для кожного наукового дослідження та кожної поліграфологічної методики.

Дослідницькі питання, які стосуються цього мета-аналізу, такі: 1) по відношенню до яких поліграфологічних методик опублікували та відтворили докази валідності, що задовольняють вимогам стандартів практики АПА 2012 року щодо точності рішень і невизначених показників, 2) яка загальна точність валідних поліграфологічних методик, інтерпретованих з припущенням незалежності стимулів RQ, 3) який рівень точності поліграфологічних методик, інтерпретованих з припущенням відсутності незалежності (*non-independence*) стимулів RQ, 4) чи існують значні відмінності або викиди серед будь-яких валідних поліграфологічних методик та 5) чи є результати-викиди, які не враховуються наявними на даний момент доказами.

## **Результати**

Альфа була встановлена на .05 для всіх статистичних порівнянь.

### **Валідні поліграфологічні методики**

Тридцять вісім наукових досліджень відповідали критеріям для включення в мета-аналіз.

Ці наукові дослідження ґрунтувались на 32 різних вибірках підтверджених випадків, з яких було отримано 45 різних вибірок оцінок. У цих дослідженнях брали участь 295 оцінювачів, які надали 11 737 оцінок за результатами 3 723 поліграфологічних досліджень, включаючи 6 109 оцінок з 2 015 підтверджених оманливих поліграфологічних досліджень, 5 628 оцінок з 1 708 підтверджених правдивих поліграфологічних досліджень. Деякі вибірки використовувалися в різних наукових дослідженнях, деякі оцінювалися за допомогою різних моделей аналізу тестових даних, а деякі вибірки оцінювали більше ніж один оцінювач. У Додатку А наведено список включених наукових досліджень і розміри вибірки. Критеріальну точність, наведену для кожного з включених наукових досліджень, можна побачити в Додатку В. Статистичні дані щодо надійності включених наукових досліджень наведено в Додатку С. Додаток D показує середнє значення та стандартні відхилення для вибірових розподілів оманливих і правдивих оцінок для включених наукових досліджень. Чотирнадцять поліграфологічних методик було визначено як такі, що обґрунтовуються опублікованими та повтореними науковими дослідженнями, які відповідають якісним і кількісним вимогам відбору для цього мета-аналізу. Вони будуть представлені у алфавітному порядку:

## **AFMGQT / Семипозиційний аналіз тестових даних (Seven-position TDA)**

Модифікована методика загальних запитань ВПС США (Air Force Modified General Question Technique, AFMGQT)<sup>22</sup> є сучасним варіантом сімейства форматів CQT, які виникли як модифікації оригінальної методики загальних запитань (General Question Technique) (Reid, 1947) і методики порівняння зон (Zone Comparison Technique) (Backster, 1963). AFMGQT можна ефективно використовувати з двома, трьома або чотирма RQ. Дослідження AFMGQT використовуються як у багатоаспектних діагностичних контекстах, пов'язаних із конкретними подіями, так і в багатотемних контекстах скринінгу (наприклад, відбір працівників сфери громадської безпеки, скринінг щодо урядової безпеки, програми скринінгу після засудження тощо).

Дослідження з AFMGQT, проведені як у багатоаспектному, так і в багатотемному контексті, інтерпретуються з правилами прийняття рішень, заснованими на припущенні, що дисперсія критерію RQ є незалежною. Три наукових дослідження описують критеріальну точність AFMGQT при оцінюванні за семипозиційною моделлю аналізу тестових даних.

Сентер, Веллер та Крепол (Senter, Waller & Krapohl, 2008), використовуючи сліпі оцінки за семипозиційною системою оцінювання 33 запрограмованих оманливих і 36 запрограмованих правдивих досліджуваних, які були протестовані за допомогою AFMGQT після їх участі в імітаційному сценарії вибуху на узбіччі дороги, повідомили про незважену точність рішення .849 з невизначеним показником .015.

Нельсон та Хендлер (Nelson & Handler, у друці) використовували методи Монте Карло для вивчення критеріальної точності семипозиційного числового оцінювання поліграфологічних досліджень з AFMGQT з двома, трьома і чотирма RQ. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків. Незважена точність прийняття рішень була .814, разом із незваженим невизначеним показником .280.

Нельсон, Хендлер, Морган та О'Барк (Nelson, Handler, Morgan & O'Burke, у друці) отримали сліпі числові оцінки від трьох досвідчених поліграфологів, найнятих урядом Іраку, які використовували семипозиційну модель аналізу тестових даних для оцінки підтвердженого зразка тестів AFMGQT (N = 22), які були вибрані з підтвердженого архіву справ Міністерства оборони США (U.S. Department of Defense confirmed case archive). Одинадцять випадків підтверджено як оманливі; решта 11 підтверджено як правдиві. Загалом було отримано 66 оцінок, незважена точність рішень була .761, а незважений невизначений показник .242.

На рисунку 1 показано графік середнього значення та стандартного відхилення оцінювання за окремо взятим релевантним запитанням (subtotal scores)<sup>23</sup> для розподілів вибірки трьох наукових досліджень щодо AFMGQT, оцінених за семипо-

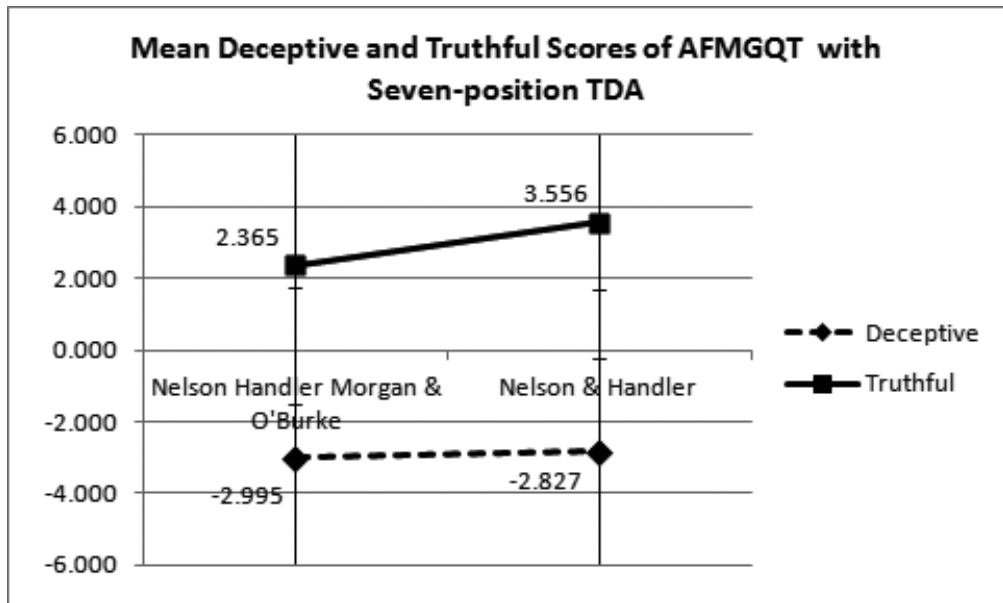
<sup>22</sup> Для AFMGQT існують дві версії: версія 1 і версія 2. Відмінності між цими двома методиками базуються на невивчених припущеннях щодо структури тесту, і ефект цих відмінностей не був повністю досліджений. Не існує переконливої гіпотези, яка б припускала, що продуктивність однієї версії буде іншою або кращою. Докази, наявні на даний момент, свідчать про те, що обидві версії працюють адекватно, і жодних істотних відмінностей виявлено не було. Тому запропоновано, що ці результати можна узагальнити для версій 1 і 2 AFMGQT, і обидві з них представлені у включених дослідженнях.

<sup>23</sup> У цьому аналізі використовувалось оцінювання за окремо взятим релевантним запитанням (Subtotal scores), оскільки AFMGQT оцінюється за допомогою правил прийняття рішень, використовуючи лише правило Subtotal, які припускають незалежність критерію серед RQ.

зиційною системою оцінювання. Для наукового дослідження щодо AFMGQT, яке завершили Сентер, Веллер та Крепол (Senter, Waller & Krapohl, 2008), середнє значення та стандартне відхилення не були доступні. Двофакторний дисперсійний аналіз ANOVA показав, що взаємодія між розподілом вибірки та статусом критерію не була значущою [ $F(1,68) = 0.263, (p = .610)$ ], і не була основним ефектом для розподілу вибірки [ $F(1,68) = 0.013, (p = .910)$ ].

Сукупний рівень точності рішень AFMGQT, оцінених за семипозиційним аналізом тестових даних, зважений за розміром вибірки та кількістю оцінювачів, становив .822 із сукупним невизначеним показником .191. Сентер, Веллер та Крепол (Senter, Waller & Krapohl, 2008) повідомили про надійність семипозиційного оцінювання AFMGQT як каппа-статистику .750. Частка сукупної згоди щодо прийняття рішення, за винятком невизначених результатів, для всіх наукових досліджень становила .965.

Рисунок 1. Середні значення для оманливих та правдивих оцінок по окремо взятому релевантному запитанню (subtotal) для AFMGQT, оціненої за семипозиційною системою оцінювання.



### AFMGQT / ESS

Три наукові дослідження описують критеріальну точність поліграфологічних досліджень з AFMGQT при оцінюванні за допомогою ESS.

Нельсон і Блелок (Nelson, Blalock, у друці) трансформували семипозиційні оцінки AFMGQT з лабораторного дослідження Сентера, Веллера та Крепола (Senter, Waller, Krapohl, 2008) на оцінки ESS, включаючи 33 результати для підтверджених оманливих випадків і 36 результатів для підтверджених правдивих випадків. Незважена точність рішення становила .839, з невизначеним показником .152.

Нельсон, Блелок і Хендлер (Nelson, Blalock, Handler, 2011) отримали сліпі оцінки ESS від двох недосвідчених поліграфологів і одного досвідченого поліграфолога,

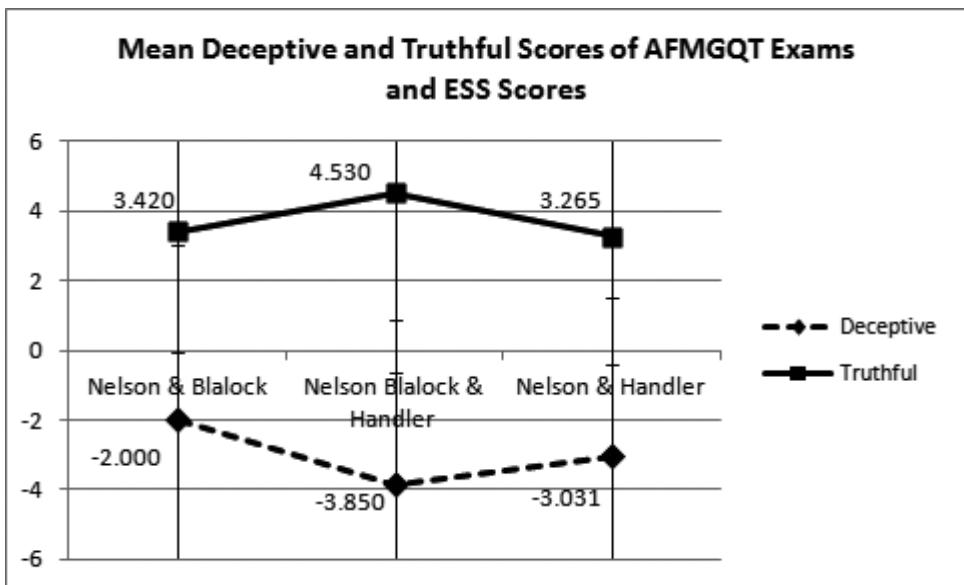
які використовували ESS для оцінки вибірки підтверджених поліграфологічних досліджень AFMGQT (N = 22), включаючи 11 поліграфологічних досліджень, які були підтверджені як оманливі, і 11 поліграфологічних досліджень, які були підтверджені як правдиві. Загалом було отримано 66 оцінок, а незважена точність рішення становила .883, з невизначеним показником .183.

Нельсон, Хендлер та Сентер (Nelson, Handler, Senter, у друці) використовували методи Монте Карло (Monte Carlo) для вивчення критеріальної точності оцінок ESS в поліграфологічних дослідженнях з AFMGQT з двома, трьома і чотирма RQ. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків. Незважена точність прийняття рішень була .876, з незваженим невизначеним показником .178.

На рисунку 2 показано графік середнього значення та стандартного відхилення балів розподілів вибірки трьох наукових досліджень щодо AFMGQT, оцінених з ESS. Двофакторний дисперсійний аналіз ANOVA показав, що взаємодія розподілу вибірки та критеріального статусу не була значущою [F (1,123) = 2.467, (p = .119)], а також не була основним ефектом для розподілу вибірки [F (1,123) = 0.009, (p = .925)].

Сукупний рівень точності прийняття рішення в цих наукових дослідженнях AFMGQT з ESS, зважений за розміром вибірки та кількістю оцінювачів, становив .875 із сукупним невизначеним показником .170. Надійність балів ESS поліграфологічних досліджень з AFMGQT, за даними Нельсона, Блелока та Хендлера (Nelson, Blalock & Handler, 2011) як бутстреп середнє значення коефіцієнтів парної кореляції (*bootstrap mean of pair-wise correlation coefficients*), становила .930.

Рисунок 2. Середні значення для оманливих та правдивих оцінок за окремо взятими релевантним запитанням (subtotal scores) для наукових досліджень щодо AFMGQT/ESS.



## **Backster You-Phase**

Методика Backster You-Phase є діагностичною методикою, яка базується на концепції порівняння зон Бакстера (Backster's Zone Comparison concept). Ця методика оцінюється за правилами аналізу тестових даних, розробленими Клівом Бакстером (Cleve Backster), та використовується майже виключно в Школі виявлення обману Бакстера (Backster School of Lie Detection, 2011). Як загальний ZCT (Department of Defense, 2006; Honts, Raskin & Kircher, 1987), так і сучасні варіанти ZCT (Gordon et al., 2000; Matte & Reuss, 1989) були розроблені на основі методики Backster You-Phase. Оцінки з двох недавніх наукових досліджень були зведені для розрахунку профілю критеріальної точності поліграфологічних досліджень з Backster You-Phase.

Нельсон (Nelson, у друці) використав початкові параметри (*seeding parameters*), розраховані на основі сукупних розподілів двох невідібраних наукових досліджень<sup>24</sup>, щоб створити простір Монте Карло (Monte Carlo) зі 100 змодельованих поліграфологічних досліджень з Backster You-Phase. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків. Незважена точність рішення для наукового дослідження Нельсона (Nelson, у друці) з Монте Карло (Monte Carlo) щодо You-Phase становила .927, з невизначеним показником .321.

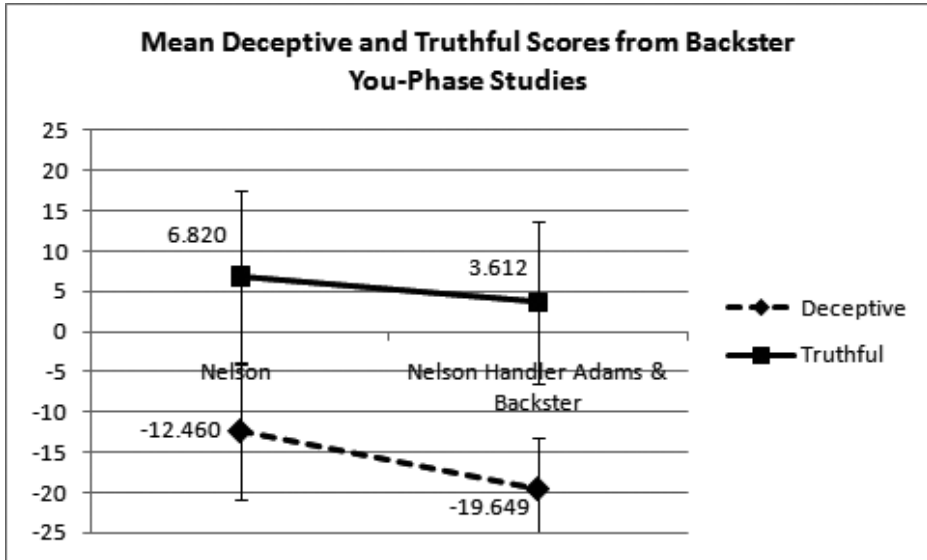
Нельсон, Хендлер, Адамс та Бакстер (Nelson, Handler, Adams & Backster, у друці) проаналізували результати семи поліграфологів, які надали 144 сліпих числових оцінок для вибірки (N = 22) з 11 підтверджених оманливих і 11 підтверджених правдивих поліграфологічних досліджень з You-Phase. Ці семеро оцінювачів мали різний досвід від менше одного року до понад 30 років. Результати сліпого оцінювання дали незважений показник точності рішень .825 з невизначеним показником .117. На рисунку 3 показано графік середнього значення та стандартного відхилення для балів оманливих і правдивих випадків. Двофакторний незбалансований<sup>25</sup> дисперсійний аналіз ANOVA показав, що взаємодія розподілу вибірки та статусу критерію не була значущою  $F(1,68) = 0.869$ , ( $p = .355$ ), а також не була основним ефектом для розподілу вибірки [ $F(1,68) = 0.164$ , ( $p = .686$ )].

Сукупні результати двох опублікованих наукових досліджень щодо Backster You-Phase, зважені за розміром вибірки та кількістю оцінювачів, дали показник точності рішення .863 і невизначений показник .196. Надійність числових балів поліграфологічних досліджень з Backster You-Phase, повідомлена Нельсоном та ін. (Nelson et al., у друці) як бутстреп середнє значення коефіцієнтів парної кореляції (*bootstrap mean of pair-wise correlation coefficients*), становила .567.

<sup>24</sup> Хонтс, Ходес та Раскін (Honts, Hodes & Raskin, 1985) використовували методику Backster You-Phase у науковому дослідженні щодо контрзаходів, для якого традиційну манжету було замінено альтернативним кардіодавачем. Мейрон, Крепол та Ешкеназі (Meiron, Krapohl & Ashkenazi, 2008) використовували методику Backster You-Phase у дослідженні правила «або-або» (Either-Or Rule), використовуючи ретельно відібрану вибірку, з якої не були включені результати проблемних поліграфологічних досліджень, у результаті чого вибірка вважається такою, що систематично позбавлена дисперсії помилок. Незважаючи на те, що жодне з цих наукових досліджень не можна було використовувати окремо, параметри, які описують зведені розподіли оманливих і правдивих оцінок, вважалися більш узагальненим представленням помилки або неконтрольованої дисперсії разом із діагностичною дисперсією для оцінок, отриманих з поліграфологічних досліджень з Backster You-Phase.

<sup>25</sup> Незбалансовані ANOVAs із використанням середнього гармонічного значення вибірки  $N_s$  використовувалися в цьому дослідженні, коли це було необхідно через відмінності в розмірах вибірки. Як наслідок, загальні ступені свободи в підсумку ANOVA можуть не відобразити суму всіх зразків так само, як збалансована конструкція ANOVA. Можна очікувати, що незбалансовані конструкції ANOVA забезпечать трохи меншу статистичну потужність, ніж збалансовані конструкції ANOVA.

Рисунок 3. Середнє значення та стандартне відхилення для числових оцінок поліграфологічних досліджень Backster You-Phase.



### Тест на приховану інформацію (Concealed Information Test, CIT)

CIT, також відомий як Тест на знання провини (Guilty Knowledge Test) (GKT; Lykken, 1959) і пов'язаний з методикою піку напруженості (Peak of Tension, POT) (Ansley, 1992), є діагностичною методикою, пов'язаною з певною подією, яка може бути використана для виявлення того, чи досліджуваний володіє знаннями або інформацією, які були б доступні лише слідчим та винному чи причетному підозрюваному. Як і CQT, CIT/GKT базується на принципі помітності (*principle of salience*), включаючи емоції, пізнання та поведінкові умови як психологічну основу фізіологічної реакції (Senter, Weatherman, Krapohl & Horvath, 2010). Крім того, CIT/GKT проводиться із застосуванням приладів, подібних як і для методик CQT, включаючи давач електродермальності активності, і може включати давачі кардіографа та пневмографа. Однак CIT/GKT не містить запитань порівняння і не є методом CQT. Таким чином, CIT/GKT не підлягає таким самим етичним занепокоєнням, як методи CQT з ймовірним обманом щодо маніпулювання досліджуваним як ознаки проведення тесту<sup>26</sup>. Гіпотетичні пояснення психофізіологічних механізмів, що лежать в основі CIT/GKT, не обмежуються лише емоціями та страхом як єдиною основою реакції. Крім того, CIT/GKT не піддавався науковій критиці щодо ролі зізнань досліджуваного як мети тесту та як форми перевірки результатів тесту<sup>27</sup>. МакЛарен (MacLaren, 2001) опублікував результати мета-аналізу 50 вибірок у 22 наукових дослідженнях щодо CIT/GKT. Тридцять дев'ять із цих вибірок включали 1070 досліджуваних, з яких 666 брали участь у поведінкових діях, щодо яких вони приховували інформацію, а

<sup>26</sup> Подібним чином доведено, що методи керованого обману працюють так само добре, як і методи ймовірного обману, і менш схильні до етичних ускладнень щодо маніпулювання суб'єктом тестування (Bell, Kircher & Bernhardt, 2008; Blalock, Nelson, Handler & Shaw, 2011; Honts & Reavy, 2009).

<sup>27</sup> Надмірна увага до підтвердження зізнання та залежний критерій призвели до критики щодо забруднення та переоцінки точності CQT.

також 404 досліджуваних, які не були залучені або не знали про поведінкові деталі. Одинадцять вибірок включали 177 досліджуваних, які володіли прихованими знаннями про досліджуваний поведінковий акт, але насправді не брали участі в ньому.

Використовуючи протокол оцінювання і методологію тестування, описану Ліккеном (Lykken, 1959), результати, опубліковані МакЛареном (MacLaren), дали рівень чутливості тесту .815 для досліджуваних, залучених до поведінкових дій, разом із рівнем специфічності тесту .832 для не залучених досліджуваних, які не мали прихованих знань. Незважена точність рішення становила .823. Однак, коли результати включали тих досліджуваних, які володіли прихованою інформацією, але не були задіяні в поведінковій діяльності, рівень чутливості тесту становив .759, а незважений коефіцієнт точності рішення становив .795.

### **Скринінгова методика керованого обману (Directed-lie Screening Test) / Семипозиційний аналіз тестових даних (Seven-position TDA)**

Скринінгова методика керованого обману (Directed-lie Screening Test, DLST) базується на методиці на шпигунство та саботаж (TES), розробленої Міністерством оборони США<sup>28</sup>. Як вказує його назва, методика DLST використовує CQ керованого обману та призначена для скринінгових досліджень, які проводяться за відсутності будь-якої відомої проблеми. Скринінгові тести часто будуються та інтерпретуються як багатотемні, для яких дисперсія критерію CQ вважається незалежною. Міністерство оборони США опублікувало два наукові дослідження, які включали DLST/TES і семипозиційний аналіз тестових даних.

Звіти про наукові дослідження були доступні, хоча їх було недостатньо для відповідності критеріям включення в дослідження через відсутність стандартних відхилень вибірки у звітах про наукові дослідження. Крім того, дані наукових досліджень Міністерства оборони були недоступні для ознайомлення комітетом. Однак нещодавно було завершено два наукові дослідження реплікації DLST/TES, і дані цих досліджень були включені в мета-аналіз.

Перше опубліковане наукове дослідження щодо DLST/TES (Research Division Staff, 1995a) було лабораторним експериментом із імітацією сценарію шпигунства. Троє досвідчених федеральних поліграфологів оцінили 94 поліграфологічних досліджень, із яких було 26 досліджуваних, запрограмованих обманувати, і 68 досліджуваних, запрограмованих говорити правду. Результати цього дослідження дали незважений рівень точності прийняття рішень .788, з невизначеним показником .155<sup>29</sup>.

У другому опублікованому науковому дослідженні щодо DLST/TES (Research Division Staff, 1995b) 10 досвідчених федеральних поліграфологів надали результати

<sup>28</sup> Назва «Скринінгова методика керованого обману» (Directed Lie Screening Test) використовується в контекстах, у яких дослідницькі цілі відрізняються від шпигунства та саботажу.

<sup>29</sup> Незважаючи на те, що вся інформація була включена в опубліковані звіти, деякі хибнопозитивні помилки та невизначені результати були виключені з попередньо наданої статистики для цього дослідження. Хибнопозитивні результати видалені з повідомлених результатів, коли досліджуваний надавав зізнання після тестування, які розглядалися б як істотні в польових умовах, через що результати не розглядалися як помилкові. Невизначені результати були видалені з попередньо повідомлених результатів через процедури DLST/TES, які вимагають негайного повторного розгляду невизначених результатів. Під час сліпого оцінювання оцінювачам було неможливо повторити дослідження з невизначеними результатами, і дослідники вирішили описати точність DLST/TES без невизначених результатів, які не могли бути предметом повторної перевірки. Усі хибнопозитивні та невизначені результати були включені в поточні результати, оскільки це вважалося більш консервативною оцінкою критеріальної точності для цієї методики.



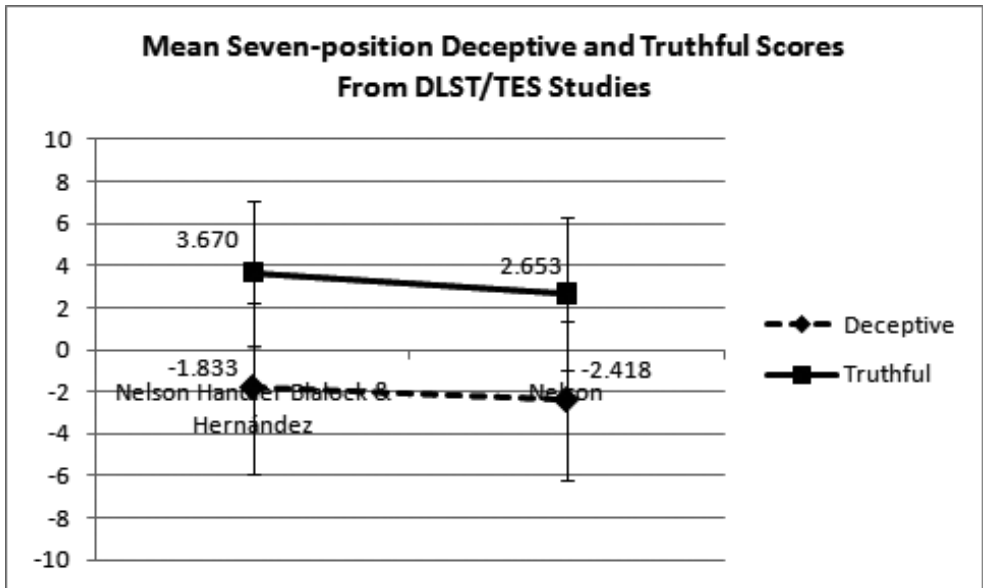
оцінювання для 30 оманливих і 55 правдивих лабораторних досліджень, які включали імітаційний сценарій шпигунства. Результати цього дослідження дали незважений рівень точності прийняття рішень .888, з невизначеним показником .009<sup>30</sup>.

Нельсон (Nelson, у друці) використовував методи Монте Карло для дослідження критеріальної точності DLST/TES і повідомив про незважений рівень точності рішення .874 з невизначеним показником .096. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків.

Нельсон, Хендлер, Блелок і Ернандес (Nelson, Handler, Blalock, Hernández, у друці) вивчали DLST/TES за імітаційним сценарієм шпигунства з поліграфологами-новачками з уряду Іраку. Двоє оцінювачів, у тому числі один досвідчений федеральний поліграфолог, який також є основним інструктором АПА, і один міжнародний поліграфолог, який є членом АПА з Мексики, провели сліпе семипозиційне оцінювання для 25 запрограмованих оманливих і 24 запрограмованих правдивих випадків. П'ятдесят оцінок було отримано для запрограмованих досліджень обману і 48 оцінок було отримано для запрограмованих досліджень правди. Незважений рівень точності рішення становив .831 з невизначеним показником .092.

На рисунку 4 показано графік середнього значення та стандартного відхилення для оманливих і правдивих оцінок за семипозиційною системою оцінювання поліграфологічних досліджень з DLST/TES. Для наукових досліджень TES, проведених Міністерством оборони США, не було доступних даних про середнє значення та стандартне відхилення. Двофакторний ANOVA продемонстрував, що взаємодія розподілу вибірки та статусу критерію не була значущою [ $F(1,128) = 0.109$ , ( $p = .742$ )], а також не була основним ефектом для розподілу вибірки [ $F(1,128) = 0.023$ , ( $p = .880$ )].

Рисунок 4. Середнє значення та стандартне відхилення для поліграфологічних досліджень з DLST/TES, оцінених за семипозиційною системою оцінювання.



<sup>30</sup> Дві хибнопозитивні помилки були видалені з раніше повідомлених оцінок точності через отримані зізнання після тестування, які, на думку головного дослідника, могли вважатися суттєвими, а не помилковими в польових умовах. Один невизначений результат було видалено з попередніх оцінок точності. Обчислення в цьому звіті включають усі помилки та невизначені результати.

Сукупні результати чотирьох опублікованих наукових досліджень щодо DLST/ TES із семипозиційним аналізом тестових даних, зважені за розміром вибірки та кількістю оцінювачів, показали коефіцієнт точності рішення .844 і невизначений показник .088. Надійність досліджень DLST/ TES, проведених Міністерством оборони США, розрахована за допомогою Cohen's Kappa, була .760. Середня частка узгодженості попарного рішення для всіх наукових досліджень щодо DLST/ TES, оцінених з семипозиційною системою аналізу тестових даних, становила .806.

### **Скринінгова методика керованого обману (Directed-lie Screening Test) / ESS**

Чотири наукові дослідження описують критеріальну точність DLST/ TES при оцінюванні за допомогою ESS.

Нельсон і Хендлер (Nelson & Handler, у друці) використовували методи Монте Карло для дослідження критеріальної точності поліграфологічних досліджень з DLST/ TES, оцінених за допомогою ESS. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків. Незважена точність рішення була .831 з невизначеним показником .104.

Нельсон, Хендлер, Морган (Nelson, Handler & Morgan, у друці) використали імітаційний сценарій шпигунства, щоб вивчити критеріальну точність поліграфологічних досліджень з DLST/ TES, оцінених з ESS, проведених сьома недосвідченими поліграфологами на восьми досвідчених поліграфологах, які були повністю обізнані з роботою та оцінюванням поліграфологічних досліджень, включаючи DLST/ TES. Сліпе оцінювання було проведено для 25 запрограмованих оманливих поліграфологічних досліджень і 24 запрограмованих правдивих поліграфологічних досліджень. Незважена точність рішення становила .854 з невизначеним показником .088.

Нельсон (Nelson, у друці), використовуючи інший метод Монте Карло, порівняв оцінки DLST/ TES ESS з оцінками семипозиційною та трипозиційною системами оцінювання. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків. Незважена точність рішення під час оцінювання з ESS була .871, з невизначеним показником .048.

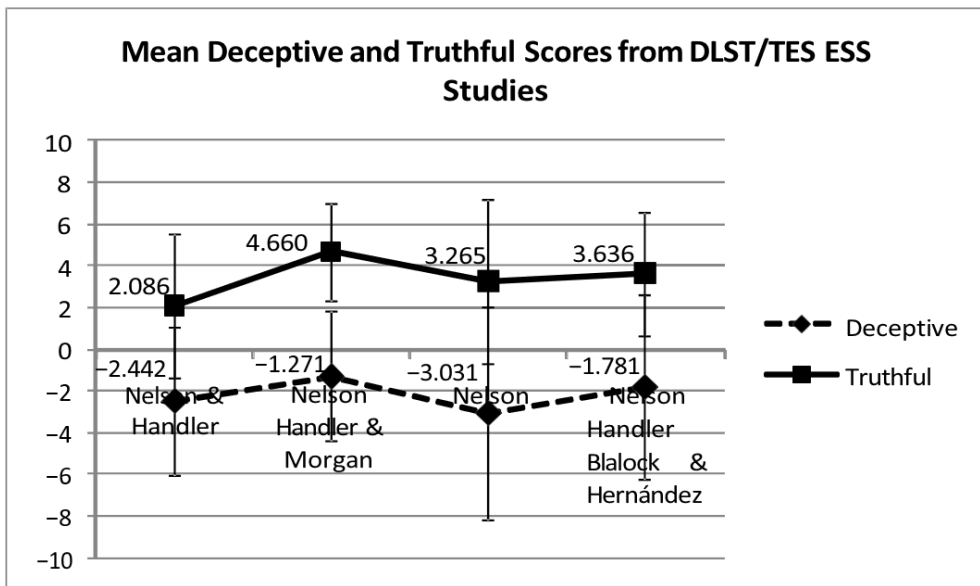
Нельсон, Хендлер, Блелок та Ернандес (Nelson, Handler, Blalock & Hernández, у друці) повідомили про результати сліпого оцінювання досліджень з DLST/ TES за ESS. Оцінювання за семипозиційною системою оцінювання від двох поліграфологів, у тому числі одного досвідченого федерального поліграфолога, який також є основним інструктором АПА, та одного міжнародного поліграфолога, який є членом АПА з Мексики, були переведені в бали ESS, включаючи 50 сліпих оцінок для 25 запрограмованих винних і 48 сліпих оцінок для 24 запрограмованих невинних досліджуваних. Незважена точність прийняття рішень становила .859, а незважений невизначений показник був 123.

На рисунку 5 показано графік середнього значення та стандартного відхилення результатів чотирьох наукових досліджень DLST/ TES, оцінених за ESS. Двофакторний ANOVA показав, що взаємодія між розподілом вибірки та статусом критерію не була значущою [ $F(1,289) = 2,396$ , ( $p = .123$ )], як і основний ефект для розподілу вибірки [ $F(3,289) = 0.156$ , ( $p = .925$ )].

Сукупний рівень точності рішень у цих дослідженнях, зважений за розміром вибірки та кількістю оцінювачів, склав .858 із сукупним невизначеним показником .090. Надійність оцінювання з ESS поліграфологічних досліджень з DLST/ TES,

представлених як бутстреп середнє значення пропорції попарної узгодженості щодо прийняття рішень (*bootstrap mean of the proportion of the pairwise decision agreement*) за винятком невизначених результатів, становила .911 для наукового дослідження Нельсона, Хендлера та Моргана (Nelson, Handler & Morgan, у друці) та .769 для наукового дослідження Нельсона, Хендлера, Блелока та Ернандеса (Nelson, Handler, Blalock & Hernández, у друці). Середня частка узгодженості попарного рішення для цих наукових досліджень .840.

Рисунок 5. Середнє значення та стандартне відхилення для оцінювання за ESS поліграфологічних досліджень DLST/TES.



### **Federal You-Phase / Семипозиційний аналіз тестових даних (Seven-position TDA)**

Методика Federal You-Phase<sup>31</sup> є діагностичною методикою для конкретної події, побудованою з двома релевантними запитаннями. Два дослідження описують критеріальну точність методики Federal You-Phase при оцінюванні з використанням семипозиційної моделі аналізу тестових даних.

Нельсон (Nelson, 2011) використовував методи Монте Карло для розрахунку критеріальної точності поліграфологічних досліджень за Federal You-Phase, оцінюючи за семипозиційною моделлю аналізу тестових даних. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків. Незважена точність прийняття рішень становила .870, а незважена точність невизначених показників становила .301.

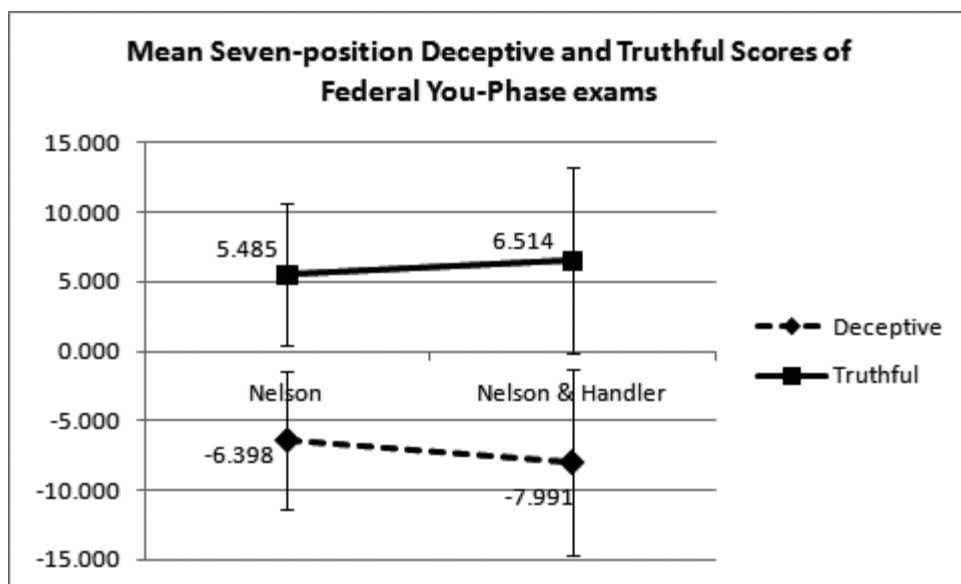
Нельсон, Хендлер, Блелок і Кушман (Nelson, Handler, Blalock & Cushman, у друці) отримали сліпі оцінки від восьми недосвідчених і двох досвідчених поліграфологів, які використовували семипозиційну модель аналізу тестових даних, щоб надати

<sup>31</sup> Іноді називають методикою Bi-Zone.

220 оцінок для вибірки з методикою Federal You-Phase (N = 22), обраних із архіву підтверджених справ Міністерства оборони США. Одинадцять випадків підтверджено як оманливі, а 11 – як правдиві. Незважена точність рішення становила .885, з незваженим невизначеним показником .108.

На рисунку 6 показано графік середнього значення та стандартного відхилення для наукових досліджень щодо Federal You-Phase, оцінених за семипозиційною системою оцінювання. Двофакторний ANOVA показав, що взаємодія розподілу вибірки та статусу критерію не була значущою [F (1,68) = 0.628, (p = .431)], також не була основним ефектом для розподілу вибірки [F (1,68) = 0.001, (p = .977)].

Рисунок 6. Середнє значення та стандартне відхилення для поліграфологічних досліджень з Federal You-Phase, оцінених за семипозиційною системою оцінювання.



Сукупний рівень точності рішень у наукових дослідженнях щодо поліграфологічних досліджень з Federal You-Phase, оцінених за допомогою семипозиційної системи оцінювання, зваженої за розміром вибірки та кількістю оцінювачів, становив .883 із сукупним невизначеним показником .168. Надійність семипозиційної системи оцінювання щодо Federal You-Phase, повідомлених як бутстреп середнє значення пропорції попарної узгодженості прийняття рішень (*bootstrap mean of the proportion of the pairwise decision agreement*), за винятком невизначених показників, становила .852.

### **Federal You-Phase / ESS**

Два дослідження описують критеріальну точність методики Federal You-Phase при оцінюванні за допомогою ESS.

Нельсон (Nelson, 2011) використовував методи Монте Карло для розрахунку критеріальної точності методики Federal You-Phase при оцінюванні за допомо-

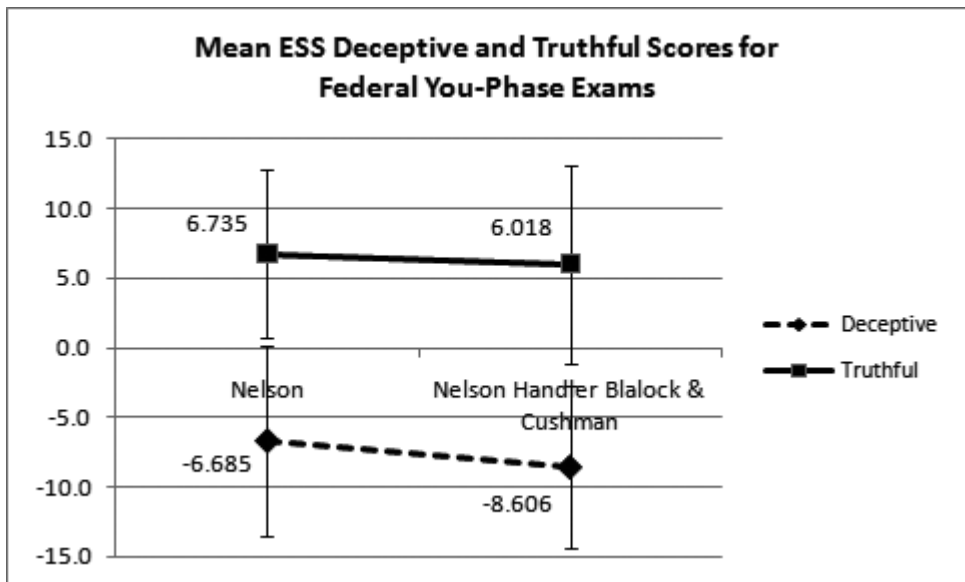
гою ESS. Простір Монте Карло складався з 50 критеріальних правдивих випадків і 50 критеріальних оманливих випадків. Незважена точність прийняття рішення становила .897, а незважений невизначений показник становив .096.

Нельсон, Хендлер, Блелок і Кушман (Nelson, Handler, Blalock & Cushman, у друці) повідомили про критеріальну точність поліграфологічних досліджень з Federal You-Phase, оцінених з ESS, які були отримані шляхом перетворення 220 семипозиційних оцінок, оцінених всліпу, отриманих від восьми недосвідчених і двох досвідчених поліграфологів, які оцінювали вибірку поліграфологічних досліджень з Federal You-Phase (N = 22), вибрані з підтвердженого архіву справ Міністерства оборони США. Одинадцять випадків підтверджено як оманливі, а 11 – як правдиві. Незважена точність прийняття рішень становила .906, а незважений невизначений показник був .235.

На рисунку 7 показано середні значення та стандартне відхилення для наукових досліджень щодо Federal You-Phase, оцінених за ESS. Двофакторний ANOVA показав, що взаємодія розподілу вибірки та статусу критерію не була значущою [F (1,68) = 0.155, (p = .695)], а також не була основним ефектом для розподілу вибірки [F (1,68) = 0.021, (p = .886)].

Сукупний рівень точності прийняття рішень у поліграфологічних досліджень з Federal You-Phase, зважений за розміром вибірки та кількістю оцінювачів, становив .904 із сукупним невизначеним показником .192, коли оцінювали за методом аналізу тестових даних ESS. Надійність, повідомлена як бутстреп середнє значення пропорції попарної узгодженості прийняття рішень (*bootstrap proportion of pairwise decision agreement*) без урахування невизначених результатів, становила .897.

Рисунок 7. Середнє значення оманливого та правдивого оцінювання для наукових досліджень щодо Federal You-Phase / ESS.



## **Federal ZCT / Семипозиційний аналіз тестових даних (Seven-position TDA)**

Federal ZCT – це діагностична методика щодо конкретної події (*event-specific diagnostic technique*), яка містить три RQ. Три наукові публікації описують критеріальну точність Federal ZCT, які оцінювались з використанням семипозиційної моделі аналізу тестових даних.

Блеквел (Blackwell, 1998) описав критеріальну точність 100 підтверджених поліграфологічних досліджень Federal ZCT, з яких 65 випадків було підтверджено як оманливі, а 35 як правдиві. Загалом було отримано 195 оцінок для критеріальних оманливих досліджень і 105 балів для критеріальних правдивих досліджень. Троє досвідчених поліграфологів, які пройшли федеральну підготовку, оцінювали всі випадки за допомогою семипозиційного методу аналізу тестових даних<sup>32</sup>. Незважена точність рішення становила .793, а незважений невизначений показник становив .159.

Крепол і Кушман (Krapohl & Cushman, 2006) повідомили про критеріальну точність поліграфологічних досліджень з Federal ZCT, оцінених групою з 10 досвідчених поліграфологів, кожен з яких оцінив 50 підтверджених оманливих польових поліграфологічних досліджень і 50 підтверджених правдивих поліграфологічних досліджень, вибраних із підтвердженого архіву справ Міністерства оборони США. Всього було отримано 1000 оцінок. Незважена точність прийняття рішень становила .852, а незважений невизначений показник становив .198

Хонтс, Амато та Гордон (Honts, Amato & Gordon, 2004) повідомили про критеріальну точність поліграфологічних досліджень з Federal ZCT, які оцінювали три оцінювача з використанням федеральної семипозиційної моделі аналізу тестових даних. Загалом було отримано 72 оцінки для 24 критеріальних оманливих поліграфологічних досліджень і 72 оцінки було отримано для 24 критеріальних правдивих поліграфологічних досліджень. Незважена точність рішення становила .958, а незважений невизначений показник становив .042.

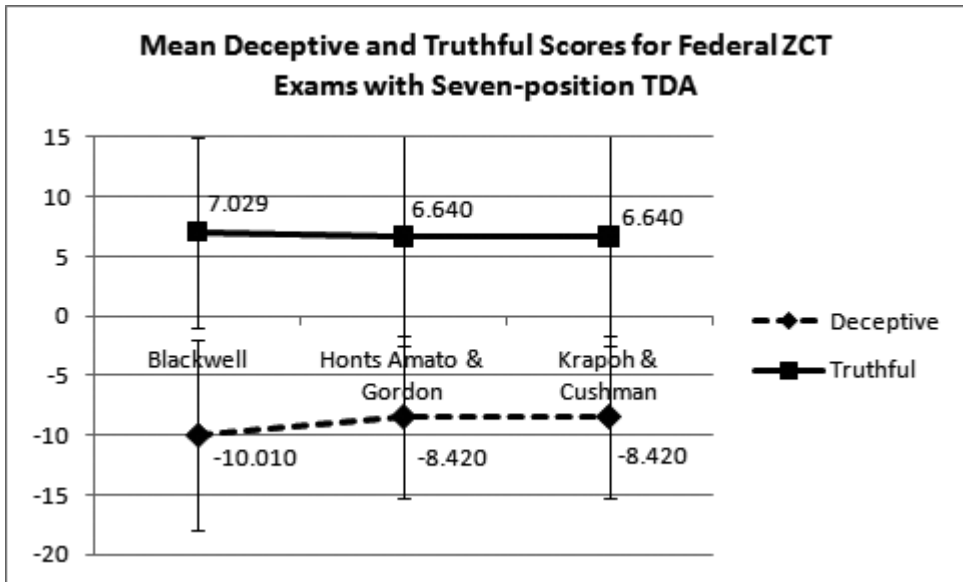
На рисунку 8 показано середнє значення та графік стандартного відхилення за семипозиційною системою оцінювання для наукових публікацій щодо Federal You-Phase. Двофакторний ANOVA показав, що взаємодія розподілу вибірки не була значущою, [ $F(1204) = 0.706$ , ( $p = .402$ )], а також не була основним ефектом для розподілу вибірки [ $F(1204) = 0.004$ , ( $p = .951$ )].

Сукупний рівень точності прийняття рішення в наукових публікаціях щодо семипозиційного аналізу тестових даних поліграфологічних досліджень з Federal ZCT, зважений за розміром вибірки та кількістю оцінювачів, склав .860 із сукупним невизначеним показником .171. Надійність семипозиційного аналізу тестових даних поліграфологічних досліджень з Federal ZCT, повідомлена як статистика каппа Флейсса (Fleiss' kappa), для категоріальних рішень кількох оцінювачів становила .570, а парна частка узгодженості рішень за винятком невизначених результатів становила .800.

---

32 Старіша федеральна модель аналізу тестових даних, створена до 2006 року, використовувала більше інформативних ознак, ніж існуюча федеральна модель аналізу тестових даних, заснована на фактах. Однак Кірчер та ін. (Kircher et al., 2005) повідомили, що досвідчені поліграфологи схильні порушувати правила, які не працюють, і наголошують на процедурах, які працюють. Тому можливо, що оцінки цих поліграфологів більш точно відображають поточну підготовку та польову практику, ніж можна було спочатку припустити.

Рисунок 8. Середні оманливі та правдиві оцінки для поліграфологічних досліджень з Federal ZCT, оцінених за семипозиційним аналізом тестових даних.



### ***Federal ZCT / семипозиційний аналіз тестових даних з доказовими правилами прийняття рішення (evidentiary decision rules)***

В двох наукових публікаціях описано критеріальну точність методики Federal You-Phase при оцінюванні за допомогою семипозиційної моделі аналізу тестових даних та доказових правил прийняття рішень.

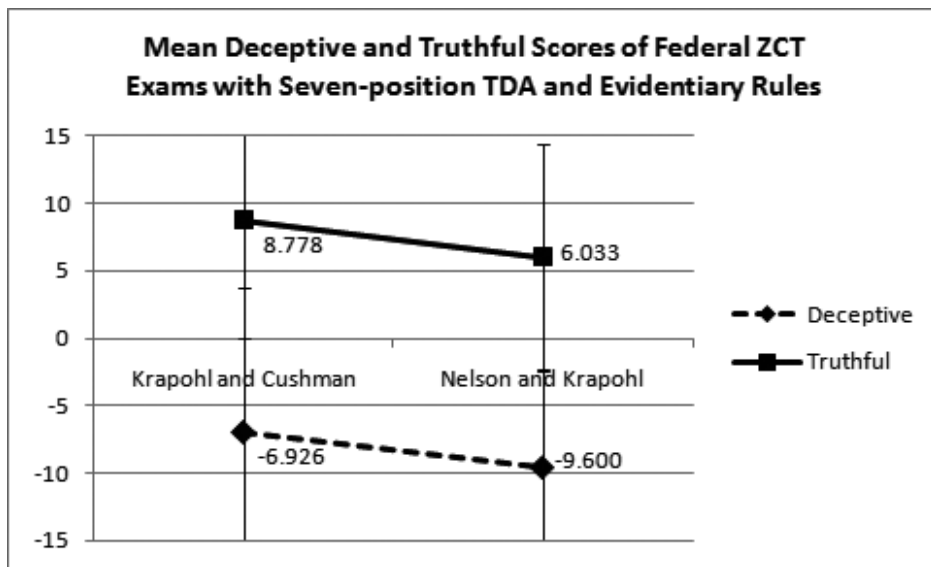
Крепол та Кушман (Krapohl & Cushman, 2006) описали критеріальну точність 100 поліграфологічних досліджень з Federal ZCT, оцінених 10 досвідченими поліграфологами з використанням семипозиційної моделі аналізу тестових даних та доказових правил прийняття рішень<sup>33</sup>. Загалом було отримано 1000 оцінених результатів. Поліграфологічні дослідження були вибрані з архіву підтверджених справ Міністерства оборони США. П'ятдесят досліджень було підтверджено як оманливі, а 50 – як правдиві. Незважена точність рішення становила .872, а незважений невизначений показник .073.

Нельсон і Крепол (Nelson & Krapohl, 2011) повідомили про критеріальну точність 60 поліграфологічних досліджень з Federal ZCT, які оцінювали шість досвідчених федеральних поліграфологів. Тридцять досліджень були підтверджені як оманливі, а 30 досліджень підтверджені як правдиві. Кожен оцінювач оцінював випадкову підмножину з 10 поліграфологічних досліджень. Результати оцінювали за допомогою федеральної семипозиційної моделі аналізу тестових даних та доказових правил прийняття рішень. Незважена точність прийняття рішень становила .870, а незважений невизначений показник становив .100.

<sup>33</sup> Крепол (Krapohl, 2005) і Крепол та Кушман (Krapohl & Cushman, 2006) встановили, що доказові правила прийняття рішень можуть суттєво зменшити кількість невизначених показників без відповідної втрати загальної точності рішень.

На рисунку 9 показано середнє значення та стандартне відхилення оцінювання методики Federal ZCT за семипозиційною системою оцінювання з використанням доказових правил. Двофакторний ANOVA показав, що взаємодія між розподілом вибірки та статусом критерію не була значущою [ $F(1,146) = 0.001$ , ( $p = .981$ )], а також не була основним ефектом для розподілу вибірки [ $F(1,146) = 0.046$ , ( $p = .830$ )].

Рисунок 9. Середнє значення оманливих та правдивих оцінок для поліграфологічних досліджень з Federal ZCT, оцінених за семипозиційним аналізом тестових даних з використанням доказових правил.



Сукупний рівень точності рішення наукових досліджень щодо поліграфологічних досліджень з Federal ZCT, оцінених із семипозиційним методом аналізу тестових даних з використанням доказових правил, зважений за розміром вибірки та кількістю оцінювачів, становив .872 із загальним показником невизначених показників .075. Надійність, розрахована як (bootstrap average of pairwise decision agreement) бутстреп середнє значення попарної узгодженості прийняття рішень без урахування невизначених результатів, становила .870.

### **Методика інтегрованого порівняння зон (Integrated Zone Comparison Technique IZCT)**

Методика інтегрованого порівняння зон (IZCT) (Gordon et al., 2000) є запатентованою діагностичною методикою щодо конкретної події, яка оцінюється за горизонтальною системою оцінювання (Horizontal Scoring System) (Gordon, 1999)<sup>34</sup>. Два наукові дослідження описують критеріальну точність цієї методики.

Гордон та ін. (Gordon et al., 2005; також описано в Mohamed et al., 2006) повідомили про результати пілотного дослідження за участю шести винних і п'яти

<sup>34</sup> Система підрахунку балів за рангом базується на унікальних інформативних ознаках, наданих розробниками.



невинних суб'єктів<sup>35</sup>, які брали участь у лабораторному дослідженні, пов'язаному з інцидентом зі стріляниною. Повідомляється, що точність рішення становила 1.000, незважений невизначений показник становив .100.

Шурані та Чавес (Shurani & Chaves, 2010) повідомили про результати вивчення 84 польових поліграфологічних досліджень, проведених з IZCT, які включали 44 оцінки підтверджених оманливих досліджень і 40 оцінок для підтверджених правдивих досліджень. Повідомляється, що всі дослідження були підтверджені зізнаннями, а деякі – доказами. Незважена точність рішення становила .988, з незваженим невизначеним показником .061. Статистика надійності для цього дослідження не надходила, і комітет не зміг розрахувати надійність між оцінювачами за наявними даними.

Шурані (Shurani, 2011) повідомив про результати польового дослідження за участю трьох поліграфологів з Коста-Ріки, які використовували IZCT разом із додатковою експериментальною методикою. Вибірка складалася з 73 випадків, для яких були протестовані всі можливі підозрювані. Було підтверджено 48 випадків, що призвело до  $N = 188$  досліджень, які були проведені за допомогою IZCT з трьома та чотирма RQ<sup>36</sup>. Два невизначені результати були видалені з повідомлених результатів. Про кількість тестів, проведених із трьома чи чотирма RQ, інформації не надходило. Однак комітету було надано дані 84 досліджень, які, за наявною інформацією, проводилися з використанням трьох RQ, включаючи бали для 36 оманливих випадків і 48 правдивих випадків. Оцінки решти 104 досліджень не були доступні. Не було повідомлено про середнє значення вибірки чи стандартне відхилення, і комітет не зміг порівняти середні значення вибірових даних, наданих комітету, з будь-якою опублікованою інформацією. Результати цього дослідження були представлені з абсолютною точністю та нульовими невизначеними результатами. Для цього дослідження не було надано статистичних даних щодо надійності, і комітет не зміг розрахувати надійність між оцінювачами за наявними даними.

На рисунку 10 показано графік середнього значення та стандартного відхилення розподілів вибірки наукових досліджень щодо IZCT. Двофакторний ANOVA показав значну взаємодію між розподілом вибірки та статусом випадку [ $F(1,173) = 533.771, (p < .001)$ ]. Post-hoc однофакторні ANOVAs показали, що відмінності у вибірці для оманливих випадків не були значними. Однак різниця в правдивих оцінках для трьох вибірок була значною [ $F(2,33) = 21.402, (p = .014)$ ]. Правдиві бали були значно вищими в дослідженнях Шурані та Чавеса (Shurani & Chaves, 2010) і Шурані (Shurani, 2011) порівняно з публікацією Гордона та ін. (Gordon et al., 2005).

Пізніше стало відомо, що наукове дослідження Гордона (Gordon et al., 2005) було проведено з використанням однотемних тестів IZCT, тоді як вибірові випадки Шурані та Чавеса (Shurani & Chaves, 2010) проводилися з використанням багатоперспективних досліджень з IZCT. Незрозуміло, чи ця різниця пояснюється значною вза-

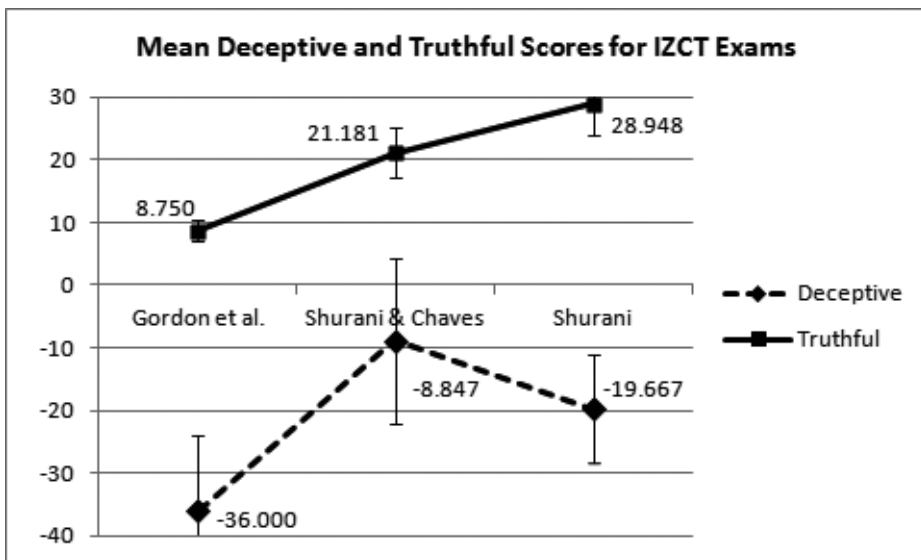
---

<sup>35</sup> Початковий план пілотного дослідження включав шість невинних суб'єктів, однак один правдивий суб'єкт зробив хибне зізнання поліграфологу (Гордон, особисте спілкування 6 липня 2011 р.), який також був основним автором публікації (Gordon et al., 2005) та розробником IZCT. Включення хибнопозитивної (хибного зізнання) помилки призвело б до меншої, ніж ідеальна точність.

<sup>36</sup> Немає опублікованого опису для використання IZCT з чотирма RQ. Оскільки IZCT оцінюється з використанням парадигми рангового порядку, можна очікувати, що включення додаткових RQ без включення еквівалентної кількості додаткових CQ по-різному (диференційовано) вплине на оцінки суми рангів для релевантних запитань і CQ. Жодне опубліковане дослідження не описувало і не досліджувало ці статистичні складності.

ємодію та відмінностями, що спостерігаються в цих розподілах вибірки<sup>37</sup>. Двофакторне порівняння ANOVA, розподілу оцінок вибірки, розподілу вибірки Шурані та Чавеса (Shurani & Chavez, 2010) і Шурані (Shurani, 2011) виявило значну взаємодію [ $F(1,164) = 43.140, (p < .001)$ ], що свідчить про те, що оцінки оманливих і правдивих випадків були виражені або інтерпретовані по-різному у вибірках наукових досліджень Шурані та Чавеса (Shurani & Chavez, 2010) та Шурані (Shurani, 2011). Однофакторні відмінності не були значними. Оцінки наукового дослідження Шурані (Shurani, 2011) були більш далекі від нуля, ніж оцінки наукового дослідження Шурані та Чавеса (Shurani & Chavez, 2010) як для оманливих, так і для правдивих випадків.

Рисунок 10. Середнє значення оманливих та правдивих оцінок для вибірки IZCT.



Сукупний рівень точності рішення в цих наукових дослідженнях щодо IZCT, зважений за розміром вибірки та кількістю оцінювачів, склав .994 із сукупним невизначеним показником .033.

Жодної статистики надійності для жодного наукового дослідження IZCT не було надано, і комітет не зміг розрахувати надійність між оцінювачами за наявними даними.

### **Методика «квадратичного» порівняння зон Матте (Matte Quadri-track Zone Comparison Technique, MQTZCT)**

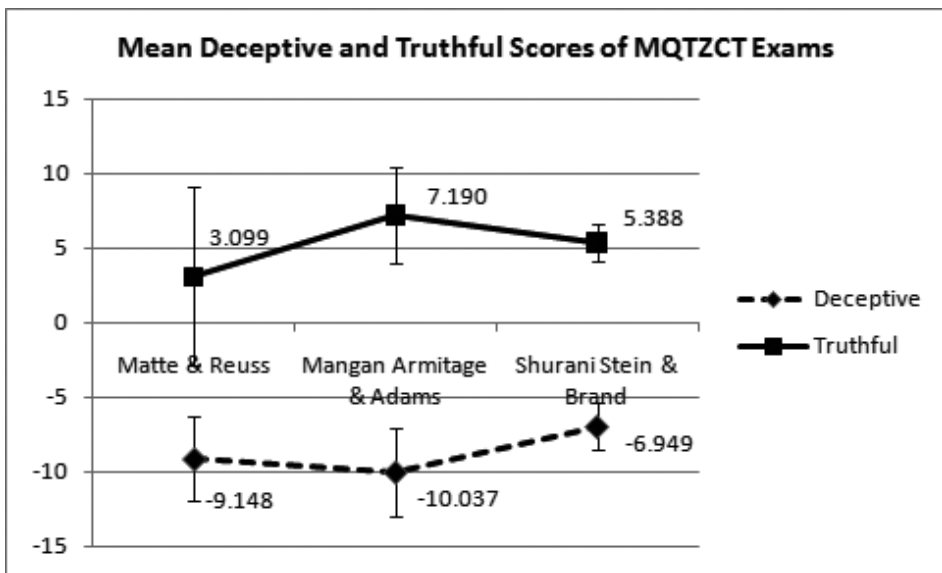
Методика квадратичного порівняння зон Матте (MQTZCT) (Matte & Reuss, 1989) є запатентованою діагностичною однотемною методикою щодо конкретної проблеми, яка оцінюється за допомогою модифікації числової системи Бакстера. Три дослідження описують критеріальну точність MQTZCT.

<sup>37</sup> У практичному сенсі відмінності в припущеннях щодо незалежності та відсутності незалежності серед запитань тесту призведуть до використання різних правил прийняття рішень, і ці відмінності можуть мати упереджений вплив на підтвердження випадку та вибір вибірки для цих польових досліджень. Рангові оцінки для всіх RQ завжди відносяться до всіх інших релевантних і порівняльних тестових стимулів. Таким чином, бали за порядком рангу за своєю суттю є залежними, і математичне обґрунтування застосування моделі оцінки за порядком рангу до багатаспектних тестів, для яких правила прийняття рішень базуються на припущенні незалежності, не є зрозумілим. Ця нетривіальна статистична та теоретична складність не була належним чином обговорена чи вивчена.

Матте і Ройс (Matte & Reuss, 1989) повідомили про результати 64 оманливих і 58 правдивих випадків, які були підтверджені комбінацією зізнань та інших доказів. Незважена точність прийняття рішень була визначена як ідеальна 1.000, з незваженим невизначеним показником .059.

Манган, Армітідж і Адамс (Mangan, Armitage & Adams, 2008) повідомили про критеріальну точність вивчення 91 оманливих випадків та 45 правдивих випадків, які були підтверджені через зізнання досліджуваного. Точність рішення знову була зареєстрована як ідеальна 1.000, з незваженим невизначеним показником .011.

Рисунок 11. Середні значення оманливих та правдивих оцінок кожного чарту (per-chart scores) для вибірки MQTZCT.



Шурані, Стейн і Бренд (Shurani, Stein & Brand, 2009) повідомили про критеріальну точність огляду 28 оманливих і 29 правдивих випадків, які були підтверджені зізнанням разом з додатковими доказами для деяких випадків. Повідомляється, що точність рішення становила .964 з нульовими невизначеними результатами. На рисунку 11 показано графік середнього значення та стандартного відхилення оцінок за окремо взятим релевантним запитанням (subtotal scores)<sup>38</sup> розподілів вибірки трьох наукових досліджень MQTZCT<sup>39</sup>. Двофакторний ANOVA виявив значну взає-

<sup>38</sup> Оцінки досліджень MQTZCT повідомляються як оцінка за окремо взятим релевантним запитанням за чартом (subtotal per chart), отриманий шляхом підсумовування всіх числових балів у кожному чарті. Оцінки за окремо взятим релевантним запитанням, описані в іншому місці цього звіту, включають оцінки за окремо взятими релевантними запитаннями між чартами (between-chart RQ subtotals), отримані шляхом підсумовування числових балів для кожного RQ для всіх чартів.

<sup>39</sup> Дані, спочатку надані спеціальному комітету для досліджень Мангана, Армітіджа та Адамса (Mangan, Armitage & Adams, 2008) і Шурані та Чавеса (Shurani & Chaves, 2009), включали лише ті оцінки, для яких оцінювачі досягли правильного результату, і не включали оцінки для невизначених або помилкових результатів. Відсутні бали пізніше були надані комітету для обох досліджень Мангана, Армітіджа та Адамса (Mangan, Armitage & Adams, 2008) та Шурані, Стейна та Бренда (Shurani, Stein & Brand, 2009). Однак отримані розподіли вибірки відрізнялися від тих, що повідомлялися для обох наукових досліджень. Через ці розбіжності, які викликають занепокоєння, статистичний аналіз не було перераховано з відсутніми балами, і звітний аналіз відображає середні бали, як повідомили Манган, Армітідж і Адамс (Mangan, Armitage & Adams, 2008) і Шурані та Чавес (Shurani & Chavez, 2009). Результатом цієї плутанини є те, що вибіркові розподіли, як повідомляється, повинні вважатися систематично позбавленими помилок або незрозумілої дисперсії, а тому не підлягають узагальненню.

модію між розподілом вибірки та статусом випадку [ $F(1,261) = 361.605$ , ( $p < .001$ )]. Незважаючи на те, що різні наукові дослідження, здавалося, здатні справлятися з оманливими та правдивими випадками з різною ефективністю, постфактум однофакторні ANOVAs показали, що відмінності в оцінках не були значущими для оманливих випадків [ $F(2,141) = 0.389$ , ( $p = .678$ ) або для правдивих випадків [ $F(2,122) = 0.264$ , ( $p < .768$ )].

Сукупний рівень точності рішення в наукових дослідженнях MQTZCT, зважений за розміром вибірки та кількістю оцінювачів, склав .994 із сукупним коефіцієнтом невизначених показників .029. Надійність поліграфологічних досліджень MQTZCT була визначена Матте та Ройссом (Matte & Reuss 1989) як .990<sup>40</sup>.

### **Utah ZCT – Probable Lie Test**

### **Юта ZCT – методика ймовірного обману**

Методика Utah ZCT Probable Lie Test<sup>41</sup> з використанням ймовірного обману, яку також називають Utah Probable Lie Test (PLT) (Handler 2006; Handler & Nelson, 2008) з системою числового оцінювання штату Utah (Bell et al., 1999; Handler & Nelson, 2008) були розроблені дослідниками з Університету Юти як модифікація Backster ZCT (Backster, 1963). Два наукові дослідження описують критеріальну точність методики Utah ZCT PLT.

Хонтс, Раскін і Кірчер (Honts, Raskin & Kircher, 1987) повідомили про результати 10 запрограмованих оманливих і 10 запрограмованих правдивих досліджуваних у науковому дослідженні щодо контрзаходів поліграфу<sup>42</sup>. Незважена точність рішення сліпих числових оцінок становила .889, з невизначеним показником .150<sup>43</sup>.

Кірчер і Раскін (Kircher & Raskin, 1988) повідомили про результати двох оцінювачів, які в лабораторних дослідженнях оцінили 50 запрограмованих оманливих і 50 запрограмованих правдивих досліджуваних. Всього було отримано 200 оцінок. Незважена точність рішення сліпих числових оцінок становила .935, з невизначеним показником .070.

На рисунку 12 показано графік середнього значення та стандартного відхилення оцінок розподілів вибірки включених наукових досліджень щодо Utah PLC. Двофакторний ANOVA показав, що взаємодія між розподілом вибірки та статусом критерію не була значущою [ $F(1,63) = 1.682$ , ( $p = .200$ )], а також не була основним ефектом для розподілу вибірки [ $F(1,63) = 0.108$ , ( $p = .743$ )].

---

<sup>40</sup> Ці статистичні дані були опубліковані в передруку дисертації Матте і Рейсса (Matte & Reuss, 1989), опублікованого в журналі *Polygraph*, але їх неможливо знайти в оригінальному дослідженні дисертації для Колумбійського Тихоокеанського університету, якого вже не існує.

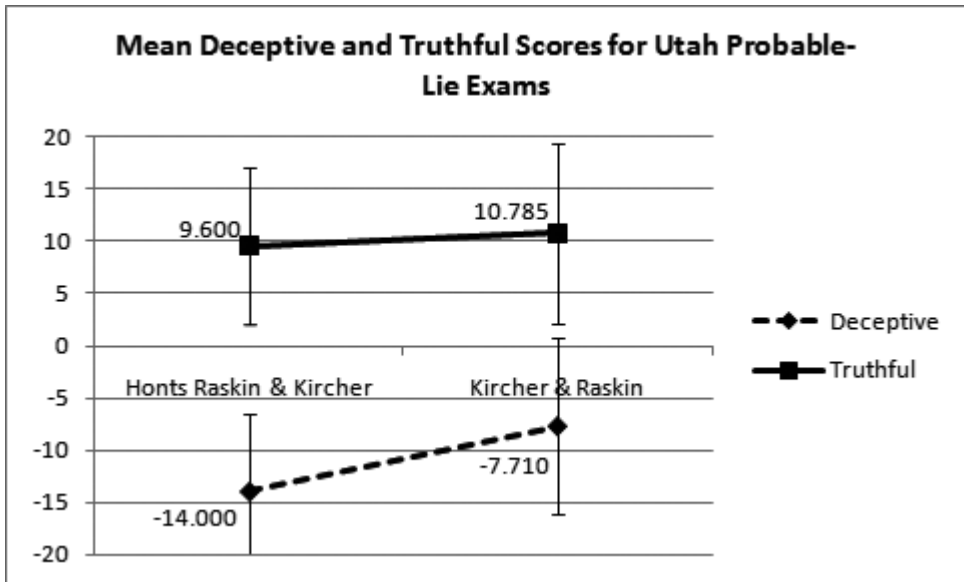
<sup>41</sup> Схоже, що розробники методики Utah мало хвилювалися за назву формату тестових запитань, і цей формат також називають версією Юта з 3 запитаннями та методики ймовірного обману штату Юта (Utah 3-question version and the Utah PLT). Поліграфологи-практики використовували термін Юта ZCT через очевидну схожість з іншими варіантами ZCT. Термін Utah ZCT використовується в цьому документі, щоб допомогти у розпізнаванні процедурної та практичної подібності між цією методикою та іншими форматами ZCT із трьома запитаннями, призначеними для одностороннього тестування щодо конкретної події (*single-issue event-specific testing*).

<sup>42</sup> У цей аналіз включено лише випадки контрольної групи без застосування контрзаходів.

<sup>43</sup> Хонтс, Раскін та Кірчер (Honts, Raskin & Kircher, 1987) повідомили середні оцінки, але редакційні та публікаційні стандарти не вимагали повідомляти про стандартні відхилення для вибіркового розподілу оманливих і правдивих оцінок і оманливі оцінки на момент публікації. Оскільки дані для обчислення цих відсутніх статистичних даних більше не були доступні, точну оцінку сукупного стандартного відхилення було розраховано на основі повідомленого  $t$ -значення для рівня значущості різниці між правдивими та оманливими оцінками.

Сукупний рівень точності рішення в цих наукових дослідженнях щодо Utah PLT, зважений за розміром вибірки та кількістю оцінювачів, склав .931 із загальним коефіцієнтом невизначеності .077. Надійність поліграфологічних досліджень Utah PLC, виражена як середнє значення Каппа-статистики для двох наукових досліджень, становила .730, з попарним показником узгодженості рішень, за винятком невизначених результатів .975.

Рисунок 12. Середнє значення оманливих та правдивих сукупних балів для наукових досліджень щодо Utah PLT



### Utah ZCT – Directed Lie Test

#### Utah ZCT – методика керованого обману

Utah ZCT Directed Lie Test (DLT) є варіацією методики Utah (PLT), в якій використовуються запитання порівняння керованого обману замість запитань порівняння ймовірного обману.

Два наукові дослідження описують критеріальну точність поліграфологічних досліджень з Utah DLC.

Хонтс і Раскін (Honts & Raskin, 1988) повідомили про критеріальну точність поліграфологічних досліджень з Utah DLC відносно 25 підозрюваних у злочинах, у тому числі 12 осіб, які брехали, і 13 осіб, які були правдиві, результати яких з часом були підтверджені їхніми зізнаннями, зізнаннями інших підозрюваних, доказами або зняттям звинувачень. Незважена точність рішення сліпих числових оцінок становила .958, з невизначеним показником .077<sup>44</sup>.

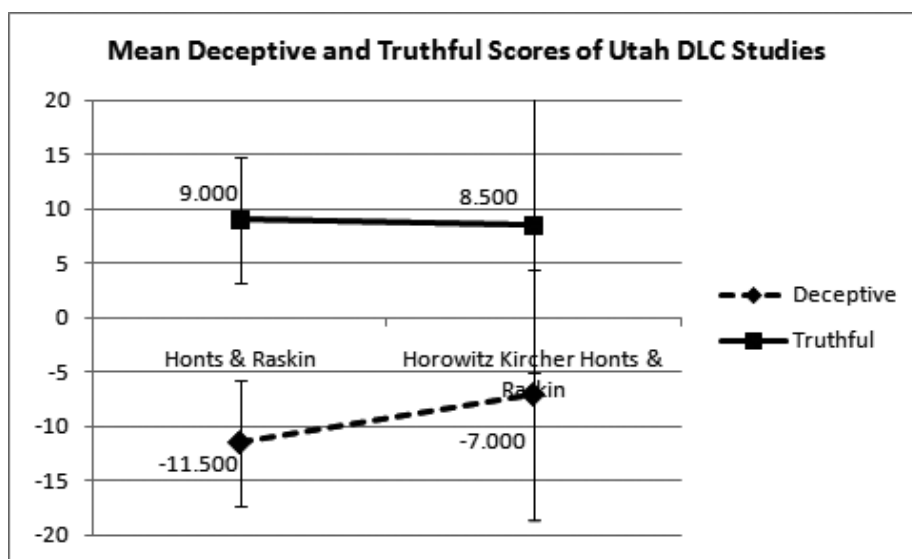
<sup>44</sup> Хонтс та Раскін (Honts & Raskin, 1988) повідомили про середні бали, але редакційні та публікаційні стандарти не вимагали звітувати про стандартні відхилення для вибіркового розподілу оманливих і правдивих оцінок, і оманливі оцінки на момент публікації. Оскільки дані для обчислення цих відсутніх статистичних даних більше не були доступні, точну оцінку сукупного стандартного відхилення було розраховано на основі повідомленого F-коефіцієнта для рівня значущості різниці між правдивими та оманливими оцінками.

Горовіц, Кірчер, Хонтс і Раскін (Horowitz, Kircher, Honts & Raskin, 1997) повідомили про результати 15 запрограмованих оманливих і 15 запрограмованих правдивих досліджуваних, які брали участь у лабораторному експерименті. Незважена точність рішення сліпих числових оцінок становила .856, з невизначеним показником .067<sup>45</sup>.

На рисунку 13 показано графік середнього значення та стандартного відхилення оцінок розподілів вибірки включених в наукові дослідження з Utah DLT. Двофакторний ANOVA показав, що взаємодія розподілу вибірки та статусу критерію не була значущою [F (4,51) = 0.705, (p = .592)], як і основний ефект для розподілу вибірки [F (2,51) = 0.009, (p = .991)].

Сукупний рівень точності рішень у цих наукових дослідженнях з Utah DLC, зважений за розміром вибірки та кількістю оцінювачів, склав .902 із сукупним показником невизначених результатів .073. Надійність поліграфологічних досліджень з Utah DLC, виражена як середнє значення коефіцієнтів кореляції Пірсона для включених наукових досліджень, становила .930.

Рисунок 13. Середні значення оманливих та правдивих сукупних оцінок для наукових досліджень щодо методики Utah DLT.



### **Utah ZCT – Версія канадського поліцейського коледжу / Королівської канадської кінної поліції**

Канадський поліцейський коледж (Canadian Police College, CPC) і Королівська канадська кінна поліція (RCMP) розробили варіант Utah PLT, який називається RCMP Zone або CPC Series A. Три наукові дослідження описують критеріальну точність поліграфологічних досліджень з Utah RCMP Series A.

Хонтс, Ходес і Раскін (Honts, Hodes & Raskin, 1985) повідомили про критеріальну точність поліграфологічних досліджень з Utah PLT, використовуючи послідовність

<sup>45</sup> Середнє значення та статистику стандартного відхилення було виміряно до найближчої точки 1/2 від рисунку 1 у звіті про дослідження Горовіца, Кірчера, Хонтса та Раскіна (Horowitz, Kircher, Honts & Raskin, 1997).

тестових запитань RCMP Series A, які включали 19 оманливих і 19 правдивих випадків. Незважена точність рішення сліпих числових оцінок становила .833, з невизначеним показником .237.

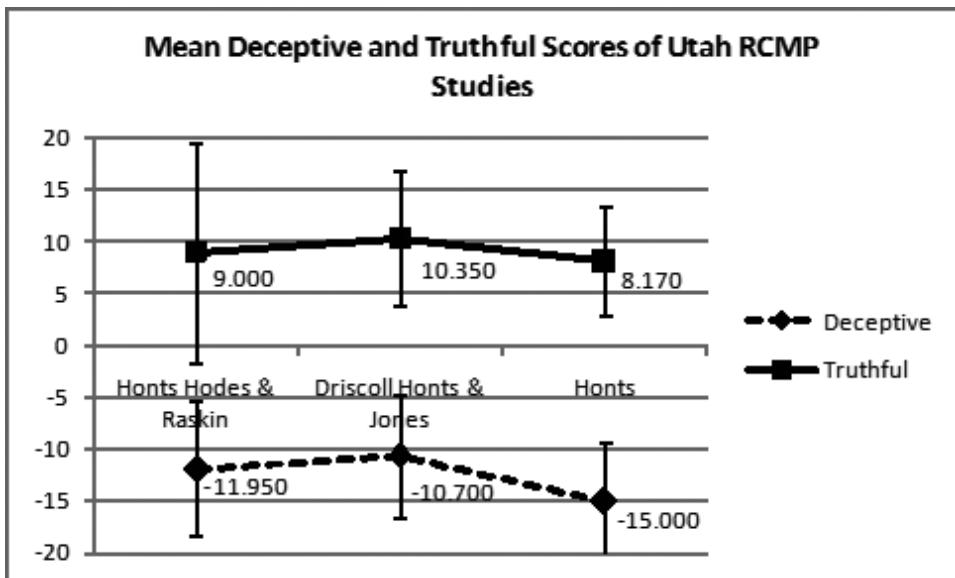
Дрісколл, Хонтс і Джонс (Driscoll, Honts & Jones, 1987) повідомили про критеріальну точність поліграфологічних досліджень Utah PLT, використовуючи результати 20 запрограмованих оманливих і 20 запрограмованих правдивих досліджуваних, які були набрані з програми групового консультування в Центрі ветеранів, використовуючи послідовність тестових запитань досліджень RCMP Series A. Повідомляється, що точність рішення становила 1.000, незважений невизначений показник становив .100.

Хонтс (Honts, 1996) повідомив про результати дослідження критеріальної точності польових перевірок, проведених канадськими правоохоронними органами з використанням версії RCMP Utah PLT. 21 випадок був підтверджений як оманливий, а 11 випадків були підтверджені як правдиві. Незважена точність рішення сліпих числових балів становила .969 з невизначеним показником .210.

На рисунку 14 показано графік середнього значення та стандартного відхилення балів розподілів вибірки включених досліджень Utah CPC-RCMP. Двофакторний ANOVA показав, що ні взаємодія розподілу вибірки та статусу критерію [ $F(1,99) = 0.562$ , ( $p = .455$ )], ні основний ефект для розподілу вибірки [ $F(1,99) = 0.109$  ( $p = .742$ )] не були статистично значущими.

Сукупний рівень точності рішень у наукових дослідженнях Utah CPC-RCMP, зважений за розміром вибірки та кількістю оцінювачів, склав .939 із сукупним показником невизначених результатів .183. Хонтс (Honts, 1996) повідомив про надійність досліджень з Utah RCMP як Каппа = .480 для категоріального рішення згоди з поправкою на випадкову згоду. Середній попарний коефіцієнт кореляції Пірсона для числових оцінок включених наукових досліджень становив .940, а середня частка узгоджених рішень, за винятком невизначених результатів, становила .883.

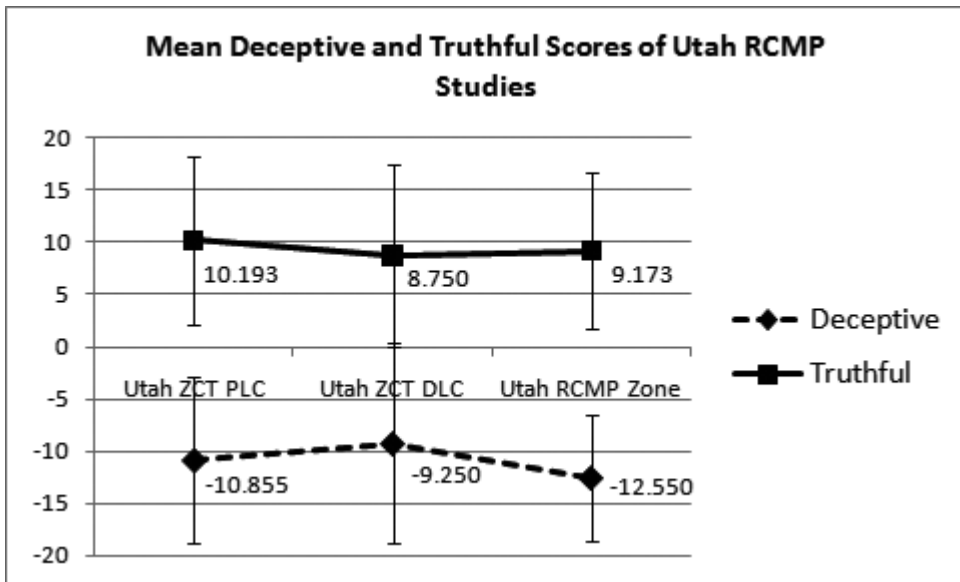
Рисунок 14. Середнє значення оманливих та правдивих сукупних оцінок для наукових досліджень Utah CPC-RCMP.



## Utah ZCT – сукупні наукові дослідження щодо PLT, DLT і RCMP

На рисунку 15 показано графік середнього значення та стандартного відхилення для трьох варіантів Utah PLT. Двофакторний ANOVA показав, що взаємодія між варіантом тесту та статусом критерію не була значущою [ $F(1,246) = 2.553, (p = .111)$ ], як і основний ефект для розподілу вибірки [ $F(2,246) = 0.02, (p = .980)$ ]. Оскільки взаємодія наближалася до значного рівня, було також постфактум завершено однофакторний *posthoc* ANOVAs. Відмінності між розподілами вибірки не були значущими для оманливих оцінок [ $F(2,100) = 0.042, (p = .959)$ ] або для правдивих оцінок [ $F(2,100) = 0.008, (p = .992)$ ].

Рисунок 15. Середні значення оманливих та правдивих сукупних оцінок з наукових досліджень щодо трьох варіантів Utah ZCT.



Незважаючи на точність рішення для семи включених наукових досліджень, що стосуються трьох варіантів методики Utah, зважена за розміром вибірки та кількістю оцінювачів, становила .930, з незваженим невизначеним показником .107. Статистику надійності було усереднено для всіх включених наукових досліджень щодо Utah ZCT, що дало середню статистику надійності  $\kappa = .647$ . Середній показник узгодженості рішень без урахування невизначених результатів становив .958, а середній коефіцієнт кореляції Пірсона для числових оцінок становив .913.

### Event-Specific ZCT / ESS

ESS – це заснована на доказах модель аналізу тестових даних, яка включає нормативні дані для досліджень ZCT та інших поліграфологічних методик. Оскільки перетворення ESS є непараметричними, її оцінки є чутливими до відмінностей (різниці) у величині реакції, але є стійкими до відмінностей у лінійності величини реакції.



Нельсон та ін. (Nelson et al. 2011) повідомили про короткий виклад п'яти попередніх наукових досліджень критеріальної точності оцінок ESS в поліграфологічних дослідженнях з ZCT, включаючи результати, повідомлені Нельсоном, Креполом і Хендлером (Nelson, Krapohl & Handler, 2008), Нельсоном, Блелоком і Кушманом (Nelson, Blalock, Oelrich & Cushman, 2009), Нельсоном, Блелоком, Оелріхом і Кушманом (Nelson, Blalock, Oelrich & Cushman, 2011), Хендлером, Нельсоном, Гудсоном і Хіксом (Handler, Nelson, Goodson & Hicks, 2010) і Нельсоном і Креполом (Nelson & Krapohl, 2011). Ці наукові дослідження включали 5192 результатів оцінок 140 оцінювачів, які оцінювали 732 індивідуальні дослідження. Ці результати склалися з 2671 результату оцінювання 384 підтверджених оманливих досліджень і 2521 результату 348 підтверджених правдивих досліджень. Дослідження включали як федеральний ZCT, так і Utah ZCT. Незважена точність прийняття рішень за цими оцінками, за винятком невизначених результатів, становила .921, а незважений невизначений показник становив .098.

Нельсон, Крепол і Хендлер (Nelson, Krapohl & Handler, 2008) повідомили про критеріальну точність оцінювання з ESS семи недосвідчених поліграфологів, які використовували ESS для оцінки вибірки зі 100 досліджень, відібраних із підтвердженого архіву справ Міністерства оборони США. П'ятдесят досліджень було підтверджено як оманливі, а 50 – як правдиві. Всього було отримано 700 оцінок. Незважена точність рішення сліпого числового оцінювання становила .872 з невизначеним показником .103.

Блелок, Кушман і Нельсон (Blalock, Cushman & Nelson, 2009) у реплікаційному дослідженні повідомили про критеріальну точність групи з дев'яти поліграфологів початківців, які використовували ESS для оцінки вибірки зі 100 досліджень, відібраних із підтвердженого архіву справ Міністерства оборони США. П'ятдесят досліджень було підтверджено як оманливі, а 50 – як правдиві. Всього було отримано 900 оцінок. Незважена точність рішення сліпого числового оцінювання становила .870, з невизначеним показником .138.

Нельсон, Блелок, Елріх і Кушман (Nelson, Blalock, Oelrich & Cushman, 2011) повідомили про результати наукового дослідження щодо надійності за участю 25 досвідчених поліграфологів, які використовували ESS для оцінки вибірки з 10 поліграфологічних досліджень, відібраних із підтвердженого архіву справ Міністерства оборони США. Шість випадків були підтверджені як оманливі, а чотири випадки підтверджені як правдиві. Всього було отримано 250 оцінок. Попарна частка узгодженості рішень становила .950, а незважене середнє значення правильних рішень без урахування невизначених результатів становило .958. Незважений невизначений показник становив .102.

Хендлер, Нельсон, Гудсон і Хікс (Handler, Nelson, Goodson & Hicks, 2010) повідомили про критеріальну точність від 19 поліграфологів-початківців з México Policía Federal, які використовували ESS для оцінки 100 поліграфологічних досліджень, вибраних із підтвердженого архіву справ Міністерства оборони США. П'ятдесят досліджень було підтверджено як оманливі, а 50 – як правдиві. Всього було отримано 1900 оцінок. Незважена точність рішення сліпого числового оцінювання становила .901, з невизначеним показником .040.

Нельсон і Крепол (Nelson & Krapohl, 2011) повідомили про критеріальну точність трансформованих балів ESS від шести досвідчених федеральних поліграфологів, які

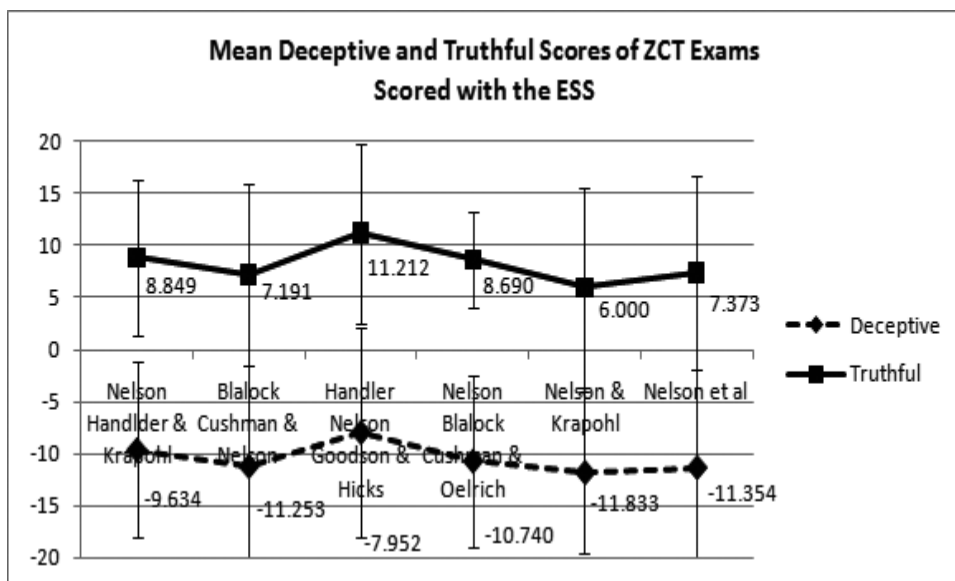
оцінювали вибірку з 60 поліграфологічних досліджень, відібраних із підтвердженого архіву справ Міністерства оборони США. Кожен поліграфолог оцінив 10 випадків. Тридцять досліджень були підтверджені як оманливі, а 30 досліджень – як правдиві. Незважена точність рішення сліпого числового оцінювання становила .913, з незваженим невизначеним показником .020.

Решта оцінок Нельсона та ін. (Nelson et al., 2011) склалися з 1382 оцінених результатів 572 індивідуальних поліграфологічних досліджень. Ці результати склалися з 741 оціненого результату 304 підтверджених оманливих досліджень і 641 оціненого результату 268 підтверджених правдивих досліджень. Дані дослідження Крепола і Кушмана (Krapohl & Cushman, 2006) були оцінені за допомогою автоматизованої версії ESS, включаючи 50 підтверджених оманливих досліджень і 50 підтверджених правдивих досліджень, вибраних із підтвердженого архіву справ Міністерства оборони США. Ці тести також оцінювала група з 11 поліграфологів початківців із підрозділу контррозвідки колумбійської армії (Army Counterintelligence Unit), які використовували ESS у парах із двох і трьох поліграфологів, щоб оцінити по 10 випадків кожен. Дані з вибірки, яку використали Крепол і Макманус (Krapohl & McManus, 1999), також оцінювали за допомогою автоматизованої версії ESS, включаючи 30 підтверджених оманливих досліджень і 30 підтверджених правдивих досліджень. Ця вибірка також була оцінена групою з 35 поліграфологів з Румунії, що складалася з 15 міжнародних поліграфологів і 20 дослідників, психологів і аспірантів з Яського Університету в Румунії, які використовували ESS під час роботи в командах, щоб оцінити підгрупи по 10 випадків кожна. Крім того відкладена вибірка також була оцінена групою з 12 поліграфологів початківців Національної поліції Панамі, які працювали в командах, щоб оцінити підмножини по 10 випадків кожна. Крім того, семеро поліграфологів початківців з поліцейських установ штату Огайо використовували ESS, щоб підрахувати підмножини з 10 випадків у кожній із відкладених вибірок. Одну підгрупу з 10 випадків оцінили двоє новачків поліції Огайо.

Числові оцінки з дослідження Кірчера, Крістжансона, Гарднера та Веба (Kircher, Kristjansson, Gardner & Web, 2005) були перетворені в оцінки ESS, включаючи 40 оцінок підтверджених оманливих досліджень і 40 оцінок підтверджених правдивих досліджень. Дані з вибірки розробки OSS (Krapohl, 2002; Krapohl & McManus, 1999; Nelson, Krapohl & Handler, 2008) були оцінені за допомогою автоматизованої моделі ESS, включаючи 149 оцінок підтверджених оманливих досліджень і 143 оцінки підтверджених правдивих тестів. Семипозиційні оцінки від двох поліграфологів, які брали участь у науковому дослідженні Кірчера та Раскіна (Kircher & Raskin 1988), були перетворені в оцінки ESS, включаючи 100 оцінок 50 запрограмованих оманливих досліджень і 100 оцінок 50 запрограмованих правдивих досліджень. Нарешті, семипозиційні оцінки від трьох поліграфологів, які оцінювали випадки для дослідження Блеквела (Blackwell, 1998), були перетворені в оцінки ESS, включаючи 195 оцінок 65 підтверджених оманливих досліджень і 105 оцінок 35 підтверджених правдивих досліджень.

На рисунку 16 показано графік середнього значення та стандартного відхилення оцінок розподілів вибірки включених наукових досліджень ZCT ESS. Двофакторний ANOVA показав, що взаємодія між розподілом вибірки та статусом критерію не була значущою [ $F(1,215) = 0.205$ , ( $p = .651$ )], як і основний ефект для розподілу вибірки [ $F(5,215) = 0.164$ , ( $p = .976$ )].

Рисунок 16. Середнє значення оманливих та правдивих сукупних оцінок для наукових досліджень щодо ZCT/ESS.



Сукупний рівень точності рішення включених наукових досліджень щодо ZCT ESS, зважений за розміром вибірки та кількістю оцінювачів, становив .922 із сукупним невизначеним показником .098, як повідомили Нельсон та ін. (Nelson et al. 2011). Статистика надійності для наукових досліджень щодо ZCT ESS була середньою для отримання попарної частки узгодженості рішень, за винятком невизначених результатів, .950, із середнім Каппа = .585.

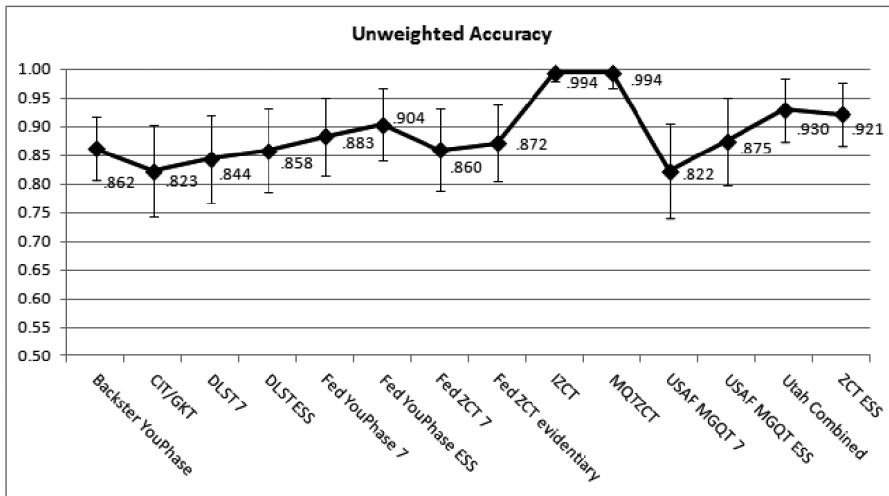
### **Критеріальна точність для всіх валідних методик**

Однофакторний ANOVA для незваженої точності рішень показав, що відмінності в незваженій точності рішень для цих 14 поліграфологічних методик були значущими [ $F(13,5119) = 2.753, (p < .001)$ ]. Однофакторні ANOVAs для статусу справи показали, що відмінності в правильних рішеннях були значущими як для критеріальних оманливих випадків [ $F(13,2494) = 1.982, (p = .019)$ ], рисунку 17 показано середнє значення та довірчі інтервали для незваженої точності 14 методик, включених до мета-аналізу.

Серія контрастів ANOVA показала, що відмінності були суттєвими лише для двох поліграфологічних методик, а саме IZCT та MQTZCT. Виключення цих двох поліграфологічних методик не призвело до суттєвих відмінностей [ $F(11,4859) = 0.949, (p = .491)$ ] у незваженій точності для решти 12 поліграфологічних методик. Т-тести для однієї вибірки додатково підтвердили статус викиду результатів цих двох методик:  $t = 212.268 (p < .001)$  як для IZCT, так і для MQTZCT<sup>46</sup>. Серія виключаючих t-тестів (leave-one-out t-tests) показала, що жодна з інших методик не дала результатів, що відрізняються від результатів усіх інших методик.

<sup>46</sup> t-значення однакові для обох цих методик, оскільки незважене середнє значення зважених середніх вибірових значень однакове для обох методик (.994).

Рисунок 17. Середнє значення та довірчі інтервали для незваженої точності рішення 14 поліграфологічних методик.



### Критеріальна валідність відповідно до Стандартів АПА 2012 року

Таблиця 1 (також наведена в Короткому описі) показує перелік із 14 поліграфологічних методик, які задовольняють вимогам для включення в цей мета-аналіз на рівнях критеріальної точності, визначених у стандартних вимогах АПА 2012 для доказового, парного та дослідницького тестувань. Також у таблиці 1 наведено незважену точність рішень і невизначені показники для кожної поліграфологічної методики. Додаткові відомості щодо розподілу вибірки та повного розмірного профілю критеріальної точності для кожної з цих методик і всіх включених досліджень можна знайти в Додатку Е.

Комбінація всіх валідних поліграфологічних методик, за винятком результатів-викидів, дала точність рішення .869 (.036) без невизначених результатів. 95% довірчий інтервал становив від .798 до .940. Середній показник невизначених результатів, за винятком результатів-викидів, становив .128 (.030) з 95% довірчим інтервалом від .068 до .187. Агрегована статистика надійності дала середню статистику Каппа .642 (.102) з 95% довірчим інтервалом від .443 до .842. Середній показник узгодженості рішень між оцінювачами, за винятком результатів-викидів, і за винятком невизначених результатів, становив .901 (.082) з 95% довірчим інтервалом від .741 до .999. Середній коефіцієнт кореляції Пірсона для числових оцінок, за винятком результатів-викидів, становив .876 (.116) з 95% довірчим інтервалом від .649 до .999. Таблиця 2 показує агрегований профіль критеріальної точності усіх валідних поліграфологічних методик CQT, зважений за розміром вибірки та кількістю оцінювачів<sup>47</sup>. У таблиці 2 також показано профіль критеріальної точності, включаючи результати-викиди.

<sup>47</sup> Більшість досліджень у польових поліграфологічних програмах проводяться з використанням поліграфологічних методик, які інтерпретуються з припущенням незалежності критеріїв серед RQ. Однак більшість досліджень поліграфологічної критеріальної валідності було проведено з використанням поліграфологічних методик, які інтерпретуються з припущенням про відсутність незалежності. Не-незалежна поліграфологічна методика має більшу статистичну дискримінаційну потужність і більшу точність, ніж незалежні поліграфологічні методики. З цієї причини незважене середнє значення вважалось більш консервативною та узагальненою оцінкою загальної точності всіх поліграфологічних методик. Це було розраховано як незважену середню величину зваженої сукупності незалежних методик і зваженої сукупності не-незалежних методик. Розрахунок середньозваженого значення призведе до переоцінки точності. Незважене середнє значення PPV і NPV може бути більш оптимістичним і прийнятним за деяких умов, але можна очікувати, що воно буде менш узагальненим для польових обставин, коли базові ставки (коефіцієнти) невідомі або відрізняються від базових ставок у вибірках наукових досліджень.

Таблиця 1. Середнє значення (стандартне відхилення) і {95% довірчі інтервали} для правильних рішень (CD) і невизначених результатів (INC) валідних поліграфологічних методик.		
Evidentiary Techniques/TDA Method Доказові методики / метод аналізу тестових даних	Paired Testing Techniques/TDA Method Методики для парного тестування / метод аналізу тестових даних	Investigative Techniques/TDA Method Дослідницькі методики / метод аналізу тестових даних
<b>Federal You-Phase / ESS<sup>1</sup></b> CD = .904 (.032) {.841 to .966} INC = .192 (.033) {.127 to .256}	<b>AFMQT<sup>4,8</sup> / ESS<sup>5</sup></b> CD = .875 (.039) {.798 to .953} INC = .170 (.036) {.100 to .241}	<b>AFMQT<sup>6,8</sup> / 7 position</b> CD = .817 (.042) {.734 to .900} INC = .197 (.030) {.138 to .255}
<b>Event-Specific ZCT / ESS</b> CD = .921 (.028) {.866 to .977} INC = .098 (.030) {.039 to .157}	<b>Backster You-Phase / Backster</b> CD = .862 (.037) {.787 to .932} INC = .196 (.040) {.117 to .275}	<b>CIT<sup>7</sup> / Lykken Scoring</b> CD = .823 (.041) {.744 to .903} INC = NA
<b>IZCT / Horizontal<sup>2</sup></b> CD = .994 (.008) {.978 to .999} INC = .033 (.019) {.001 to .069}	<b>Federal You-Phase / 7 position</b> CD = .883 (.035) {.813 to .952} INC = .168 (.037) {.096 to .241}	<b>DLST (TES)<sup>8</sup> / 7 position</b> CD = .844 (.039) {.768 to .920} INC = .088 (.028) {.034 to .142}
<b>MQTZCT / Matte<sup>3</sup></b> CD = .994 (.013) {.968 to .999} INC = .029 (.015) {.001 to .058}	<b>Federal ZCT / 7 position</b> CD = .860 (.037) {.801 to .945} INC = .171 (.040) {.113 to .269}	<b>DLST (TES)<sup>8</sup> / ESS</b> CD = .858 (.037) {.786 to .930} INC = .090 (.026) {.039 to .142}
<b>Utah ZCT DLT / Utah</b> CD = .902 (.031) {.841 to .962} INC = .073 (.025) {.023 to .122}	<b>Federal ZCT / 7 pos. evidentiary</b> CD = .880 (.034) {.813 to .948} INC = .085 (.029) {.028 to .141}	–
<b>Utah ZCT PLI / Utah</b> CD = .931 (.026) {.879 to .983} INC = .077 (.028) {.022 to .133}	–	–
<b>Utah ZCT Combined / Utah</b> CD = .930 (.026) {.875 to .984} INC = .107 (.028) {.048 to .165}	–	–

<b>Evidentiary Techniques/LDA Method</b> <b>Доказові методики / метод аналізу тестових даних</b>	<b>Paired Testing Techniques/TDA Method</b> <b>Методики для парного тестування / метод аналізу тестових даних</b>	<b>Investigative Techniques/TDA Method</b> <b>Дослідницькі методики / метод аналізу тестових даних</b>
<p><b>Utah ZCT CPC-RCMP Series A / Utah</b>  <b>CD = .939 (.038) { .864 to .999 }</b>  <b>INC = .185 (.041) { .104 to .266 }</b></p> <p><sup>1</sup> Емпірична система оцінювання (Empirical Scoring System).  <sup>2</sup> Можливість узагальнення цього результату-викиду обмежена тим фактом, що для цієї методики не було опубліковано показників надійності тесту. Крім того, були виявлені значні відмінності у розподілі вибірки включених наукових досліджень, що свідчить про те, що дані вибірки не є репрезентативними один для одного, або що дослідження проводилися та/або оцінювалися по-різному. Одне з досліджень включало невелику вибірку (N = 12), про яку повідомлялося у двох статтях, для яких особа, що брала участь у оцінюванні результатів, також була розробником методики. В одній із публікацій описане дослідження як пілотне без сліпого оцінювання. В обох звітах зазначено, що одного з шести правдивих учасників було виключено з дослідження після того, як він зробив неправдиве зізнання. Повідомлений рівень абсолютної точності не включав неправдиве зізнання. Ані абсолютна точність, ані показник хибних зізнань .167 наврайд чи будуть узагальнюватись у польових умовах.  <sup>3</sup> Узагальнення цього аномального результату (викиду) обмежена тим фактом, що розробники та дослідники повідомили про необхідність інтенсивного навчання, доступного лише для досвідчених практиків методики, і припустили, що складність методики перевищує ту складність, якої інші професіонали можуть навчитися у опублікованих ресурсах. Розробник повідомив про майже ідеальний коефіцієнт кореляції .99 для числового оцінювання, що свідчить про безпрецедентно високий рівень узгодження між оцінювачами, що є неочікуваним, враховуючи передбачувану складність методу. Крім того, дані, спочатку надані комітету для реплікаційних досліджень, включали лише ті випадки, щодо яких оцінювачі прийшли до правильного рішення, за винятком оцінок тих випадків, щодо яких оцінювачі не досягли правильного рішення. Відсутні оцінки пізніше були надані комітету як для досліджень Моргана та ін. (Morgan та ін. 2008), так і для Шурані та Чавеса (Shurani та Chavez, 2009). Однак отримані середні значення вибірки відрізнялися від тих, що були зареєстровані для обох реплікаційних досліджень. Через ці розбіжності статистичний аналіз не було повторно обчислено з урахуванням відсутніх оцінок і звіт відображає середні значення розподілу вибірки, які було зазначено. Вибірку для повторних досліджень слід вважати позбавленими помилок або неконтрольованої дисперсії.  <sup>4</sup> Для AFMGQT існують дві версії з незначними структурними відмінностями між ними. Немає доказів того, що продуктивність однієї версії є кращою за іншу. Оскільки для відхилення нульової гіпотези про те, що відмінності є безглуздими, потрібні повторні докази, а також оскільки вибрані дослідження включають поєднання обох версій AFMGQT, ці результати надаються як узагальнені для обох версій. Дослідження з AFMGQT використовуються як у багатоаспектних дослідженнях щодо конкретної події, так і в багатотемних скринінгових дослідженнях. І багатоаспектні, і багатотемні дослідження були інтерпретовані з правилами прийняття рішень, заснованими на припущенні про незалежність критеріїв серед релевантних запитань RQ.  <sup>5</sup> AFMGQT продемонстрував точність, яка є задовільною для парного тестування, лише якщо оцінювати його за допомогою емпіричної системи оцінювання.  <sup>6</sup> Існує дві методики, щодо яких немає опублікованих досліджень, але які структурно майже ідентичні AFMGQT: LEPET і Utah MGQT. Валідність AFMGQT можна узагальнити для цих методик, якщо оцінювати за допомогою тих самих методів оцінювання тестових даних (TDA).  <sup>7</sup> Concealed Information Test (CIT), також відомий як Guilty Knowledge Test (GKT) (GKT) і Peak of Tension Test (POT). Дані, використані тут, були надані в звіті про мета-аналіз лабораторних досліджень МакЛареном (MacLaren, 2001).  <sup>8</sup> Дослідження цих поліграфологічних методик проводилися з використанням правил прийняття рішень, заснованих на припущенні про незалежність критеріїв між тестовими цілями. Точність скринінгових методик може бути додатково покращена шляхом систематичного застосування підходу послідовних бар'єрів.</p>	<p>—</p>	<p>—</p>

<b>Таблиця 2. Середнє значення (стандартне відхилення) та {95% довірчий інтервал} для профілів критеріальної точності для всіх валідних поліграфологічних методик разом</b>		
	За виключенням результатів викиду (Excluding outlier results)	Всі включені дослідження (All included studies)
Number of PDD Techniques	12	14
Number of Studies	39	45
N Deceptive	2,067	2,336
N Truthful	1,802	2,031
Total N	3,869	4,367
Number Scorers	280	295
N of Deceptive Scores	5,840	6,109
N of Truthful Scores	5,399	5,628
Total Scores	11,239	11,737
Percent Correct	.869 (.036) {.798 to .940}	.887 (.033) {.823 to .951}
Inconclusive	.128 (.030) {.068 to .187}	.114 (.028) {.058 to .170}
Sensitivity	.812 (.056) {.702 to .923}	.835 (.051) {.734 to .936}
Specificity	.717 (.061) {.597 to .838}	.751 (.058) {.638 to .864}
FN Errors	.083 (.038) {.008 to .157}	.072 (.035) {.004 to .141}
FP Errors	.144 (.049) {.048 to .239}	.123 (.043) {.039 to .208}
D Inc	.105 (.042) {.022 to .187}	.092 (.037) {.020 to .165}
T Inc	.151 (.042) {.068 to .234}	.136 (.041) {.056 to .216}
PPV	.854 (.049) {.757 to .950}	.874 (.044) {.789 to .960}
NPV	.899 (.047) {.807 to .990}	.911 (.043) {.827 to .995}
D Correct	.909 (.042) {.826 to .992}	.921 (.039) {.844 to .997}
T Correct	.829 (.056) {.721 to .938}	.854 (.049) {.757 to .950}

У таблиці 3 наведено профіль критеріальної точності зваженої агрегації на рівні доказових (Evidentiary), для парного тестування (Paired-testing) та дослідницьких (Investigative) поліграфологічних методик відповідно до стандартів АПА 2012. Також у таблиці 3 наведено профіль критеріальної точності для доказових методик без результатів двох методик, що є викидами.

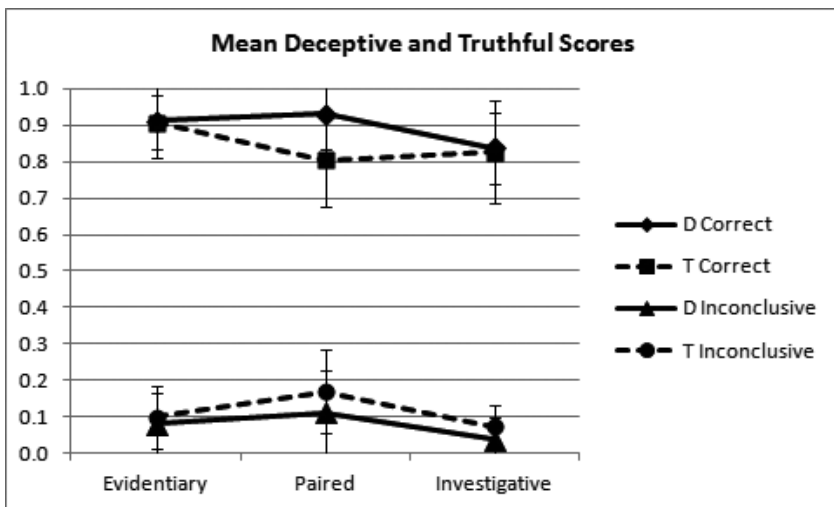
<b>Таблиця 3. Профілі критеріальної точності для доказових, парних та дослідницьких методик</b>				
	Доказові методики Evidentiary Techniques	Доказові методики без наукових досліджень з викидами Evidentiary w/o Outlier Studies	Методики для парного тестування Paired-testing Techniques	Дослідницькі методики Investigative Techniques
Number of Techniques	5	3	5	4
Number of Studies	21	15	12	12
N Deceptive	861	592	435	1,040
N Truthful	776	547	408	847
Total N	1,637	1,139	843	1,887
Number Scorers	174	159	56	65
N of Deceptive Scores	3,297	3,028	1,700	1,112
N of Truthful Scores	3,098	2,869	1,613	917
Total Scores	6,395	5,897	3,313	2,029
Unweighted Average Accuracy	.910 (.027) {.857 to .963}	.903 (.028) {.847 to .958}	.867 (.036) {.796 to .938}	.844 (.039) {.767 to .920}
Unweighted Average Inconclusives	.090 (.029) {.032 to .147}	.095 (.030) {.035 to .154}	.142 (.036) {.071 to .213}	.114 (.028) {.060 to .168}
Sensitivity	.843 (.050) {.745 to .941}	.832 (.053) {.729 to .935}	.828 (.051) {.728 to .928}	.802 (.047) {.710 to .893}
Specificity	.826 (.054) {.721 to .931}	.816 (.055) {.708 to .923}	.670 (.071) {.531 to .809}	.771 (.073) {.627 to .915}
FN Errors	.082 (.033) {.018 to .147}	.089 (.034) {.021 to .156}	.060 (.032) {.001 to .123}	.158 (.042) {.076 to .240}
FP Errors	.083 (.035) {.014 to .152}	.090 (.037) {.018 to .162}	.159 (.052) {.056 to .261}	.159 (.070) {.022 to .296}
D Inc	.080 (.038) {.005 to .155}	.086 (.041) {.004 to .167}	.112 (.043) {.028 to .195}	.038 (.020) {.001 to .077}
T Inc	.099 (.044) {.014 to .185}	.104 (.044) {.017 to .191}	.170 (.058) {.056 to .284}	.073 (.015) {.043 to .102}



	Доказові методики Evidentiary Techniques	Доказові методики без наукових досліджень з викидами Evidentiary w/o Outlier Studies	Методики для парного тестування Paired-testing Techniques	Дослідницькі методики Investigative Techniques
PPV	.915 (.034) {.848 to .982}	.908 (.037) {.836 to .979}	.847 (.050) {.749 to .945}	.860 (.037) {.788 to .933}
NPV	.904 (.042) {.823 to .986}	.898 (.043) {.814 to .982}	.920 (.046) {.829 to .999}	.812 (.082) {.651 to .973}
D Correct	.911 (.037) {.839 to .983}	.904 (.039) {.828 to .98}	.932 (.036) {.862 to .999}	.837 (.045) {.749 to .924}
T Correct	.908 (.038) {.833 to .983}	.901 (.040) {.822 to .980}	.804 (.064) {.678 to .930}	.827 (.072) {.686 to .968}

На рисунку 18 показано середні значення та статистичні довірчі інтервали для правильних рішень і невизначених результатів для трьох рівнів критеріальної валідності, описаних стандартами АПА 2012 року. Двофакторний ANOVAs, включно з викидами, показав значну взаємодію між статусом критерію та категорією валідності для правильних рішень [ $F(1,10896) = 7433.144, (p < .001)$ ] і для невизначених результатів [ $F(1,10896) = 3562.384, (p < .001)$ ], що свідчить про те, що методики на цих різних категоріальних рівнях можуть обробляти оманливі та правдиві випадки з різною ефективністю. Проте односторонній post hoc ANOVAs показав, що немає суттєвої різниці у частці правильних рішень і невизначених результатів або помилок для оманливих чи правдивих випадків.

Рисунок 18. Середні значення та довірчий інтервал для валідованих категорій АПА (APA validation categories).



### Методики для доказового тестування

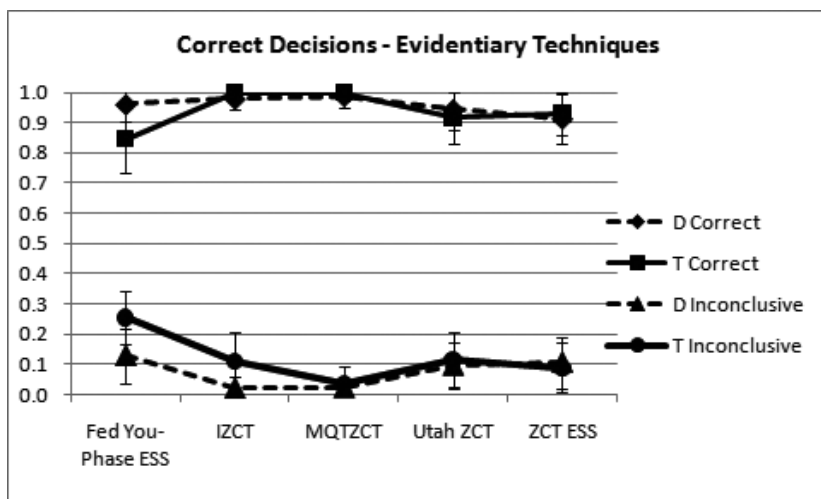
Повідомлялося, що п'ять поліграфологічних методик дають достатньо високий рівень діагностичної точності та низький рівень невизначених показників, які відповідають стандартним вимогам АПА 2012 року щодо доказового тестування. Оцінки 21 опитування та експерименту були узагальнені, щоб описати критеріальну валідність цих доказових методик. Наукові дослідження, які підтверджують ці доказові методики, включали 174 оцінювача, які надали 7407 числових оцінок для 1637 підтверджених поліграфологічних досліджень, включаючи 3821 числовий бал для 861 підтвердженого оманливого поліграфологічного дослідження та 3586 числових балів для 776 підтверджених правдивих поліграфологічних досліджень.

Таблиця 4 показує профілі критеріальної точності для п'яти поліграфологічних методик, які задовольняють вимогам АПА 2012 року щодо доказового/діагностичного тестування. Також у таблиці 4 наведено кількість включених наукових досліджень для кожної поліграфологічної методики, загальну кількість оцінених результатів, надійність, а також середнє значення та стандартні відхилення середніх оманливих і правдивих оцінок, зазначених у включених наукових дослідженнях. Середня чутливість тесту, специфічність тесту та незважена точність були зареєстровані на рівнях, які статистично значно перевищують шанс (50%) для кожного з цих п'яти поліграфологічних методик.

Двох-п'ятифакторний ANOVA, критеріальний статус х методики, для правильних рішень показав значну взаємодію між методикою та статусом випадку [ $F(1,7397) = 8944.964$ , ( $p < .001$ )], що вказує на те, що ці п'ять різних доказових методик по-різному розглядали оманливі та правдиві справи.

Post-hoc однофакторний ANOVAs показав, що відмінності в показниках правильних рішень для критеріальних правдивих випадків були значущими [ $F(4,828) = 3.118$  ( $p = .015$ )], тоді як відмінності в показниках правильних рішень для оманливих випадків не були значущими. Серія контрастів ANOVA показала, що відмінності були значущими лише для двох методик-викидів. Не було суттєвих відмінностей, якщо викиди не були включені. На рисунку 19 показано середнє значення та довірчий інтервал для правильних рішень і невизначених результатів п'яти доказових методик.

Рисунок 19. Середні значення та довірчі інтервали для доказових методик.



Таблиця 4. Профілі критеріальної точності для доказових/діагностичних поліграфологічних методик					
Technique	Federal You-Phase	IZCT*	MQTZCT*	Utah PLT (combined)	ZCT ESS
TDA Method	ESS	Horizontal	Matte	Utah	ESS
Number of Studies	2	3	3	7	6
N Deceptive	61	86	183	147	384
N Truthful	61	93	139	138	348
Total N	122	179	319	285	732
Number Scorers	11	8	7	8	140
N of Deceptive Scores	160	86	183	197	2671
N of Truthful Scores	160	93	136	188	2521
Total Scores	320	179	319	385	5192
Mean D	-7.512	-21.505	-8.711	-10.885	-10.46
StDev D	6.184	12.606	2.489	7.878	8.949
Mean T	6.146	19.626	5.226	9.372	8.219
StDev T	6.217	4.232	3.479	8.066	8.051
Reliability - Kappa	-	.†	-	.650	0.59
Reliability - Agreement	.900	.†	-	.960	.950
Reliability - Correlation	-	.†	.990†	.910	-
Unweighted Average Accuracy	.904 (.032) {.841 to .966}	.994 (.008) {.978 to .999}	.994 (.013)* {.968 to .999}	.930 (.028) {.875 to .984}	.921 (.028) {.866 to .977}
Unweighted Average Inconclusives	.192 (.033) {.127 to .256}	.033 (.019) {.001 to .069}	.029 (.015) {.001 to .058}	.107 (.030) {.048 to .165}	.098 (.030) {.039 to .157}
Sensitivity	.845 (.052) {.742 to .948}	.977 (.020) {.937 to .999}	.967 (.021) {.926 to .999}	.853 (.049) {.757 to .948}	.817 (.056) {.706 to .927}

Technique	Federal You-Phase	IZCT*	MQTZCT*	Utah PLT (combined)	ZCT ESS
Specificity	.757 (.064) {.633 to .882}	.946 (.035) {.878 to .999}	.963 (.033) {.899 to .999}	.809 (.056) {.699 to .918}	.846 (.051) {.747 to .946}
FN Errors	.034 (.026) {.001 to .085}	.012 (.015) {.001 to .041}	.011 (.021) {.001 to .052}	.051 (.031) {.001 to .112}	.077 (.037) {.004 to .151}
FP Errors	.138 (.050) {.039 to .236}	.001 (.005) {.001 to .01}	.001 (.015) {.001 to .03}	.074 (.038) {.001 to .148}	.064 (.034) {.001 to .130}
D INC	.128 (.046) {.037 to .219}	.012 (.014) {.001 to .040}	.022 (.001) {.022 to .022}	.096 (.040) {.017 to .176}	.106 (.044) {.020 to .192}
T INC	.255 (.044) {.170 to .341}	.054 (.035) {.001 to .122}	.037 (.029) {.001 to .094}	.117 (.046) {.027 to .207}	.089 (.042) {.008 to .171}
PPV	.860 (.050) {.761 to .958}	.999 (.004) {.991 to .999}	.999 (.015) {.970 to .999}	.923 (.039) {.847 to .999}	.931 (.038) {.857 to .999}
NPV	.957 (.033) {.892 to .999}	.989 (.019) {.952 to .999}	.985 (.021) {.944 to .999}	.938 (.036) {.867 to .999}	.912 (.042) {.830 to .993}
D Correct	.961 (.029) {.903 to .999}	.988 (.015) {.959 to .999}	.989 (.021) {.948 to .999}	.944 (.034) {.877 to .999}	.913 (.042) {.831 to .996}
T Correct	.846 (.056) {.736 to .956}	.999 (.005) {.989 to .999}	.999 (.015) {.969 to .999}	.916 (.043) {.832 to .999}	.929 (.037) {.857 to .999}

\* Результати-викиди, які суттєво відрізняються від нормального діапазону інших методик.

† Не було опубліковано жодних даних про надійність жодного з наукових досліджень IZCT.

‡ Коefіцієнт кореляції .990 є надзвичайним і видатним відкриттям у будь-якій галузі досліджень і свідчить про надзвичайно низький рівень розбіжностей між числовими оцінками сліпих оцінювачів, які використовують MQTZCT. Цю статистику не можна знайти в дисертаційній роботі Матте і Ройсса (Matte & Reuss, 1989) нині неіснуючого Колумбійського Тихоокеанського університету, але вона була опублікована у включеному перевиданні Матте і Ройсса (Matte & Reuss, 1989) у журналі *Polygraph*. Незважаючи на цю надзвичайно високу кореляцію числових оцінок від різних оцінювачів, розробники та дослідники MQTZCT неодноразово висловлювали застереження щодо відсутності можливості узагальнення результатів MQTZCT без інтенсивного спеціального навчання.

### **Методики для парного тестування**

П'ять методик було визначено як такі, що забезпечують достатній рівень точності, які задовольняють вимогам АПА для парного тестування<sup>48</sup>. Оцінки з 12 опитувань та експериментів були узагальнені, щоб описати валідність критеріїв цих методик для парного тестування. Наукові дослідження, які підтверджують ці методики парного тестування, включали 56 оцінювачів, які надали 3313 числових балів для 843 підтверджених поліграфологічних досліджень, включаючи 1700 числових оцінок для 435 підтверджених оманливих поліграфологічних досліджень і 1613 числових оцінок для 408 підтверджених правдивих поліграфологічних досліджень.

У таблиці 5 наведено профілі критеріальної точності для п'яти поліграфологічних методик, які задовольняють вимогам АПА 2012 року для парного тестування. У таблиці 5 також наведено кількість включених наукових досліджень для кожної поліграфологічної методики, загальну кількість оцінених результатів, надійність, а також середнє значення та стандартні відхилення середніх оманливих і правдивих оцінок із включених наукових досліджень.

Незважаючи на те, що чутливість тесту та точність незваженого рішення значно перевищують шанс (випадковість) для всіх п'яти методик для парного тестування, три з цих методик дали рівні специфічності тесту, які не були значно вищими за шанс (Backster You-Phase/Backster, Federal You-Phase/7-позиційна система оцінювання, та Federal ZCT/7-позиційна система оцінювання).

На рисунку 20 показано середнє значення та довірчий інтервал для правильних рішень та невизначених результатів п'яти методик для парного тестування. Двох-п'ятифакторний ANOVA, критеріальний статус  $x$  методики, для правильних рішень показав значний взаємозв'язок [ $F(1,3303) = 5891.333, (p < .001)$ ], вказуючи, що ці п'ять методик для парного тестування демонструють різні показники правильних рішень для випадків обману і правди. Post-hoc однофакторний ANOVAs не показали суттєвих відмінностей у показниках правильних рішень для критеріїв випадків обману або критеріїв випадків правди.

Двох-п'ятифакторний ANOVA, статус критерію  $x$  методик, для невизначених результатів показав значну взаємодію [ $F(1,3303) = 5891.333, (p < .001)$ ], що вказує на те, що ці п'ять методик для парного тестування дають різні показники невизначених результатів для випадків обману і правди. Post-hoc однофакторні ANOVA показали, що відмінності в невизначених показниках не були значущими для критеріїв випадків обману або правди.

---

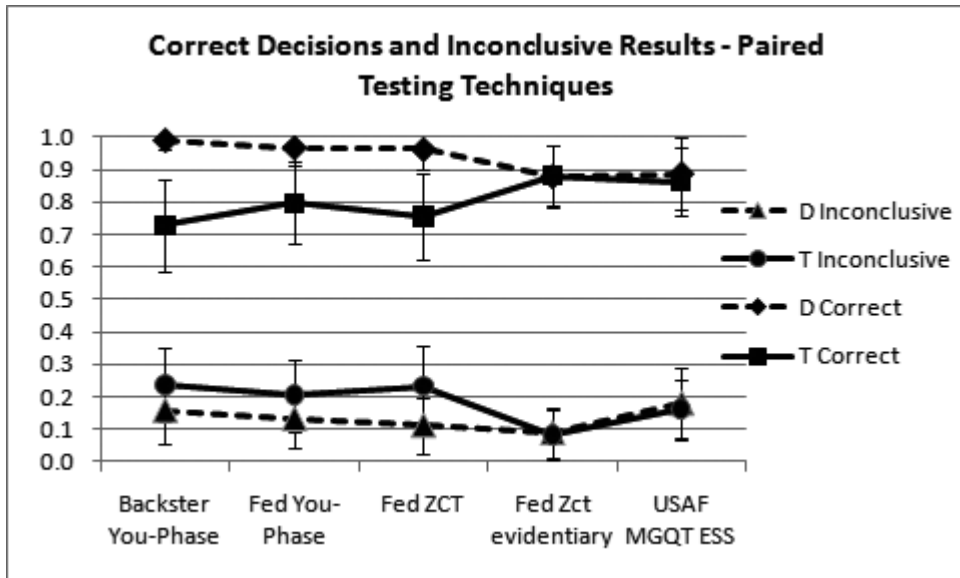
<sup>48</sup> Усі поліграфологічні методики, які відповідають стандартним вимогам АПА 2012 року щодо доказового тестування, також відповідають вимогам щодо парного та дослідницького тестування. Ті поліграфологічні методики, які відповідають вимогам щодо критеріальної точності для парного тестування, також є дійсними для дослідницького тестування.

**Таблиця 5. Профілі критеріальної точності методик для парного тестування**

<b>Technique</b>	<b>Backster You-Phase</b>	<b>Federal You-Phase</b>	<b>Federal ZCT</b>	<b>Federal ZCT</b>	<b>AFMGQT</b>
TDA Method	Backster	7-position	7-position	7-position evidentiary	ESS
Number of Studies	2	2	3	2	3
N Deceptive	61	61	139	80	94
N Truthful	61	61	109	80	97
Total N	122	122	248	160	191
Number Scorers	8	11	16	16	5
N of Deceptive Scores	127	160	767	530	116
N of Truthful Scores	127	160	677	530	119
Total Scores	254	320	1,444	1,060	235
Mean D	-16.055	-7.195	-8.577	-8.263	-2.960
StDev D	7.417	5.824	9.018	9.032	4.765
Mean T	5.216	5.999	7.466	7.852	3.738
StDev T	10.291	5.893	8.472	9.721	4.104
Reliability - Kappa	-	-	.570	-	-
Reliability - Agreement	-	.850	.800	.870	1
Reliability - Correlation	.567	-	-	-	.930
Unweighted Average Accuracy	.862 (.037) {.787 to .932}	.883 (.035) {.813 to .952}	.860 (.037) {.788 to .931}	.880 (.034) {.813 to .948}	.875 (.039) {.798 to .953}
Unweighted Average Inconclusives	.196 (.040) {.117 to .275}	.168 (.037) {.096 to .241}	.171 (.040) {.093 to .249}	.085 (.029) {.028 to .141}	.170 (.036) {.100 to .241}

Technique	Backster You-Phase	Federal You-Phase	Federal ZCT	Federal ZCT	AFMGQT
Sensitivity	.836 (.052) {.734 to .938}	.841 (.050) {.742 to .939}	.858 (.051) {.759 to .957}	.804 (.054) {.697 to .911}	.729 (.065) {.603 to .856}
Specificity	.556 (.070) {.418 to .694}	.632 (.069) {.497 to .768}	.581 (.073) {.438 to .723}	.809 (.057) {.698 to .920}	.700 (.063) {.577 to .823}
FN Errors	.007 (.012) {-.016 to .03}	.028 (.023) {.001 to .073}	.033 (.029) {.001 to .090}	.110 (.044) {.024 to .197}	.092 (.046) {.002 to .182}
FP Errors	.207 (.058) {.091 to .322}	.161 (.051) {.061 to .261}	.188 (.051) {.089 to .287}	.109 (.044) {.022 to .196}	.112 (.047) {.02 to .204}
D INC	.156 (.051) {.055 to .257}	.131 (.046) {.041 to .221}	.110 (.044) {.023 to .196}	.087 (.039) {.010 to .163}	.178 (.056) {.068 to .289}
T INC	.236 (.059) {.119 to .354}	.205 (.057) {.093 to .318}	.232 (.064) {.106 to .358}	.083 (.040) {.003 to .162}	.162 (.047) {.071 to .254}
PPV	.801 (.055) {.693 to .909}	.840 (.053) {.736 to .943}	.838 (.053) {.734 to .943}	.880 (.048) {.786 to .974}	.864 (.058) {.751 to .977}
NPV	.987 (.021) {.945 to .999}	.958 (.035) {.889 to .999}	.940 (.046) {.851 to .999}	.880 (.048) {.786 to .974}	.887 (.052) {.785 to .989}
D Correct	.991 (.014) {.963 to .999}	.968 (.027) {.916 to .999}	.963 (.033) {.898 to .999}	.879 (.048) {.786 to .973}	.888 (.057) {.777 to .999}
T Correct	.728 (.073) {.584 to .873}	.797 (.064) {.672 to .923}	.756 (.067) {.625 to .887}	.881 (.048) {.786 to .976}	.862 (.053) {.758 to .967}

Рисунок 20. Середнє значення та довірчі інтервали методик для парного тестування.



#### Дослідницькі методики тестування.

Чотири поліграфологічні методики утворили критеріальну точність, яка задовольняє вимогам АПА для дослідницького тестування. Оцінки з 12 опитувань та експериментів були узагальнені, щоб описати критеріальну валідність цих дослідницьких методик<sup>49</sup>. Наукові дослідження, які підтверджують ці дослідницькі методики, включали 65 оцінювачів, які надали 2029 числових оцінок для 1887 підтверджених поліграфологічних досліджень, включаючи 1112 числових оцінок для 1040 підтверджених випадків обману і 917 числових оцінок для 847 підтверджених випадків правди.

Таблиця 6 показує профілі критеріальної точності для чотирьох поліграфологічних методик, які задовольняють вимогам АПА 2012 року для дослідницького тестування. У таблиці 6 також наведено кількість включених наукових досліджень для кожної поліграфологічної методики, загальну кількість оцінених результатів, надійність, а також середнє значення та стандартні відхилення середніх оманливих і правдивих оцінок включених наукових досліджень. Повідомляється, що незважаючи на точність рішень і чутливість тесту значно перевищують шанс для всіх чотирьох дослідницьких методик. Три з цих дослідницьких методик, СІТ, а також DLST/ТЕС, оцінені як за семипозиційною моделлю оцінювання, так і моделлю ESS, показали специфічність тесту, яка була значно більшою, ніж шанс (випадковість). Специфічність тесту для однієї дослідницької методики, АFMGQT, оціненою семипозиційною моделлю, не була значно більшою, ніж шанс (випадковість).

<sup>49</sup> Одним із включених наукових досліджень був мета-аналіз, який підсумовував результати лабораторних досліджень з використанням СІТ.



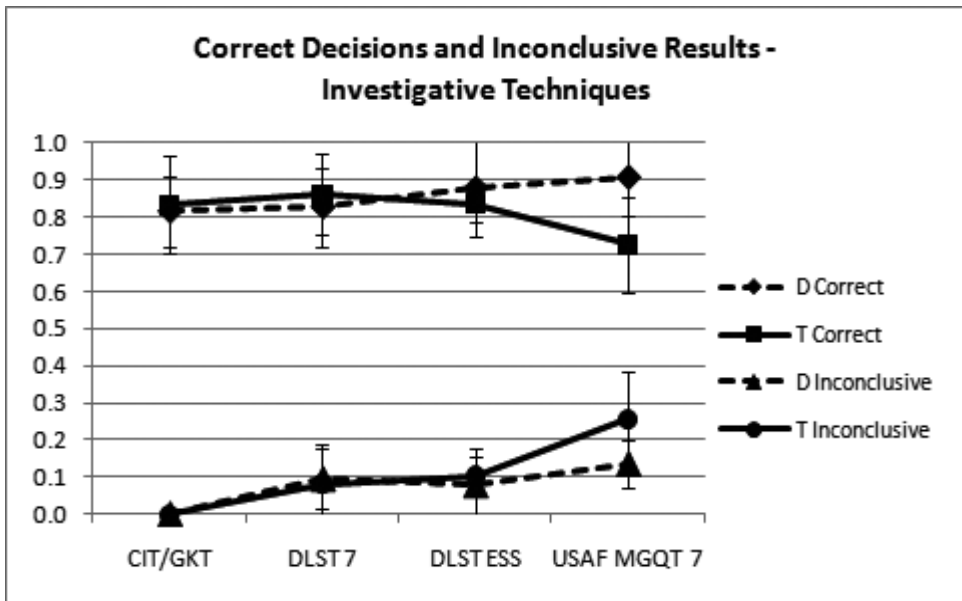
**Таблиця 6. Критеріальна точність дослідницьких методик**

Technique	CIT/GKT	DLST/TES	DLST/TES	AFMGQT
TDA Method	Lykken	7-position	ESS	7-position
Number of Studies	39	4	4	3
N Deceptive	666	131	149	94
N Truthful	404	197	149	97
Total N	1070	328	298	191
Number Scorers	39	16	5	5
N of Deceptive Scores	666	156	174	116
N of Truthful Scores	404	221	173	119
Total Scores	1070	377	347	235
Mean D	-	-2.126	-2.131	-2.607
StDev D	-	3.959	3.801	4.754
Mean T	-	3.162	3.412	3.114
StDev T	-	3.531	3.153	3.705
Reliability - Kappa	-	.760	-	.750
Reliability - Agreement	-	.806	.840	.965
Reliability - Correlation	-	-	-	.940
Unweighted Average Accuracy	.823 (.041) {.744 to .903}	.844 (.039) {.768 to .920}	.858 (.037) {.786 to .930}	.817 (.042) {.734 to .900}
Unweighted Average Inconclusives	.001 (.001) {.001 to .001}	.088 (.028) {.034 to .142}	.090 (.026) {.039 to .142}	.197 (.030) {.138 to .255}
Sensitivity	.815 (.048) {.721 to .910}	.748 (.062) {.626 to .869}	.809 (.069) {.674 to .945}	.783 (.058) {.669 to .896}
Specificity	.832 (.067) {.700 to .963}	.792 (.060) {.674 to .909}	.751 (.031) {.691 to .811}	.538 (.068) {.405 to .672}
FN Errors	.185 (.048) {.090 to .279}	.156 (.050) {.058 to .255}	.112 (.057) {.001 to .224}	.079 (.050) {.001 to .177}
FP Errors	.168 (.067) {.037 to .300}	.127 (.052) {.026 to .229}	.146 (.027) {.093 to .2}	.203 (.057) {.090 to .315}
D INC	.001 (.001) {.001 to .001}	.096 (.041) {.016 to .175}	.078 (.052) {.001 to .180}	.137 (.033) {.071 to .202}
T INC	.001 (.001) {.001 to .001}	.081 (.037) {.008 to .153}	.102 (.014) {.075 to .130}	.257 (.049) {.160 to .354}
PPV	.889 (.037) {.816 to .961}	.806 (.055) {.698 to .914}	.848 (.041) {.767 to .928}	.79 (.059) {.675 to .905}

Technique	CIT/GKT	DLST/TES	DLST/ES	AFMGQT
NPV	.732 (.076) {.583 to .881}	.878 (.054) {.772 to .983}	.870 (.052) {.768 to .971}	.874 (.062) {.753 to .996}
D Correct	.815 (.048) {.721 to .910}	.827 (.055) {.719 to .935}	.878 (.067) {.746 to .999}	.908 (.053) {.804 to .999}
T Correct	.832 (.067) {.700 to .963}	.861 (.055) {.753 to .969}	.837 (.027) {.783 to .891}	.726 (.066) {.597 to .856}

На рисунку 21 показано середні значення та довірчий інтервал для правильних рішень і невизначених результатів методик для дослідницького тестування. Двох-п'ятифакторний ANOVA, статусу критерію x методики, для правильних рішень показав значну взаємодію [F (1,2021) = 1320.745, (p < .001)], що вказує на те, що ці чотири дослідницькі методики відрізняються за своїми можливостями правильно класифікувати випадки обману та правди. Однак post-hoc однофакторні ANOVAs не показали суттєвих відмінностей у показнику правильних рішень для критеріїв випадків обману або правди.

Рисунок 21. Середні значення та довірчі інтервали дослідницьких методик.



Двох-п'ятифакторний ANOVA, статусу критерію x методики, для невизначених результатів показав значну взаємодію [F (1,953) = 404.177, (p < .001)], що вказує на те, що ці три дослідницькі методики дають різні показники невизначених результатів для випадків обману та правди. Post-hoc однофакторні ANOVAs показали, що відмінності в невизначених показниках не були значущими для випадків обману, але були значущими для випадків правди [F (2,478) = 3.418, (p = .034)]. Результати CIT/GKT не включають невизначену категорію, і ця методика не була включена в двофакторний аналіз для невизначених результатів.

### **Незалежні та не-незалежні (non-independent) поліграфологічні методики**

Таблиця 7 показує профіль критеріальної точності чотирьох поліграфологічних методик, які інтерпретуються з правилами прийняття рішень на основі припущення про незалежну дисперсію критерію релевантних запитань, разом із профілем критеріальної точності поліграфологічних методик, які інтерпретуються з правилами прийняття рішень на основі припущення про відсутність незалежності. Оцінки були узагальнені з 14 оглядів та експериментів із застосуванням поліграфологічних методик, які інтерпретуються з припущенням незалежного критерію дисперсії серед релевантних запитань. Ці наукові дослідження включали 31 оцінювача, які надали 1194 числові оцінки для 1008 підтверджених поліграфологічних досліджень, які включали 562 числові оцінки для 468 підтверджених оманливих поліграфологічних досліджень і 632 числові оцінки для 540 підтверджених правдивих поліграфологічних досліджень. За винятком результатів викидів, оцінки з 24 оглядів та експериментів були узагальнені для опису валідності критеріїв поліграфологічних методик, для яких результати інтерпретуються за допомогою правил прийняття рішень на основі припущення про незалежність дисперсії критерію для релевантних запитань. Ці наукові дослідження включали 210 оцінювачів, які надали 8975 числових оцінок для 1791 підтвердженого поліграфологічного дослідження, включаючи 4612 числових оцінок для 933 підтверджених оманливих поліграфологічних досліджень і 4363 числових оцінок для 858 підтверджених правдивих поліграфологічних досліджень.

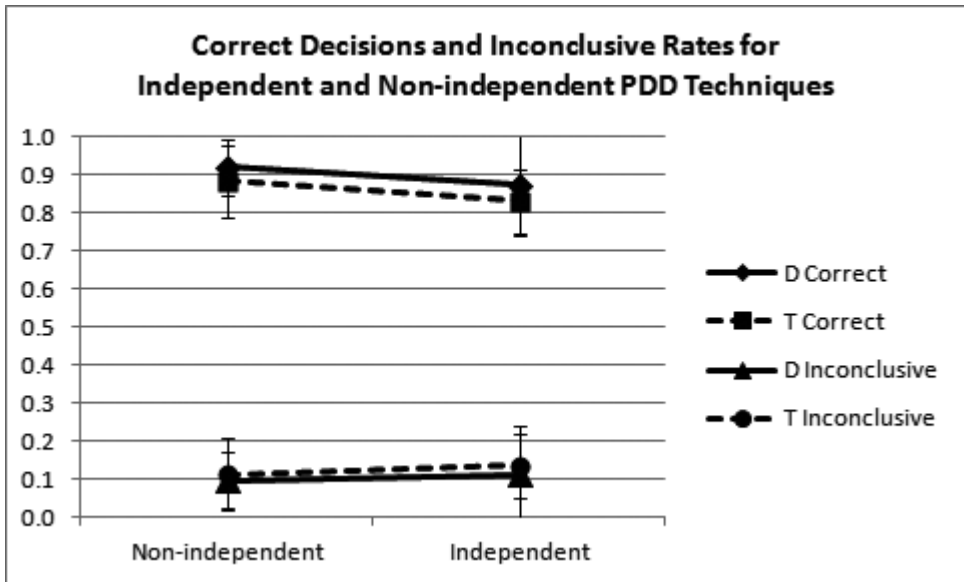
За винятком результатів викидів, методики із запитаннями порівняння, призначені для діагностичного тестування на конкретну подію (event specific) (single-issue, однотемного одноаспектного), у яких вважається, що дисперсія критерію кількох релевантних запитань не є незалежною, дали загальний показник точності рішення .890 (.829 – .951), із загальним невизначеним показником .110 (.047 – .173). Поліграфологічні методики із запитаннями порівняння, розроблені для інтерпретації з припущенням незалежності дисперсії критерію кількох релевантних запитань, дали загальний показник точності рішень .850 (.773 – .926) із загальним невизначеним показником .125 (.068 – .183). Незважене середнє значення точності для незалежних і не-незалежних поліграфологічних методик, за винятком результатів викидів, дало рівень точності рішення .869 (.798 – .940) з невизначеним показником .128 (.068 – .187), як показано у таблиці 7.

На рисунку 22 показано середнє значення та статистичні довірчі інтервали для правильних рішень і невизначених показників для поліграфологічних методик, інтерпретованих за допомогою правил прийняття рішень, заснованих на припущеннях дисперсії незалежного та не-незалежного критерію, за винятком результатів викидів. Двофакторні ANOVAs, незалежність стану критерію  $x$ , за винятком викидів, показав значну взаємодію для правильних рішень [ $F(1,10165) = 2656.637$ , ( $p < .001$ )] і невизначених результатів [ $F(1,10165) = 806.839$ , ( $p < .001$ )]. Проте post-hoc ANOVAs показали відсутність суттєвої однофакторної різниці у частці правильних рішень або невизначених результатів для критеріальних оманливих випадків або критеріальних правдивих випадків.

**Таблиця 7. Профіль критеріальної точності незалежних і не-незалежних поліграфологічних методик**

	<b>Критерій незалежних поліграфологічних методик Criterion Independent PDD Techniques</b>	<b>Критерій не-незалежних поліграфологічних методик Non-independent PDD Techniques</b>	<b>Критерій не-незалежних поліграфологічних методик з результатами викидами Non-independent Techniques with Outlier Results</b>
Number of Techniques	4	7	9
Number of Studies	14	24	30
N Deceptive	468	933	1,202
N Truthful	540	858	1,087
Total N	1,008	1,791	2,289
Number Scorers	31	210	225
N of Deceptive Scores	562	4,612	4,881
N of Truthful Scores	632	4,363	4,592
Total Scores	1,194	8,975	9,473
Percent Correct	.850 (.039) {.773 to .926}	.890 (.031) {.829 to .951}	.896 (.030) {.837 to .955}
Inconclusive	.125 (.029) {.068 to .183}	.110 (.032) {.047 to .173}	.106 (.031) {.044 to .167}
Sensitivity	.771 (.072) {.630 to .911}	.833 (.052) {.731 to .934}	.840 (.050) {.743 to .938}
Specificity	.719 (.047) {.626 to .811}	.765 (.061) {.646 to .884}	.775 (.059) {.658 to .891}
FN Errors	.113 (.058) {.001 to .226}	.078 (.033) {.013 to .143}	.074 (.032) {.011 to .138}
FP Errors	.144 (.039) {.066 to .221}	.115 (.042) {.032 to .197}	.109 (.041) {.029 to .189}
D Inc	.112 (.051) {.013 to .212}	.093 (.041) {.012 to .174}	.089 (.039) {.011 to .166}
T Inc	.136 (.031) {.076 to .196}	.127 (.049) {.030 to .223}	.122 (.049) {.027 to .218}
PPV	.828 (.059) {.712 to .943}	.886 (.041) {.806 to .967}	.893 (.039) {.816 to .969}
NPV	.878 (.049) {.782 to .973}	.906 (.044) {.820 to .993}	.910 (.043) {.826 to .995}
D Correct	.873 (.066) {.744 to .999}	.915 (.037) {.842 to .988}	.919 (.036) {.849 to .989}
T Correct	.831 (.043) {.746 to .915}	.866 (.049) {.770 to .962}	.873 (.047) {.780 to .965}

Рисунок 22. Середнє значення та довірчий інтервал критеріїв поліграфологічних методик, незалежних і не-незалежних (за винятком результатів викидів).



### Обговорення

Чотирнадцять поліграфологічних методик (показані на рисунку 17) відповідають вимогам стандартів АПА 2012 року для валідації тестування. Ці методики підтримуються опублікованими описами протоколу проведення тестування та аналізу тестових даних із використанням інструментарію, що є репрезентативним до того, що використовується у польовій практиці, а також методики та метод аналізу тестових даних підтримано публікаціями емпірично щодо критеріальної точності.

Щодо п'яти поліграфологічних методик є опубліковані докази валідності, які відповідають вимогам АПА 2012 року щодо доказового тестування, включаючи незважену точність рішень понад .900 разом із невизначеним показником, меншим .200. Ці п'ять поліграфологічних методик розташовані в алфавітному порядку; Federal You-Phase, оцінена за допомогою емпіричної системи оцінювання (ESS), IZCT, MQTZCT, Utah ZCT (включаючи варіанти PLT, DLT і CPC-RCMP), оцінена за допомогою числової системи оцінювання штату Юта, і будь-який варіант ZCT з трьома релевантними запитаннями, що стосуються конкретної події, оцінені за допомогою ESS. Статистичний аналіз виявив два статистичні викиди, IZCT та MQTZCT. Двофакторні ANOVAs показали, що не було суттєвих відмінностей між іншими доказовими методиками, коли результати викиди не були включені.

Також встановлено, що п'ять інших поліграфологічних методик забезпечують критеріальну точність, яка відповідає вимогам стандарту АПА 2012 року для парного тестування, з незваженою точністю рішення понад .860 разом із невизначеними показниками нижче .200. Ці поліграфологічні методики в алфавітному порядку: методика AFMGQT, оцінена за допомогою ESS, методика Backster You-Phase, оцінена за допомогою числової системи оцінювання Backster, методика Federal You-Phase, оцінена за допомогою федеральної семипозиційної моделі TDA, методика Federal ZCT, оцінена за федеральною семипозиційною моделлю TDA, і методика Federal ZCT, оці-

нена за семипозиційною моделлю TDA та інтерпретована за допомогою доказових правил прийняття рішень. Незважаючи на те, що цей рівень валідації призначений для задоволення потреб у критеріальній точності в ситуаціях парного тестування, більшість поліграфологічних досліджень не призначені для парного тестування чи використання як докази у залі суду. Тому неминуче, що багато польових досліджень можуть проводитися з використанням поліграфологічних методик з цього списку, хоча вони не призначені для використання в залі судових засідань. Незважаючи на те, що спостерігалася значна взаємодія між критеріальним статусом та поліграфологічними методиками, що вказує на те, що різні поліграфологічні методики можуть забезпечувати незначні відмінності в критеріальній точності з випадками обману та правди, жоден із однофакторних основних ефектів не був значущим для точності рішень або невизначених результатів серед випадків правди та обману. Нинішні докази не підтверджують висновок про те, що будь-який із цих методів забезпечує точність, яка відрізняється від інших методів, і натомість припускають, що ця група поліграфологічних методів забезпечує загальну точність критерію подібної ефективності.

Було встановлено, що чотири додаткові поліграфологічні методики задовольняють вимогам стандартів АПА 2012 року для дослідницького тестування. Ці чотири поліграфологічні методики в алфавітному порядку: CIT/GKT; DLST/TES, оцінені за допомогою семипозиційного методу TDA; DLST/TES, оцінені за допомогою ESS, і AFMGQT<sup>50</sup>, оцінені за допомогою семипозиційного методу TDA.

Хоча можуть існувати незначні відмінності в точності цих методик у критеріях випадків обману та правди, не було суттєвих відмінностей основного ефекту щодо точності рішень або невизначених результатів серед випадків обману чи правди. Ці результати свідчать про те, що ця група поліграфологічних методик забезпечує загальну критеріальну точність подібної ефективності, і наявні дані не підтверджують висновок про те, що будь-яка із цих методик має точність, відмінну від інших методик.

### **Результати викиди**

Було виявлено два викиди: IZCT і MQTZCT. Дослідження обох цих методик показали майже ідеальну точність, і було виявлено, що ці результати є статистичними викидами щодо розподілу результатів, передбачених усіма іншими дослідженнями щодо всіх інших методик, включаючи інші доказові методики, у які згруповано ці два наукові дослідження. Ці дві методики ґрунтуються на найбільш проблематичних дослідженнях з усіх наявних досліджень, включених до мета-аналізу.

Одне з двох наукових досліджень, включених на підтримку IZCT (Gordon et al., 2005; також описане Mohamed et al., 2006), є дуже невеликим дослідженням, в одній публікації як несліпе (*прим. перекл.: тобто, коли сліпого оцінювання незалежними оцінювачами не проводилось*) пілотне наукове дослідження. Використання пілотних досліджень для відповідей на питання про критеріальну точність викликає труднощі. Крім того, в обох звітах зазначено, що один із 12 учасників наукового дослідження (Gordon et al. 2005), запрограмований невинний учасник, зробив неправдиве зізнання поліграфологу, також головному автору, під час передтестового інтерв'ю. Цього учасника було вилучено з експерименту, що пояснюється *несліпим*

---

<sup>50</sup> Оскільки Utah MGQT та LEPET структурно практично ідентичні з AFMGQT і використовують той самий підхід в оцінюванні, доцільно узагальнити результати валідазації AFMGQT і на ці дві методики.

(без сліпого оцінювання) задумом дослідження. Неправдиве зізнання в польових поліграфологічних програмах не можна відразу відрізнити від автентичного зізнання. У польових поліграфологічних програмах передтестове зізнання розглядається як практична та успішна форма вирішення розслідуваної справи. Достовірні зізнання вважаються успіхом в поліграфології, і тому необхідно розглядати неправдиві зізнання як проблеми. У польовій ситуації лише пізніше, коли будуть доступні додаткові докази, зізнання буде визначено як помилка та розглядатиметься як проблематичне.

Включення цієї помилки в результати дослідження призвело б до хибно-позитивного результату (тобто хибного зізнання) на рівні .167 і менш ніж ідеальної точності тесту. Натомість результати наукового дослідження Гордона та ін. (Gordon et al. 2005) було надано без неправдивих зізнань разом із заявленим показником точності рішення 1.000. Цілком можливо, що ані зареєстрований показник точності рішення 1.000, ані показник помилкових зізнань .167 не є репрезентативними для продуктивності IZCT у польових умовах. Можна запропонувати аргумент, що, оскільки це було несліпе пілотне наукове дослідження, яке не було розроблено, щоб служити дослідженням критеріальної точності, видалення помилок із повідомленого результату дослідження було виправданим. Подібні пілотні дослідження допомагають приймати рішення щодо фінансування та планування більш ретельних досліджень у таких сферах, як fMRI або інших методах виявлення обману. Однак вибіркоче виключення несприятливих даних із дослідження критеріальної точності вимагає серйозного обґрунтування.

Додатковим занепокоєнням щодо доказів, що підтверджують IZCT, є той факт, що розподіли вибірки з трьох включених досліджень суттєво відрізняються. Значні відмінності є результатом кількох можливих умов, у тому числі: 1) зразки були відібрані з різних популяцій, 2) IZCT проводили по-різному для різних досліджуваних зразків, або 3) досліджувані зразки оцінювали та інтерпретували з застосуванням різних правил TDA. Також можливо, що спостережувані суттєві відмінності є результатом високоселективної методології вибірки, до якої включено дослідження на основі думки поліграфолога або дослідників щодо хороших або впевнених результатів, як, наприклад, у контексті прямого визнання щодо дослідницьких цілей. Надмірна залежність від підтвердження зізнання може призвести до систематичного виключення як хибнонегативних, так і хибнопозитивних випадків помилки, для яких зізнання, ймовірно, не буде отримано.

Незалежно від причини, оцінки обману та правди були виражені суттєво різними способами в трьох різних наукових дослідженнях щодо IZCT. Значення цих суттєвих відмінностей у цьому мета-аналізі полягає в тому, що включені дослідження, здається, базуються на вибірках, які не є репрезентативними одне для одного, і невідомо, чи одне чи кілька наукових досліджень є не репрезентативними для популяції досліджуваних.

Третє проблемне занепокоєння щодо IZCT полягає в тому, що жодне з опублікованих наукових досліджень не включало жодної статистики надійності, а розрахунки надійності між оцінювачами не могли бути завершені на основі доступних даних. Відсутність статистики надійності не дозволяє оцінити можливість узагальнення результатів наукового дослідження як результати, які були б отримані від інших поліграфологів або інших оцінювачів. У поєднанні зі значущими ефектами взаємодії між розподілом вибірки та статусом випадку, наявних доказів недостатньо, щоб підтвер-

дити уявлення про те, що інші практики отримують оцінки або результати, подібні до тих, що повідомляються в опублікованих наукових дослідженнях.

Наукові дослідження, що підтверджують можливість узагальнення MQTZCT, іншого статистичного викиду, обмежені деякими цікавими та унікальними факторами. По-перше, розробник MQTZCT і самі попередні автори, застерігали щодо складності цієї методики, наголошуючи на необхідності інтенсивного та спеціалізованого навчання, доступного лише для практиків цього методу. Дійсно, вони стверджували, що складність методики та пов'язаних з нею психологічних гіпотез є такими, що інші навчені поліграфологи не повинні очікувати, що навчаться або належним чином виконають MQTZCT на основі інформації, доступної в опублікованих джерелах. Акцент на суворому дотриманні складної та точної системи багатьох правил створює враження, що методику слід розглядати як крихку та ненадійну, яку легко порушити навіть незначними відхиленнями від передбачених процедур.

Друге, не менш важливе занепокоєння, пов'язане з тим, що було виявлено значну взаємодію між розподілом вибірки та статусом випадку. Незважаючи на те, що однофакторні відмінності не були значущими в межах оманливих чи правдивих груп, значний ефект взаємодії вказує на те, що бали критеріальних оманливих і критеріальних правдивих випадків виражаються або інтерпретуються по-різному в розподілі вибірки трьох включених досліджень MQTZCT. Іншими словами, дані не збігаються навіть серед наукових досліджень, які використовуються для підтримки MQTZCT. Ця значуща взаємодія передбачає можливість того, що включені наукові дослідження базуються на вибірках, які не є репрезентативними одна для одної. Невідомо, чи одне або кілька наукових досліджень є не репрезентативними для популяції всіх досліджуваних, що зменшує нашу впевненість у потенціалі узагальнення повідомлених результатів.

Третє занепокоєння, пов'язане з MQTZCT, полягає в тому, що зареєстрований коефіцієнт надійності .990 був опублікований у переддруку дисертації Матте та Ройса (Matte & Reuss 1989), опублікованого в журналі *Polygraph*, але його неможливо знайти в оригінальному дослідженні дисертації Колумбійського Тихоокеанського університету, якого вже не існує. Це водночас і прикро, і викликає занепокоєння, оскільки безпрецедентно високий рівень узгодження між оцінками є неочікуваним, враховуючи передбачувану складність методу.

Останнім недоліком узагальнення результатів включених наукових досліджень щодо MQTZCT є те, що дані, надані комітету, спочатку включали числові оцінки лише для тих випадків, для яких оцінювачі досягли правильного результату. Дані, доступні спеціальному комітету, спочатку не включали числові оцінки випадків, в яких оцінювачі досягли помилкових або невизначених результатів. Відсутні оцінки пізніше були надані комітету як для наукових досліджень Мангана, Армітіджа та Адамса (Mangan, Armitage & Adams, 2008), так і Шурані, Стейна та Брендта (Shurani, Stein & Brand, 2009). Однак отримані середні значення розподілу вибірки, розраховані з відсутніми оцінками, відрізнялися від тих, що були надані для обох наукових досліджень. Через ці розбіжності статистичний аналіз не було перераховано з відсутніми оцінками, і цей аналіз відображає середні бали, як повідомили Манган, Армітідж і Адамс (Mangan, Armitage & Adams, 2008), а також Шурані та Чавес (Shurani & Chavez, 2009). Польові дані завжди являють собою комбінацію діагностичної (тобто контрольованої або поясненої) дисперсії та дисперсії помилки (тобто неконтрольованої або непоясненої дисперсії). Середні значення вибірки, зазначені в дослі-



дженнях Мангана, Армітіджа та Адамса (Mangan, Armitage & Adams, 2008) і Шурані, Стейна та Бренда (Shurani, Stein & Brand, 2009), систематично позбавлені дисперсії помилок. Враховуючи, що спостерігався значущий ефект взаємодії між розподілом вибірки та статусом випадку, наявних доказів недостатньо, щоб підтвердити можливість узагальнення наведених результатів дослідження.

Можливі посередники цих результатів викидів включають можливість того, що ці методики просто перевершують інші. Однак не можна ігнорувати роль майнових, особистих і фінансових інтересів, включаючи ділові відносини між розробниками методики та головними дослідниками, а серйозна методологічна та емпірична плутанина навколо підтверджуючих досліджень підривають довіру до результатів наукового дослідження та повідомленої точності цих методик. З наукової точки зору, навіть добре сплановане дослідження, проведене прихильниками методу, які мають особисту зацікавленість у результатах і які діють як учасники та автори звіту про наукове дослідження, не має переконливої сили дослідження, яке не настільки обтяжене цими потенційними компрометуючими факторами<sup>51</sup>.

Незалежно від того, які фактори сприяють отриманню цих виняткових результатів, плутанина, пов'язана з супровідними дослідженнями, підриває впевненість у тому, що вони представляють справжню точність. Очікування того, що ці викиди будуть узагальнюватись для польових умов, слід відкласти до завершення більш повних незалежних досліджень реплікації та розширеного аналізу.

### **Критеріальна точність**

Без урахування викидів, сукупна незважена точність<sup>52</sup> усіх поліграфологічних методик становила .869 (.798 – .940), з незваженим показником .128 (.068 – .187). Усі 14 поліграфологічних методик, включені до цього мета-аналізу, дали незважені рівні точності рішень, які були значно вищими за шанс (випадковість). За винятком викидів, не було значних однофакторних відмінностей у незваженій точності рішень будь-якої з 14 поліграфологічних методик, а також значущих однофакторних відмінностей у правильних рішеннях, невизначених результатах або помилках для випадків оманливого або правдивого критерію. Також не було жодних суттєвих відмінностей у сукупній критеріальній точності поліграфологічних методик для доказового, парного та дослідницького тестування. Деякі значущі відмінності спостерігалися в профілях критеріальної точності цих методик. Усі п'ять методик, включених для доказових тестувань дали статистично значущі розміри ефекту тесту щодо чутливості до обману, так і для тесту щодо специфічності щодо правдивості.

На рівні парного тестування всі п'ять методик також продемонстрували чутливість тесту до обману, яка була значно більшою, ніж випадковість, хоча лише дві з цих методик, Federal ZCT, що оцінювався за семипозиційними доказовими правилами, і AFMGQT, що оцінювався за допомогою ESS, продемонстрували специфічність

---

<sup>51</sup> Можуть виникнути запитання щодо того, чому ці наукові дослідження та методики були включені в мета-аналіз після виявлення такої кількості серйозних недоліків. Зрештою методики були включені в мета-аналіз, оскільки вони відповідали більш загальним вимогам, викладеним у Стандартах практики АПА. Також було визначено, що мета-аналіз буде більш повним, а отже більш корисним та інформативним для зацікавлених читачів, якщо включити ці дослідження та методики.

<sup>52</sup> Незважене середнє значення вважалося більш консервативним і реалістичним розрахунком загальної точності всіх поліграфологічних методик. Розрахунок середньозваженої величини або простої частки правильних рішень часто призводить до вищих статистичних результатів, які є менш стійкими до різниці у базових ставках і, отже, менш узагальненими.

для правди, яка була значно більшою, ніж випадковість. Специфічність щодо правди не була суттєво більшою, ніж шанс, для методики Backster You-Phase, яка оцінювалась з системою оцінювання Backster, Federal ZCT, яка оцінювалась з семипозиційною системою оцінювання, або Federal You-Phase, яка оцінювалась з семипозиційною системою оцінювання.

Для дослідницьких методик всі чотири методики продемонстрували чутливість тесту до обману, яка була значно більшою, ніж шанс/випадковість. Специфічність щодо правдивості була значно вищою, ніж шанс для CIT і формату DLST/TES, який оцінювався як за семипозиційною моделлю, так і за моделлю ESS, але не була значно вищою за шанс для AFMGQT, коли оцінювалась за семипозиційною моделлю.

Виключення результатів викидів, опублікованих і відтворених емпіричних доказів для семи форматів CQT, призначених для діагностичного тестування конкретної події, для яких результати інтерпретуються з використанням правил прийняття рішень на основі припущення про відсутність незалежності (не-незалежність) дисперсії критерію релевантних запитань, дали сукупну незважену точність .890 (.829 – .951) разом із невизначеним показником .110 (.047 – .173). Ці методики в алфавітному порядку: методика AFMGQT, оцінена за допомогою ESS, методика Backster You-Phase, оцінена за допомогою числової системи оцінювання Backster, методика Federal You-Phase, оцінена за допомогою федеральної семипозиційної моделі TDA, методика Federal You-Phase, оцінена за допомогою ESS, методика Federal ZCT, оцінена за допомогою федеральної семипозиційної моделі TDA, методика Federal ZCT, оцінена за допомогою семипозиційної моделі TDA та інтерпретована за допомогою доказових правил прийняття рішень, методика Utah ZCT (включаючи варіанти PLT, DLT і CPC-RCMP), оцінена системою числового оцінювання штату Юта, та будь-який варіант методики ZCT із трьох запитань (event-specific three-question ZCT), пов'язаних із конкретною подією, оцінений за допомогою ESS.

Опубліковані та відтворені емпіричні докази існують для чотирьох поліграфологічних методик, які інтерпретуються за допомогою правил прийняття рішень, заснованих на припущенні дисперсії незалежного критерію серед релевантних запитань. Ці методики дали загальний незважений рівень точності .850 (.773 – .926) з невизначеним показником .125 (.068 – .183). В алфавітному порядку ці методики: методика DLST/TES, оцінена за допомогою семипозиційного методу TDA, методика DLST/TES, оцінена за допомогою ESS, методика AFMGQT, оцінена за допомогою семипозиційного методу TDA, та методика AFMGQT, оцінена за допомогою ESS.

Незважаючи на ці спостережувані та практичні відмінності, за винятком результатів викидів, не було виявлено значущих відмінностей щодо точності рішень або невизначених результатів серед поліграфологічних методик, які задовольняють вимогам стандартів практики АПА 2012 року. Подібним чином не було виявлено суттєвих відмінностей щодо точності рішень або невизначених результатів для поліграфологічних методик, інтерпретованих із припущенням незалежності чи не-незалежності серед релевантних запитань.

Не всі розглянуті методики мали достатню емпіричну підтримку, щоб відповідати стандартам АПА для включення в дослідження. Було виявлено, що в деяких названих поліграфологічних методиках відсутні будь-які опубліковані обґрунтовані докази, які можна було б використати для розрахунку розподілів вибірки, надійності та профілів критеріальної точності, необхідних для включення в цей мета-аналіз; вони перераховані в Додатку F. У Додатку G наведено висновок опублікованих до-

сліджень, які не можна було включити в мета-аналіз. Додаток Н містить опис методик, для яких існує єдине неповторене дослідження, що відповідає вимогам для включення в цей мета-аналіз. Ці методики не можна було включити в мета-аналіз, оскільки стандарт АПА вимагає наявності мінімум двох опублікованих досліджень. У Додатку І перераховано поліграфологічні методики, які опублікували та відтворили обґрунтовані докази, але критеріальна точність, що була повідомлена, не відповідала вимогам стандартів АПА 2012 року.

Поліграфологічні методики, які використовують трипозиційну модель TDA, не включені в мета-аналіз і тому не зазначені в Таблиці 1. Профілі критеріальної точності поліграфологічних методик, які використовують трипозиційну модель TDA, наведені в Додатку І. Незважена точність прийняття рішень була значною для всіх методик, заснованих на трипозиційних методах TDA, але не була однаковою для оманливих і правдивих випадків. Усі методики, які використовували трипозиційні методи TDA, постійно перевищували ліміт 2012 року для невизначених показників (20%). Оскільки показники критеріальної точності для трипозиційної системи оцінювання тестових даних TDA суттєво не відрізнялися від критеріальної точності семипозиційної системи оцінювання тестових даних TDA, початковий аналіз із застосуванням методу трипозиційної системи оцінювання TDA можна вважати прийнятним, якщо невизначені результати вирішуються за допомогою наступного аналізу за допомогою методу TDA, який забезпечує як точність, так і невизначені показники, які відповідають вимогам стандартів АПА 2012 року.

Деякі читачі можуть зауважити, що існує дві версії AFMGQT з незначними структурними відмінностями між ними<sup>53</sup>. Немає доказів того, що продуктивність однієї версії краща за іншу. Враховуючи те, що для відхилення нульової гіпотези про те, що відмінності є безглуздими, знадобляться обґрунтовані та повторені докази (дослідження). Враховуючи, що досліджувались обидві версії AFMGQT, то отримані результати надаються як узагальнення для обох версій. Дві широко використовувані та впізнавані методики, LEPET і Utah MGQT (версія методики Юти з чотирма релевантними запитаннями), не були включені в мета-аналіз, оскільки не вдалося знайти опублікованих досліджень на підтримку цих методик. Однак обидві ці поліграфологічні методики структурно майже ідентичні AFMGQT. Ми не можемо знайти причин, чому дані щодо обґрунтованості AFMGQT не можна узагальнити для цих методик, якщо оцінювати їх за допомогою тих самих методів числового оцінювання.

### **Порівняння з попередніми систематизованими оглядами**

Ми не перевіряли рівень значущості різниці між поточними оцінками точності та оцінками ОТА (1983), хоча можна легко побачити, що середній рівень точності польових досліджень .847 знаходиться за межами 95% довірчого інтервалу (.865 до .977) для поліграфологічних методик, які відповідають вимогам АПА 2012 року щодо доказового тестування. Ця відмінність, швидше за все, пов'язана з виключенням результатів досліджень, які не відповідають загальноприйнятій польовій практиці, і виключенням результатів поліграфологічних методик, які не дають задовільних результатів відповідно до стандартів практики АПА 2012. Існує мало обґрунтувань

<sup>53</sup> AFMGQT використовуються як у багатоаспектних розслідуваннях щодо відомих інцидентів, так і в скринінгових контекстах. **Обидва типи поліграфологічних досліджень, багатоаспектний і багатотемний, інтерпретуються з правилами прийняття рішень, заснованими на припущенні незалежного критерію дисперсії серед релевантних запитань.**

для використання в польовій практиці, а отже, недостатньо підстав для включення в оцінки точності поліграфологічних методик, які були витіснені більш ефективними методами. Включення таємничих або нестандартних методів в оцінку точності було б еквівалентом спроб відповісти на питання автомобільної промисловості щодо корпоративної економії палива, одночасно включаючи всі марки та моделі з 1960-х і 1970-х років епохи поглинання газу з розрахунку сучасної економіки. Тому методики, які забезпечують нестандартну та незадовільну критеріальну точність, були виключені з мета-аналізу.

Дані результати досліджень узгоджуються з результатами Хонтса та Петерсона (Honts & Peterson, 1997), Раскіна та Подлесного (Raskin & Podlesny, 1979), Абрамса (Abrams 1977, 1989) і Анслі (Ansley, 1990) щодо сліпого оцінювання тестових даних. Результати цього мета-аналізу також узгоджуються з результатами Національної дослідницької ради (2003), яка повідомила про рівень точності лабораторних досліджень як .860 разом із сукупним показником .890 для польових досліджень, використовуючи наукові дослідження, які відповідали їх критеріям відбору. Оскільки даний аналіз включає лише методики, якщо вони задокументовані та використовуються в польових умовах, ми припускаємо, що ці результати нададуть більш корисну та практичну відповідь поліграфологам, керівникам програм та професійним споживачам поліграфологічних результатів, які стикаються з необхідністю приймати рішення на основі обґрунтованих щодо вибору та використання в практиці доступних на даний момент поліграфологічних методик.

Ці результати більш консервативні, ніж результати Анслі (Ansley, 1983; 1990) і Абрамса (Abrams, 1973), що вимагає подальшого обговорення. Навряд чи поліграфологічний тест став менш точним протягом останніх трьох десятиліть. Більш реалістична можливість полягає в тому, що вибірки, включені в ранні публікації Анслі, були більш уразливими до переоцінки точності тесту в результаті методології відбору вибірок. Анслі (Ansley, 1990) зазначив, що судові рішення та докази іноді є ненадійними, і висловив перевагу підтвердженню зізнань результатів поліграфологічного дослідження. Надмірний акцент на підтвердженні зізнання включає в себе можливість ненавмисного систематичного виключення хибнонегативних і хибнопозитивних помилок, обидві умови навряд чи призведуть до зізнання, і, отже, критерій підтвердження. Самі зізнання є результатом невинного рішення продовжити обговорення з досліджуваним і розкрити його. Якщо рішення отримати зізнання частково базується на результатах поліграфологічного дослідження, то підтвердження за допомогою зізнання не залежить від результату тесту, а отже, виконується самостійно.

Вплив цих методологічних проблем може полягати в розподілі вибірки, який завищить точність поліграфологічного тесту<sup>54</sup>. Це явище може не обмежуватися підтвердженням через визнання; усі польові вибірки, відібрані через наявність і якість даних підтвердження, є потенційно невинними та нерепрезентативними. Це ж занепокоєння щодо відсутності незалежності (не-незалежності) підтверджуючих даних стосується результатів розслідування та судових рішень, які частково ґрунтуються на інформації, отриманій під час поліграфологічних досліджень. Цілком можливо, що це явище лежить в основі загальної тенденції в літературі, згідно з якою

---

<sup>54</sup> Можливий приклад цього явища можна побачити у Мангана та ін. (Mangan et al., 2008), які повідомили про результати опитування підтверджених зізнанням результатів тесту одного досвідченого поліграфолога. Повідомлені результати були на 100% точними, знахідка відповідала тому, що можна було б очікувати від упередженого відбору на основі визнання.

результати польових поліграфологічних досліджень загалом перевершують результати лабораторних експериментів<sup>55</sup>.

Незважаючи на потенційні або спостережувані відмінності вибірки між польовими та лабораторними дослідженнями, NRC (2003) не знайшов жодні значущі відмінності між результатами якісних польових і лабораторних досліджень.

Результати цього мета-аналізу узгоджуються з систематичним оглядом Крюсона (Crewson, 2003) щодо точності діагностичних поліграфологічних досліджень. Однак ці результати відрізняються від висновку Крюсона щодо скринінгових поліграфологічних досліджень. Рівень точності, знайдений для скринінгових поліграфологічних досліджень в цьому мета-аналізі, був вищим, ніж той, про який повідомив Крюсон, і різниця є статистично значущою ( $t [1008] = .002$ ). Хоча точну причину цієї різниці неможливо визначити з наявних даних, ми зазначаємо, що наукові дослідження та методики, використані Крюсоном, не можуть бути включені в цей мета-аналіз. Чотири з наукових досліджень щодо скринінгу, про які повідомив Крюсон, включали методику релевантних-нейтральних запитань (Relevant-Irrelevant technique) (Ansley, 1989; Brownlie, Johnson & Knill, 1997; Honts & Amato, 1999; Jayne, 1989), а решта наукового дослідження включала методику Ріда. Включені наукові дослідження, що стосуються критеріїв незалежності скринінгових поліграфологічних досліджень, були недоступні на момент огляду Крюсоном опублікованої наукової літератури.

### **Модератори та посередники**

Не було жодних ознак того, що результати наукових досліджень залежали від розміру вибірки або впливали на неї. Результати не кодувалися за характеристиками досліджуваних, включаючи вік, стать, етнічну приналежність, культуру, освіту чи соціально-економічний статус, а також дослідження не кодувалися за їхньою якістю чи методологією. Результати вибірки, які включали досліджуваних, які були піддані тій чи іншій формі експериментальних маніпуляцій (наприклад, ліки, втома, хронічні фізичні проблеми або хронічні проблеми з психічним здоров'ям, рівень функціонування, навчання контрзаходам або інструкції тощо), не були включені, і ці фактори не були оцінені.

За винятком викидів, суттєвих відмінностей у критеріальній точності поліграфологічних методик, придатних для доказового, парного та дослідницького тестувань, виявлено не було. Це свідчить про те, що ці категоріальні відмінності є довільними і тому безглуздими/несуттєвими в науковому сенсі. Однак значення стандартизованих вимог до точності тесту стає більш зрозумілим при розгляді адміністративних рішень, які підкреслюють або вимагають використання доказових методів і обмежують використання невалідних або експериментальних методів. Наукова цінність категоріальних відмінностей стає більш очевидною, якщо взяти до уваги складність відповідей на питання про наукову точність і ускладнення, які є результатом включення в оцінку точності менш точних і містичних методів, які були витіснені або замінені більш ефективними сучасними альтернативами. З етичної точки зору, важко уявити певне виправдання, коли рішення впливають на особисте життя, безпеку су-

---

<sup>55</sup> Альтернативне пояснення стверджує, що різниця є результатом відмінностей в екологічній та зовнішній валідності обставин тестування. Ці гіпотези не були ретельно оцінені, і було б нерозумно намагатися дійти будь-яких висновків з поточним станом розуміння. NRC (2003) повідомила, що ця тенденція не суперечить досвіду в інших галузях тестування та науки і повинна бути в центрі уваги майбутніх досліджень.

спільства та національну безпеку, для використання методів, які, згідно з науковими доказами, є неоптимальними або нестандартними.

Порівняння рівнів точності для поліграфологічних методик, інтерпретованих з припущенням про відсутність незалежності (не-незалежності) критерію на протилежну незалежності, не показало значних відмінностей у точності рішень. Однак значний ефект взаємодії для невизначених результатів свідчить про те, що можуть бути незначні відмінності в невизначених показниках для цих типів поліграфологічних досліджень. Здавалося б, це означає, що вибір різних стратегій поліграфологічних досліджень, включаючи незалежні або не-незалежні релевантні запитання, є питанням практики, яке має визначатися потребами обставин тестування.

### **Допоміжний аналіз**

Було виконано один допоміжний аналіз. Результати були розраховані для форматів CQT за виключенням тих наукових досліджень, які не задовольняли більш суворій низці критеріїв відбору. По-перше, поліграфологічні методики були виключені з допоміжного аналізу, в разі коли чутливість тесту до обману та специфічність тесту до правдивості не були статистично значущими, ніж випадковість. Це призвело до виключення методик AFMGQT, Federal ZCT і Federal You-Phase, в разі коли вони оцінюються за семіпозиційною моделлю TDA на додаток до методики Backster You-Phase. Статистичні викиди, не підтверджені наявними доказами, також були виключені. Це призвело до виключення IZCT і MQTZCT і кількох наукових досліджень, які були сильно заплутані. Виключення тих методик, для яких немає опублікованих статистичних даних, що описують надійність тесту, також призвело до виключення методики IZCT. Подібним чином, поліграфологічні методики та наукові дослідження були виключені, якщо існували значні відмінності між розподілами вибірки або основними ефектами, що вказувало на те, що розподіли вибірки не є репрезентативними один для одного. Це також призвело до виключення IZCT і MQTZCT. Дослідження також виключалися, якщо статистичні описи розподілів вибірки були недоступні або не могли бути розраховані за наявними даними. Це призвело до видалення двох наукових досліджень щодо DLST (Research Division Staff, 1995a; 1995b), одного наукового дослідження щодо AFMGQT (Senter, Waller & Krapohl, 2008) і двох наукових досліджень щодо MQTZCT (Shurani, Stein & Brand, 2009); Shurani, 2011).

Формати CQT, збережені для допоміжного аналізу, дали загальний показник точності прийняття рішень .898 (.840 – .955) і невизначений показник .092 (.033 – .150)<sup>56</sup> для поліграфологічних методик, інтерпретованих за допомогою правил прийняття рішень, заснованих на припущенні про відсутність незалежності критеріальної дисперсії серед запитань порівняння. Поліграфологічні методики, інтерпретовані з правилами прийняття рішень, заснованими на припущенні дисперсії незалежного критерію, дали показник точності рішення .857 (.782 – .932) і невизначений показник .117 (.058 – .177). Сукупний показник точності рішень для всіх наукових досліджень і поліграфологічних методик, включених до допоміжного аналізу, становила .883 (.817 – .950) з невизначеним показником .116 (.056 – .175)<sup>57</sup>. Двофакторні ANOVAs показали, що жодна основна ефективна взаємодія не була значущою при порівнянні точності рішення для допоміжного аналізу з точністю всього мета-аналі-

<sup>56</sup> Розраховується як середньозважене значення точності незваженого рішення та незваженого невизначеного показника.

<sup>57</sup> Розраховується як незважене середнє значення всіх наукових досліджень, включених у допоміжний аналіз.

зу. Ефект взаємодії був значущим для невизначених результатів [ $F(1,1992) = 17.335$ , ( $p < .001$ )]. Невизначені показники були дещо вищими для правдивих випадків з не-незалежними методиками та трохи вищими для оманливих випадків з незалежними від критеріїв поліграфологічними методиками. Post-hoc ANOVA показав, що жодна з однофакторних відмінностей не була суттєвою, що вказує на те, що ці невеликі відмінності навряд чи будуть помічені польовими дослідниками.

Однофакторні ANOVAs показали, що результати допоміжного аналізу суттєво не відрізнялися від результатів усього мета-аналізу щодо правильних рішень [ $F(1,5471) = 0.08$ , ( $p = 0.777$ )] або невизначених результатів [ $F(1,5471) = 0.08$ , ( $p = 0.777$ )]. Це вказує на те, що застосування більш суворих вимог до відбору досліджень навряд чи дасть мета-аналітичні результати, які відрізнятимуться від результатів цього дослідження.

### **Обмеження**

Два очевидних обмеження стосуються цього аналізу. По-перше, наукові дослідження не були закодовані для польових і лабораторних наукових досліджень, і не було зроблено жодних спроб дослідити будь-який вплив відмінностей у дизайні наукового дослідження. Натомість польові і лабораторні результати були включені з однаковою увагою, а результати всіх досліджень були об'єднані незалежно від змісту. По-друге, не було жодної спроби дослідити точність рішення на рівні окремих питань для будь-якої з включених поліграфологічних методик. З цією другою плутаниною пов'язаний той факт, що результати наукових досліджень із залученням DLST/TES і AFMGQT були отримані з використанням правил прийняття рішень, які базуються на припущенні про незалежність критеріїв серед релевантних запитань. Можливість узагальнення результатів цього мета-аналізу може частково залежати від правильності цього припущення.

Деякі з включених наукових досліджень порушені очевидними дослідницькими недоліками, найбільш помітним з яких є те, що деякі зразки були відібрані з наголосом на зізнанні досліджуваного як центральній ознаці критерію. Ще одна важлива помилка, яка спостерігалася в деяких із включених досліджень, полягала в тому, що основний автор також був розробником поліграфологічної методики, для якої існує певна форма права власності або фінансового інтересу. Дійсно, здається, що одним із маркерів для такого роду досліджень (і типових для досліджень з адвокації в інших місцях) є те, що повідомлені демонстрації майже ідеальної точності є статистичними викидами щодо розподілу результатів менш плутаних досліджень.

Відсутність критичної інформації та критичних коментарів у деяких звітах про дослідження справляє враження необ'єктивності файлів, коли менш ніж сприятливі результати не подаються для публікації. Інша версія цієї проблеми, схоже, виникла в контексті цього мета-аналізу, в якому деякі дані дослідження, спочатку надані Комітету, і деякі з опублікованих засобів вибірки включали лише ті результати, для яких були досягнуті оцінювачами правильні результати, спочатку приховуючи результати невизначених і помилкових випадків. Результатом цього є те, що опубліковані середні значення вибірки для деяких досліджень систематично позбавлені помилок або неконтрольованої дисперсії, і тому їх не можна вважати узагальненими.

Плутанина, пов'язана з окремими науковими дослідженнями, може ускладнити значення та інтерпретацію результатів мета-аналізу. Ці занепокоєння є прикладом цінності та необхідності наукової ретельності та незалежності при оцінці ефектив-

ності поліграфології та методів виявлення обману. Правила відбору та включення наукового дослідження для мета-аналізу мали бути максимально інклюзивними, але підтримувати рівень наукової строгості. Щоб зменшити вплив цих спотворень на мета-аналіз, було надано зведені результати як із викидами, так і без них.

Ще одна проблема, яка викликає плутанину в деяких дослідженнях, полягає в тому, що рівень освіти, підготовки та знань щодо психологічних, фізіологічних принципів і принципів підходів тестування може бути значно вищим для поліграфологів-учасників, ніж для більшості поліграфологів-практиків. До більшості сліпих досліджень поліграфологічної точності залучали досвідчених поліграфологів. Наукові дослідження щодо ESS були винятком із цієї тенденції, використовуючи недосвідчених спеціалістів та оцінювачів, а в останніх дослідженнях Хонтса (Honts) та його колег залучали студентів, навчених збирати наукові дані.

Мета-аналіз завжди передбачає встановлення правил відбору наукового дослідження, і цілком можливо, що мета-аналіз, заснований на іншому наборі критеріїв для включення, призведе до інших результатів. Усі наукові дослідження в цьому аналізі розглядалися однаково, в разі якщо вони відповідали вимогам до публікації та надавали достатньо інформації для оцінки критеріальної точності та надійності або можливості узагальнення результатів наукового дослідження. Якісні вимоги для включення в це наукове дослідження стосувалися лише того, чи включені наукові дослідження задовільно представляли прилади та компоненти для польових тестувань, а також задовільно представляли поліграфологічну методику, для якої існує опублікований опис послідовності тестових запитань і методу аналізу тестових даних. Мета-аналітичні вагові значення були призначені відповідно до розміру вибірки та кількості оцінювачів для кожного наукового дослідження, хоча очевидних ефектів, пов'язаних із розміром вибірки, не було. Хоча попередні статистичні аналізи не виявили суттєвих відмінностей у результатах польових і лабораторних досліджень, цілком можливо, що мета-аналітичні результати дещо відрізнялися, якби включені дослідження були закодовані та зважені за іншими параметрами, включаючи якість наукового дослідження, структуру, вибірку методології, майнові інтереси або включення основного автора, як учасника дослідження.

Іншим обмеженням цього аналізу є те, що жодне з включених наукових досліджень не включало неповнолітніх. Як наслідок, консервативна оцінка поточних результатів свідчить про те, що нашу наявну базу знань можна вважати застосовною лише до фізично та психічно здорових дорослих людей з нормальними функціональними характеристиками. У більш широкому тлумаченні було б визнано, що існує невелика різниця між дорослими та більш старшими неповнолітніми з точки зору фізіології, зареєстрованої або використаної сучасними давачами поліграфа, і невелика різниця в психологічних основах реакцій в поліграфологічному дослідженні між дорослими та старшими, зрілими в розвитку неповнолітніми. Узагальнення цих результатів на осіб, які є відомими викидами порівняно з очікуваним розподілом осіб із нормативної популяції (тобто осіб, чії функціональні характеристики виходять за межі нормального діапазону), слід проводити з великою обережністю. Деякі з включених досліджень не мали повної інформації, хоча можна було обчислити достовірність, розподіл вибірки, і розмірний профіль критеріальної точності з вихідних даних, які були надані тимчасовій комісії. Вибіркові дані не були доступні для деяких наукових досліджень, зокрема, від уряду США. Ці дослідження включають в себе дослідження розробки та перевірки методик TES (DLST) та AFMGQT. Ці дослід-



дження показали достовірність і точність даних, які були достатніми для включення їх до мета-аналізу, і всі ці дослідження були відтворені незалежно.

Очевидним обмеженням цього мета-аналізу є те, що він не включав результати комп'ютеризованих алгоритмів оцінювання.

Варто згадати ще одне питання. Головний дослідник цього мета-аналізу також був основним автором ряду включених наукових досліджень<sup>58</sup>. Комітет знав, що його дослідження значною мірою, а часом і виключно, було залучено до досліджень, які пропонувалися для перевірки деяких із включених методик. Деякі з цих методик не були б включені без цих наукових досліджень. Таким чином, обов'язком комітету було зважити його судження щодо факторів, які могли зменшити його незалежність. Це викликало важливе занепокоєння, оскільки упередженість у виборі наукового дослідження або аналізі даних серйозно поставила під загрозу об'єктивність кінцевого звіту. Після більш ретельного вивчення застереження щодо цих двох потенційно конфліктуючих ролей (основного автора мета-аналізу та автора досліджень, включених до мета-аналізу) були пом'якшені відсутністю будь-яких очевидних особистих інтересів у результатах цих досліджень та його обмеженою участю у будь-якому дослідженні (ніколи не проводив і не оцінював жодного з поліграфологічних досліджень). Його опубліковані чи незавершені дослідження не виявили жодної помітної моделі переваги чи неприйняття певних поліграфологічних методик. Нарешті, усі рішення про включення наукового дослідження приймалися колективно членами комітету на основі обґрунтованості дослідження: жоден член комітету не мав абсолютних повноважень виключити або включити дане наукове дослідження. Хоча обґрунтовані особисті думки можуть відрізнитися в деяких частинах цього звіту, комітет свідомо подбав про те, щоб особисті інтереси членів комітету не стали приводом для критики звіту. Повне розкриття відношення між головним автором цього звіту та його дослідженням надається тут для виконання стандартних етичних зобов'язань у наукових звітах.

### **Рекомендації**

Оскільки серед 14 поліграфологічних методик, включених до цього мета-аналізу, не було виявлено суттєвих відмінностей, не слід робити спроби описати ці методики в термінах рангового порядку щодо ефективності. Наявні докази не підтверджують перевагу будь-якої поліграфологічної методики порівняно з іншими. Тому спроби встановити будь-яку ієрархію ефективності є невиправданими. Натомість слід приділяти менше уваги названим поліграфологічним методикам та беззмисловим відмінностям у форматах поліграфологічного тестування. Більше уваги слід приділяти деталям конструкції тесту, для яких є реплікації доказів їх внеску в критеріальну точність. Більше уваги слід приділити важливим практичним і теоретичним відмінностям у поліграфологічних методиках, для яких RQ інтерпретуються як незалежні або не-незалежні.

Однією з практичних областей необхідного дослідження є можливість узагальнення нормативних даних і оцінок точності поліграфологічних методик, інтерпретованих з припущенням незалежності критеріїв, включаючи як багатоаспектні, так і багатотемні поліграфологічні дослідження. Інша практична область необхідних до-

---

<sup>58</sup> Пан Нельсон не має фінансових, майнових чи особистих інтересів у будь-якій поліграфологічній методиці або методології, включених до цього мета-аналізу.

сліджень передбачає використання DLC з додатковими поліграфологічними тестовими форматами.

Необхідно продовжувати постійні наукові дослідження та вдосконалювати всі поліграфологічні методики, і ці вдосконалення мають бути повністю інтегровані як у навчальну, так і в польову практику. Необхідно провести додаткові дослідження, щоб розширити базу знань щодо таких змінних, як характеристики досліджуваних (наприклад, неповнолітні, люди похилого віку, особи з психічними захворюваннями та особи з медичними проблемами) на додаток до деталей злочину або характеристик, які призводять до найбільш ефективного використання поліграфологічних досліджень. Потрібні додаткові дослідження для скринінгових досліджень, включаючи дослідження, що стосуються складності теорії прийняття рішень, притаманних поліграфологічним дослідженням з кількома незалежними цілями. Дослідники повинні продовжувати збільшувати використання моделей Монте-Карло та інших статистичних методів, які можна використовувати для надання відповідей на складні дослідницькі проблеми, які важко дослідити іншими методами. Результати наукових досліджень за методом Монте-Карло слід порівнювати з результатами експериментів у польових і лабораторних умовах.

Було запропоновано, що низка посередницьких змінних має суттєвий вплив на точність поліграфологічних досліджень, і деякі з них включають складні психологічні та лінгвістичні припущення, які можуть або не можуть бути повністю перевірені. Гіпотези, які не підлягають перевірці, слід відкинути на користь гіпотез, які можна перевірити, і слід провести додаткове вивчення, щоб зрозуміти переваги процедурних і структурних гіпотез, які були запропоновані як пов'язані з точністю тесту. Докази наукових досліджень мають стати постійним очікуванням, і розробники та практики поліграфологічних досліджень повинні протистояти спокусі включити авторитетні та анекдотичні теорії, які не були перевірені.

Потрібні підвищені стандарти щодо дослідження, включаючи вимоги щодо прозорості та заяв про зацікавленість від усіх авторів та учасників. Що ще важливіше, оскільки дослідження ефективності поліграфологічного тесту є процесом тестування тесту, первинні автори повинні утримуватися від участі в дослідженні. Особливо важливо, коли основний автор також є розробником поліграфологічної методики або не є незалежним через фінансові чи ділові стосунки з розробником, потрібно щоб дослідницькі дані та методологія піддавалися суворій об'єктивній зовнішній перевірці перед тим, як професію чи спільноту спонукають поклагатися на результати дослідження.

Слід вимагати від дослідників надання статистичних описів розподілу вибірки. Це сприятиме більш ефективному співставленню розподілу вибірки та підвищить здатність оцінювати та розуміти репрезентативність та можливість узагальнення результатів наукового дослідження. Подібно до того, як результати окремих неповторених (не реплікованих) наукових досліджень мають невелику реальну цінність для мета-аналітичного дослідження, результати досліджень, у яких використовується один експертний оцінювач, не мають дійсної цінності для професії. Слід заохочувати дослідників або вимагати від них використання низки учасників, які мають різну підготовку та досвід. Як поліграфологів, так і досліджуваних слід вибирати випадковим чином, коли це можливо. Це збільшить здатність вивчати та розуміти можливість узагальнення поліграфологічних методів.

Деякі наукові дослідження, включені до мета-аналізу, не мають належного визначення щодо типу дослідження, і ефект потенційно вводить в оману професію.

Пілотні наукові дослідження та огляди мають бути чітко визначені як такі, і їх не слід включати в майбутні систематичні огляди чи мета-аналіз критеріальної точності. Наукові дослідження щодо критеріїв також слід чітко виділити з досліджень, призначених для оцінки змінних модераторів чи медіаторів або питань конструкту та причинності.

Статистика надійності повинна вимагатися для всіх досліджень, якщо це не заборонено структурою дослідження (наприклад, дослідження комп'ютерного алгоритму або моделювання). Від основних авторів слід вимагати надання всіх необроблених даних і числових оцінок для перегляду та розширеного аналізу.

Результати комп'ютеризованих статистичних алгоритмів TDA повинні бути включені в майбутні дослідження такого типу. Використання комп'ютерів і статистичної теорії прийняття рішень все ще не є поширеним у методах TDA для поліграфологічних досліджень. Натомість методи TDA у польовому застосуванні наголошують на ручних методах підрахунку балів із точністю на цілочисельному та ранговому рівнях, які слід вважати грубими та ненадійними порівняно з точністю та надійністю, які можна отримати за допомогою автоматизованих вимірювань та статистичного аналізу. Зростає кількість обґрунтувань, які вказують на те, що комп'ютерні алгоритми можуть бути такими ж або більш ефективними, ніж ручні методи TDA, якщо дані мають задовільну якість. Оскільки результати поліграфологічного дослідження можуть відігравати роль підтримки прийняття рішень у питаннях, які впливають на життя людей, безпеку громади та національну безпеку, від розробників комп'ютерних алгоритмів слід вимагати надання повного опису операційних процедур, на додаток до критеріїв оцінки, перетворення даних та методології агрегування, нормативні дані та статистичну основу в теорії прийняття рішень, теорію виявлення сигналу, теорію розрізнення сигналу, регресійного аналізу або машинного навчання.

Необхідно продовжувати розвиток та вдосконалення методології поліграфологічних тестувань. Процедури поліграфологічного тестування мало змінилися за останнє десятиліття. Зусилля щодо розробки протягом цього часу були зосереджені на вдосконаленні методів аналізу тестових даних, включаючи правила прийняття рішень (Senter, 2003; Senter & Dollins, 2002; 2004; 2008), розробку статистичних алгоритмів (Nelson, Krapohl & Handler, 2008), числові перетворення (Krapohl, 2010; Nelson, Krapohl & Handler, 2008; Nelson et al., 2011), а також збільшення використання нормативних даних для розрахунку частоти помилок і оптимальних показників порогових значень для рішень (Krapohl, 2010; Nelson & Handler, 2010; Nelson, Krapohl & Handler, 2008; Nelson et al., 2011). Необхідні додаткові дослідження, більш детальне вивчення та порівняння моделей числових перетворень, включаючи семипозиційні, трьохпозиційні методи, ESS, методи ранжування, комп'ютерні алгоритми, а також застосування цих перетворень до поліграфологічних досліджень, побудованих з незалежних і не-незалежних цілей дослідження.

Давачі компонентів поліграфа мало змінилися протягом кількох десятиліть. Це може бути змішане благо. Хоча критики можуть вказати на це як на застій у дослідженнях і розробках, існує значна опублікована база знань, що підтверджує та описує ефективність використовуваного на даний момент масиву давачів компонентів поліграфа. Було б передчасно відмовлятися від цієї бази знань, намагаючись задовольнити колективний голод нових методів. Будь-яка заміна компонентів давачів, які зараз використовуються, повинна супроводжуватися опублікованими та реплікованими доказами того, що дані та інформація, надані новими давачами,

такі ж або кращі, ніж дані та інформація від давачів, які зараз використовуються. Крім того, використання нових і вдосконалених давачів зіштовхнеться зі значним і нетривіальним тягарем розробки та щодо включення цих даних у нові або існуючі моделі структурованих рішень та нормативні дані. Незважаючи на ці загальні застереження щодо заміни компонентів в поліграфології та фізіологічних заходів, також очевидно, що поліграфологічне тестування залишається недосконалим і потребує подальшого вдосконалення. Методи поліграфологічного тестування не можна вдосконалити без заміни малоефективних методів і методик на більш ефективні.

Для більшої впевненості в ефективності IZCT і MQTZCT (або будь-яких патентованих особистих методів<sup>59</sup>), вони повинні бути відтворені незалежними дослідниками, які не розробляли методики, не мали ділових стосунків з розробниками, не проводили поліграфологічні дослідження, проаналізувати дані або повідомити про результати під час поліграфологічних досліджень. Усі необроблені дані та числові оцінки мають бути доступними для розширеного аналізу. Якщо ці методики за своєю суттю перевершують інші, не повинно виникати особливих труднощів підтвердити це за допомогою високоякісних незалежних досліджень.

Нарешті, цей мета-аналіз слід повторити в майбутньому з включенням нової виникаючої інформації. Майбутні мета-аналітичні дослідження повинні кодувати та оцінювати такі показники, як характеристики поліграфолога, характеристики досліджуваного та показники щодо якості дослідження, фінансові інтереси та інші можливі показники.

## **Висновки**

Результати цього мета-аналізу показують, що низка досліджень має задовільну вимогу до якості, щоб забезпечити основу емпіричної підтримки, що описує можливість узагальнення ряду поліграфологічних методик на рівнях критеріальної точності, які значно перевищують випадкові очікування. Хоча стандарти практики АПА 2012 року і вимоги до точності тестів є дещо довільними, вони корисні для професії. Метою професійних стандартів є сприяння у використанні ефективних методів і запобігання використанню менш ефективних і неперевіраних методів. Було встановлено, що чотирнадцять поліграфологічних методик підтверджуються кількома опублікованими дослідженнями та відповідають вимогам стандартів практики АПА 2012 року. Нормативні дані доступні для кожного з цих 14 поліграфологічних методик. Зазначимо, що незважаючи на недосконалість поліграфа, NRC (2003) повідомила, що жодна з потенційно нових технологій не була готова замінити поліграф, і ця умова, здається, не змінилася на даний момент.

Самі стандарти АПА не встановлюють якісних або методологічних вимог до наукових доказів<sup>60</sup>, і жодних кількісних вимог, окрім вимог щодо двох публікацій, які

---

<sup>59</sup> Чесно кажучи, щодо розробників цих методів, кожен метод «детектору брехні» за останні 100 років, який досліджував розробник або захоплений користувач, оцінюючи його чи його власні дослідження, повідомляв про точність, що наближалася до досконалості. Це одна з характерних рис адвокаційних досліджень у всіх сферах, а не лише у виявленні обману. Тенденція включає методіку переривчастого артеріального тиску Марстона (Marston's discontinuous blood pressure technique), патометр Саммерса (Summer's Pathometer) в 1930-х роках, МакНітта з методикою Релевантних-нейтральних запитань в 1940-х роках, ГКТ Ліккена, «Відбитки пальців мозку» Фарвелла (Farwell's "Brain Fingerprinting,") та аналізатор стресу комп'ютерного голосу. У кожному випадку автори повідомили про надзвичайну точність, яка зазвичай перевищує 99%. Однак у всіх цих випадках подальші дослідження або були відсутні, або продемонстрували точність, значно нижчу за первинні звіти.

<sup>60</sup> АПА прийняла стандарт щодо дослідження, який можна знайти в Інтернеті на [www.polygraph.org](http://www.polygraph.org) і надрукувала в журналі *Polygraph*.

вказують на певні рівні точності для висновків поліграфологічного дослідження. У цьому полягає потенційна слабкість стандартів АПА: спрощене тлумачення вимог передбачає, що будь-хто може двічі натиснути чорнилом на папір у самовиданих томах, заявляючи про ідеальну або майже ідеальну точність, а потім стверджувати про відповідність стандартам АПА для підтвердження тесту за вищим категорійним рівнем – для доказових тестувань. Хоча науковці сприйняли б це з деяким скептицизмом, цей приклад ілюструє потребу мета-аналітичних досліджень у визначенні більш жорстких критеріїв включення та виключення наукового дослідження. Так само, як не всі докази є хорошими доказами, так не всі публікації корисні. Сама публікація не є схваленням факту, а лише вказує на те, що редактори та рецензенти погодилися, що робота буде певною мірою цікава для професії.

Нічого в цьому документі не слід сприймати як припущення про те, що ми знаємо все, що нам потрібно знати, або все, що потрібно знати, про поліграфологічне дослідження. Завжди є чому навчитися, і завжди є потреба в продовженні досліджень. Більше інформації, безсумнівно, стане доступною у майбутньому, і професіонали повинні продовжувати впроваджувати практики, засновані на нових і вдосконалених обґрунтованих доказах наукових досліджень. Вчинити інакше означає підпорядкувати майбутнє професії думці, яка буде вразливою до особистостей, політики та особистих інтересів. У найсуворішому розумінні поліграфологічні методики, для яких існує недостатня база опублікованих і повторених наукових досліджень, повинні вважатися експериментальними, незалежно від того, як довго вони існують. Проте пропозиція відмовитися від неперевіраних, неефективних або експериментальних поліграфологічних методик, які давно використовуються в польовій практиці, не позбавлена суперечок.

Можна стверджувати, і було стверджено, що деякі недосліджені або експериментальні методи можуть працювати так само добре або краще, ніж деякі перевірені методи, і можуть певним чином надавати спеціальні переваги. І навпаки, також можливо, що експериментальні та неперевірені методи працюють не так добре, як ті, що мають докази наукової валідності. У найгірших обставинах використання експериментальних методів може призвести до несприятливих наслідків, яких в іншому разі можна уникнути. Консервативна оцінка припускає, що практика проведення експериментальних методів серед населення, коли ефективні обґрунтовані методи доступні та легко впроваджуються без додаткових витрат (зусиль), може вважатися розумним за деяких обставин, але вимагає переконливого обґрунтування та включає етичні вимоги до отримання інформованої згоди та попередження про це.

Нарешті, цей мета-аналіз слід розглядати лише як інформаційний ресурс, а результати цього дослідження не слід інтерпретувати як політику АПА. Цей документ або результати цього дослідження не являються єдиною або остаточною авторитетною думкою щодо валідації поліграфологічного тесту. Інші, настільки ж обґрунтовані підходи також можливі щодо оцінки наукової літератури щодо поліграфологічного тестування.

Цей проєкт було завершено з метою підсумувати наявну опубліковану наукову літературу щодо поліграфологічних методик і критеріальної точності, а також забезпечити зручний ресурс для тих, хто, можливо, бажає уникнути тягаря перегляду дослідницької літератури самостійно. Було докладено всіх зусиль, щоб надати не лише висновки, але й поглиблені пояснення, які лежать в основі цих висновків, щоб читачі могли краще зрозуміти їх основу.

Інформація, наведена в цьому документі, надається Раді директорів АПА, щоб повідомити її професійним членам про показники валідності поліграфологічних методик, які використовуються на даний момент. Ця інформація призначена лише для того, щоб допомогти фахівцям-поліграфологам прийняти обґрунтовані рішення щодо вибору поліграфологічних методик для використання в польових умовах. Це також може допомогти адміністраторам програм, політикам і судам приймати рішення на основі доказів щодо інформаційної цінності результатів поліграфологічних тестів загалом. Ніщо не повинно перешкоджати використанню будь-якої поліграфологічної методики, яка підтверджується науковими дослідженнями, які демонструють рівень точності, значно більший за випадковість, за умови, що таке використання відповідає вимогам місцевих законів, нормативних актів і чинним стандартам.

Як цитувати цей документ:

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. [Electronic version] Retrieved <DATE>, from <http://www.polygraph.org>.

**Список використаних джерел**

\* вказує на дослідження, які були включені в Мета-аналіз.

□ вказує на дослідження, які були вказані лише в додатках.

- Abrams, S. (1973). Polygraph validity and reliability: A review. *Journal of Forensic Sciences*, 18, 313–326.
- Abrams, S. (1977). *A polygraph handbook for attorneys*. Lexington, MA: Lexington Books.
- Abrams, S. (1989). *The complete polygraph handbook*. Lexington, MA: Lexington Books.
- Abrams, S. (1984). *The question of the intent question*. *Polygraph*, 13, 326–332.
- Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions In Psychological Science*, 8, 3–9.
- Ansley, N. (1983). A compendium on polygraph validity. *Polygraph*, 12, 53–61.
- Ansley, N. (1989). *Accuracy and utility of RI screening by student examiners at DODPI*. Polygraph and Personnel Security Research. Office of Security. National Security Agency. Fort George G. Meade, MD.
- Ansley, N. (1990). The validity and reliability of polygraph decisions in real cases. *Polygraph*, 19, 169–181.
- Ansley, N. (1992). The history and accuracy of guilty knowledge and peak of tension tests. *Polygraph*, 21, 174–247.
- Backster, C. (1963). *Standardized polygraph notepad and technique guide: Backster zone comparison technique*. Cleve Backster: New York.
- Backster School of Lie Detection (2011). *Basic polygraph examiner's course chart interpretation notebook*. Backster School of Lie Detection: San Diego.
- Barland, G. H., Honts, C. R., & Barger, S. D. (1989). Studies of the accuracy of security screening polygraph examinations. Department of Defense Polygraph Institute.
- Barland, G. H. & Raskin, D. C. (1975). Psychopathy and detection of deception in criminal suspects. *Psychophysiology*, 12, 224.
- Bell, B. G., Kircher, J. C., & Bernhardt, P. C. (2008). New measures improve the accuracy of the direct-lie test when detecting deception using a mock crime. *Physiology and Behavior*, 94, 331–340.

- Bell, B. G., Raskin, D. C., Honts, C. R., & Kircher, J. C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1–9.
- Blackstone, K. (2011). *Polygraph, Sex Offenders, and the Court: What Professionals Should Know About Polygraph., and a Lot More*. Concord, MA: Emerson Books.
- \*Blackwell, J. N. (1998). PolyScore 33 and psychophysiological detection of deception examiner rates of accuracy when scoring examination from actual criminal investigations. Available at the Defense Technical Information Center. DTIC AD Number A355504/PAA. Reprinted in *Polygraph*, 28(2) 149–175.
- \*Blalock, B., Cushman, B., & Nelson, R. (2009). A replication and validation study on an empirically based manual scoring system. *Polygraph*, 38, 281–288.
- Blalock, B., Nelson, R., Handler, M., & Shaw, P. (2011). A position paper on the use of directed lie comparison questions in diagnostic and screening polygraphs. *Police Polygraph Digest*, 2–5.
- Brownlie, C., Johnson, G. J., & Knill, B. (1997). Validation study of the relevant/irrelevant screening format. Unpublished report.
- Capps, M. H. (1991). Predictive value of the sacrifice relevant. *Polygraph*, 20(1), 1–8.
- Capps, M. H. & Ansley, N. (1992). Comparison of two scoring scales. *Polygraph*, 21, 39–43.
- Capps, M. H., Knill, B. L., & Evans, R. K. (1993). Effectiveness of the symptomatic questions. *Polygraph*, 22, 285–298.
- Correa, E. J. & Adams, H. E. (1981). The validity of the pre-employment polygraph examination and the effects of motivation. *Polygraph*, 10, 143–155.
- Crewson, P. E. (2001). A comparative analysis of polygraph with other screening and diagnostic tools. Research Support Service. Report No. DoDPI01-R-0003. Reprinted in *Polygraph* 32, 57–85.
- Department of Defense (2006). Federal psychophysiological detection of deception examiner handbook. Reprinted in *Polygraph*, 40(1), 2–66.
- \*Driscoll, L. N., Honts, C. R., & Jones, D. (1987). The validity of the positive control physiological detection of deception technique. *Journal of Police Science and Administration*, 15, 46–50. Reprinted in *Polygraph*, 16(3), 218–225.
- Forman, R. F. & McCauley, C. (1986). Validity of the positive control polygraph test using the field practice model. *Journal of Applied Psychology*, 71, 691–698. Reprinted in *Polygraph*, 16(2), 145–160.
- Ganguly, A. K., Lahri, S. K., & Bhaseen, V. (1986). Detection of deception by conventional qualitative method and its confirmation by quantitative method - An experimental study in polygraphy. *Polygraph*, 15, 203–210.
- Ginton, A., Daie, N., Elaad, E., & Ben-Shakhar, G. (1982). A method for evaluating the use of the polygraph in a real-life situation. *Journal of Applied Psychology*, 67, 131–137.
- Gordon, N. J. (1999). The academy for scientific investigative training's horizontal scoring system and examiner's algorithm system for chart interpretation. *Polygraph*, 28, 56–64.
- Gordon, N. J., Fleisher, W. L., Morsie, H., Habib, W., & Salah, K. (2000). A field validity study of the integrated zone comparison technique. *Polygraph*, 29, 220–225.
- \*Gordon, N. J., Mohamed, F. B., Faro, S. H., Platek, S. M., Ahmad, H., & Williams, J. M. (2005). Integrated zone comparison polygraph technique accuracy with scoring algorithms. *Physiology & behavior*, 87(2), 251–254. (Same study is described in Mohamed, F. B., Faro, S. H., Gordon, N. J., Platek, S. M., Ahmad, H. & Williams, J.M. (2006).)
- Handler, M. (2006). The Utah PLC. *Polygraph*, 35, 139–148.
- Handler, M. & Nelson, R. (2008). Utah approach to comparison question polygraph testing. *European Polygraph*, 2, 83–119.

- Handler, M. & Nelson, R. (In press). Criterion validity of the United States Air Force Modified General Question Technique and three position scoring. *Polygraph*.
- \*Handler, M., Nelson, R., Goodson, W., & Hicks, M. (2010). Empirical Scoring System: A cross-cultural replication and extension study of manual scoring and decision policies. *Polygraph*, 39, 200–215.
- Harwell, E. M. (2000). A comparison of 3- and 7-position scoring scales with field examinations. *Polygraph*, 29, 195–197.
- Hilliard, D. L. (1979). A cross analysis between relevant questions and a generalized intent to answer truthfully question. *Polygraph*, 8, 73–77.
- \*Honts, C. R. (1996). Criterion development and validity of the CQT in field application. *The Journal of General Psychology*, 123, 309–324.
- Honts, C. R., & Amato, S. L. (1999). The automated polygraph examination: Final report of U. S. Government Contract No. 110224-1998-MO. Boise State University.
- \*Honts, C. R., Amato, S. & Gordon, A. (2004). Effects of outside issues on the comparison question test. *Journal of General Psychology*, 131(1), 53–74.
- Honts, C. R. & Driscoll, L. N. (1987). An evaluation of the reliability and validity of rank order and standard numerical scoring of polygraph charts. *Polygraph*, 16, 241–257.
- Honts, C. R. & Hodes, R. L. (1983). The detection of physical countermeasures. *Polygraph*, 12, 7–17.
- \*Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, 70(1), 177–187.
- Honts, C. R. & Peterson, C. F. (1997). Brief of the Committee of Concerned Social Scientists as Amicus Curiae United States v Scheffer. Available from the author.
- \*Honts, C. R. & Raskin, D. (1988). A field study of the validity of the directed lie control question. *Journal of Police Science and Administration*, 16(1), 56–61.
- \*Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Psychophysiology*, 1, 241–247.
- Honts, C. R., & Reavy, R. (2009). Effects of Comparison Question Type and Between Test Stimulation on the Validity of Comparison Question Test. US Army Research Office: Grant Number W911NF-07-1-0670.
- \*Horowitz, S. W., Kircher, J. C., Honts, C. R., & Raskin, D. C. (1997). The role of comparison questions in physiological detection of deception. *Psychophysiology*, 34, 108–115.
- Horvath, F. S. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology*, 62, 127–136.
- Horvath F. S. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. *Journal of Police Science and Administration*, 16(3), 198–209. Reprinted in *Polygraph*, 20, 7–25.
- Horvath, F. S. (1994). The value and effectiveness of the sacrifice relevant question: An empirical assessment. *Polygraph*, 23, 261–279.
- Horvath, F. & Palmatier, J. (2008). Effect of two types of control questions and two question formats on the outcomes of polygraph examinations. *Journal of Forensic Sciences*, 53(4), 1–11.
- Horvath, F. S. & Reid, J. E. (1971). The reliability of polygraph examiner diagnosis of truth and deception. *Journal of Criminal Law, Criminology and Police Science*, 62, 276–281.
- Hunter, F. L., & Ash, P. (1973). The accuracy and consistency of polygraph examiners' diagnosis. *Journal of Police Science and Administration*, 1, 370–375.
- Jayne, B. (1989). A comparison between the predictive value of two common preemployment screening procedures. *The Investigator*, 5(3).



- Jayne, B. C. (1990). Contributions of physiological recordings in the polygraph technique. *Polygraph*, 19, 105–117.
- Kircher, J. C., Kristjansson, S. D., Gardner, M. K., & Webb, A. (2005). Human and computer decision-making in the psychophysiological detection of deception. University of Utah.
- \*Kircher, J. C. & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73, 291–302.
- Kokish, R., Levenson, J. S., & Blasingame, G. D. (2005). Post-conviction sex offender polygraph examination: client-reported perceptions of utility and accuracy. *Sexual Abuse : A Journal of Research and Treatment*, 17, 211–21.
- Krapohl, D. J. (1998). A comparison of 3- and 7- position scoring scales with laboratory data. *Polygraph*, 27, 210–218.
- Krapohl, D. J. (2002). Short report: Update for the objective scoring system. *Polygraph*, 31, 298–302.
- Krapohl, D. J. (2005). Polygraph decision rules for evidentiary and paired testing (Marin protocol) applications. *Polygraph*, 34, 184–192.
- Krapohl, D. J. (2006). Validated polygraph techniques. *Polygraph*, 35(3), 149–155.
- Krapohl, D. J. (2010). Short report: A test of the ESS with two-question field cases. *Polygraph*, 39, 124–126.
- \*Krapohl, D. J. & Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: A replication. *Polygraph*, 35(1), 55–63.
- Krapohl, D. J., Dutton, D. W. & Ryan, A. H. (2001). The rank order scoring system: Replication and extension with field data. *Polygraph*, 30, 172–181.
- Krapohl, D. J. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209–222.
- Krapohl, D. J. & Norris, W. F. (2000). An exploratory study of traditional and objective scoring systems with MGQT field cases. *Polygraph*, 29, 185–194.
- Krapohl, D. J. & Ryan, A. H. (2001). A belated look at symptomatic questions. *Polygraph*, 30, 206–212.
- Krapohl, D. J., Senter, S. M., & Stern, B. A. (2005). An exploration of methods for the analysis of multiple-issue Relevant/Irrelevant screening data. *Polygraph*, 34(1), 47–62.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, 43, 385–388.
- \*MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *The Journal of Applied Psychology*, 86, 674–683.
- \*Mangan, D. J., Armitage, T. E., & Adams, G. C. (2008). A field study on the validity of the Quadri-Track Zone Comparison Technique. *Physiology and Behavior*, 17–23.
- Matte, J. A. (1990). Validation study on the polygraph Quadri-Zone Comparison Technique. Research Abstract LD 01452, Vol. 1502, 1989, University Microfilm International (UMI), Ann Arbor, MI.
- Matte, J. A. (2010). A field study of the Backster Zone Comparison Technique's Either Or Rule and scoring system versus two other scoring systems when relevant question elicits strong response. *European Polygraph*, 4, 53–69.
- \*Matte, J. A. & Reuss, R. M. (1989). A field validation study of the Quadri-Zone Comparison Technique. *Polygraph*, 18, 187–202.
- Meiron, E., Krapohl, D. J., & Ashkenazi, T. (2008). An assessment of the Backster "Either-Or" Rule in polygraph scoring. *Polygraph*, 37, 240–249.
- Mohamed, F. B., Faro, S. H., Gordon, N. J., Platek, S. M., Ahmad, H., & Williams, J. M. (2006). Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation-initial experience. *Radiology*, 238, 679–88.

National Research Council (2003). *The Polygraph and Lie Detection*. Washington, D.C.: National Academy of Sciences.

\*Nelson, R. (In press). Monte Carlo study of criterion validity of Backster You-Phase examinations. *Polygraph*.

\*Nelson, R. (In press). Monte Carlo study of criterion validity of the Directed Lie Screening Test using the seven-position, three-position and Empirical Scoring Systems. *Polygraph*.

\*Nelson, R. (2011). Monte Carlo study of criterion validity for two-question zone comparison tests with the Empirical Scoring System, seven-position, and three-position scoring models. *Polygraph*, 40, 146–156.

\*Nelson, R. & Blalock, B. (In press). Extended analysis of Senter, Waller and Krapohl's AFMGQT examination data with the Empirical Scoring System and the Objective Scoring System, version 3. *Polygraph*, (In press).

\*Nelson, R., Blalock, B., & Handler, M. (2011). Criterion validity of the Empirical Scoring System and the Objective Scoring System, version 3 with the USAF Modified General Question Technique. *Polygraph*, 40, 172–179.

\*Nelson, R., Blalock, B., Oelrich, M., & Cushman, B. (2011). Reliability of the Empirical Scoring System with expert examiners. *Polygraph*, 40, 131–139.

Nelson, R. & Handler, M. (2010). *Empirical Scoring System*. Lafayette Instrument Company.

\*Nelson, R. & Handler, M. (In press). Monte Carlo study of the United States Air Force Modified General Question Technique with two three and four questions. *Polygraph*.

\*Nelson, R., Handler, M., Adams, G., & Backster, C. (In press). Survey of reliability and criterion validity of Backster numerical scores of You-Phase exams from confirmed field investigations. *Polygraph*.

\*Nelson, R., Handler, M., Blalock, B., & Cushman, B. (In press). Blind scoring of confirmed federal You-Phase examinations by experienced and inexperienced examiners: Criterion validity with the Empirical Scoring System and the seven-position model. *Polygraph*.

\*Nelson, R., Handler, M., Blalock, B., & Hernández, N. (In press). Replication and extension study of Directed Lie Screening Tests: Criterion validity with the seven- and three-position models and the Empirical Scoring System. *Polygraph*.

\*Nelson, R., Handler, M., & Morgan, C. (In press). Criterion validity of the Directed Lie Screening Test and the Empirical Scoring System with inexperienced examiners and non-naive examinees in a laboratory setting. *Polygraph*.

\*Nelson, R., Handler, M., Morgan, C., & O'Burke, P. (In press). Criterion validity of the United States Air Force Modified General Question Technique and Iraqi scorers. *Polygraph*.

\*Nelson, R., Handler, M., & Senter, S. (In press). Monte Carlo study of criterion validity of the Directed Lie Screening Test using the Empirical Scoring System and the Objective Scoring System version 3. *Polygraph*.

\*Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B., & Oelrich, M. (2011). Using the Empirical Scoring System. *Polygraph*, 40(2), 67–78.

\*Nelson, R. & Krapohl, D. (2011). Criterion validity of the Empirical Scoring System with experienced examiners: Comparison with the seven-position evidentiary model using the Federal Zone Comparison Technique. *Polygraph*, 40, 79–85.

\*Nelson, R., Krapohl, D., & Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. *Polygraph*, 37, 185–215.

Office of Technology Assessment (1983). *The validity of polygraph testing: A research review and evaluation*. Washington, D.C.: U.S. Congress, Office of Technology Assessment.

- Patrick, C. J. & Iacono, W. G. (1989). Psychopathy, threat and polygraph test accuracy. *Journal of Applied Psychology*, 74, 347–355.
- Patrick, C. J. & Iacono, W. G. (1991). Validity of the control question polygraph test: The problem of sampling bias. *Journal of Applied Psychology*, 76, 229–238.
- Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344–359.
- Podlesny, J., Raskin, D., & Barland, G. (1976). Effectiveness of Techniques and Physiological Measures in the Detection of Deception. Report No. 76–5, Contract 75–N1-99-001 LEAA (available through Department of Psychology, University of Utah, Salt Lake City).
- Podlesny, J. A. & Truslow, C. M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology*, 78, 788–797.
- Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J. & Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *Journal of applied psychology*, 89, 1099–1105.
- Raskin, D. C. & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15, 126–136.
- Raskin, D. C. & Honts, C. R. (2002). Handbook of polygraph testing. In M. Kleiner (Ed.), *Handbook of Polygraph Testing*. San Diego: Academic Press.
- Raskin, D. C. & Podlesny, J. A. (1979). Truth and deception: A reply to Lykken. *Psychological Bulletin*, 86, 54–59.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37, 542–547. Reprinted in *Polygraph* 11, 17–21.
- Reid, J. E. & Inbau, F. E. (1977). Truth and deception: The polygraph ('lie detector') technique.
- \*Research Division Staff (1995a). A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope Polygraph and the test for espionage and sabotage question formats. DTIC AD Number A319333. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 26(2), 79–106.
- \*Research Division Staff (1995b). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. DTIC AD Number A330774. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 27(3), 171–180.
- Research Division Staff (2001). Test of a mock theft scenario for use in the Psychophysiological Detection of Deception: IV. Report No. DoDPI00-R-0002. Department of Defense Polygraph Institute. Reprinted in *Polygraph* 30(4), 244–253.
- Rovner, L. I. (1986). Accuracy of physiological detection of deception for subjects with prior knowledge. *Polygraph*, 15(1), 1–39.
- Senter, S. M. (2003). Modified general question test decision rule exploration. *Polygraph*, 32, 251–263.
- Senter, S. M. & Dollins, A. B. (2002). New Decision Rule Development: Exploration of a two-stage approach. Report number DoDPI00-R-0001. Department of Defense Polygraph Institute Research Division, Fort Jackson, SC. Reprinted in *Polygraph* 37, 149–164.
- Senter, S. & Dollins, A. B. (2004). Comparison of question series and decision rules: A replication. *Polygraph*, 33, 223–233.
- Senter, S. M. & Dollins, A. B. (2008). Optimal decision rules for evaluating psychophysiological detection of deception data: an exploration. *Polygraph*, 37(2), 112–124.
- \*Senter, S., Waller, J., & Krapohl, D. (2008). Air Force Modified General Question Test validation study. *Polygraph*, 37(3), 174–184.

- Senter, S., Weatherman, D., Krapohl, D., & Horvath, F. (2010). Psychological set or differential salience: A proposal for reconciling theory and terminology in polygraph testing. *Polygraph*, 39 (2), 109–117.
- \*Shurani, T. (2011). Polygraph verification test. *European Polygraph*, 16.
- \*Shurani, T. & Chaves, F. (2010). Integrated Zone Comparison Technique and ASIT PolySuite algorithm: A field validity study. *European Polygraph*, 4(2), 71–80.
- \*Shurani, T., Stein, E. & Brand, E. (2009). A Field Study on the Validity of the Quadri-Track Zone Comparison Technique. *European Polygraph*, 1, 5–24.
- Slowik, S. M. & Buckley, J. P., III (1975). Relative accuracy of polygraph examiner diagnosis of respiration, blood pressure and GSR recordings. *Journal of Police Science and Administration*, 3, 305–309.
- Van Herk, M. (1990). Numerical evaluation: Seven point scale +/-6 and possible alternatives: A discussion. The Newsletter of the Canadian Association of Police Polygraphists, 7, 28-47. Reprinted in *Polygraph*, 20(2), 70–79.
- Verschuere, B., Meijer, E., & Merckelbach, H. (2008). The Quadri-Track Zone Comparison Technique: It's just not science. A critique to Mangan, Armitage, and Adams (2008). *Physiology and Behavior*, 1–2, 27–28.
- Wicklander, D. E. & Hunter, F. L. (1975). The influence of auxiliary sources of information in polygraph diagnosis. *Journal of Police Science and Administration*, 3, 405–409.

## Розміри вибірки включених наукових досліджень

Поліграфологічні методики	Наукові дослідження	Total N	N Deceptive	N Truthful	Total Scores	Deceptive Scores	Truthful Scores	Scorers
AFMGQT (7-position)	Senter, Waller & Krapohl (2008) <sup>1</sup>	69	33	36	69	33	36	1
AFMGQT (7-position)	Nelson, Handler, Morgan & O'Burke (In press) <sup>2</sup>	22	11	11	66	33	33	3
AFMGQT (7-position)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	-	-	-	100	50	50	1
AFMGQT (ESS)	Nelson, Blalock & Handler (2011) <sup>2</sup>	-	-	-	66	33	33	3
AFMGQT (ESS)	Nelson & Blalock (In press) <sup>1</sup>	-	-	-	69	33	36	1
AFMGQT (ESS)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	100	50	50	100	50	50	1
Backster You-Phase (Backster)	Nelson, Handler, Adams & Backster (In press) <sup>4</sup>	22	11	11	154	77	77	7
Backster You-Phase (Backster)	Nelson (In press)	100	50	50	100	50	50	1
CIT	MacLaren 2001	1,070	666	404	1,070	666	404	39
DLST/TES (7-position)	Research Division Staff 1995a	94	26	68	94	26	68	3
DLST/TES (7-position)	Research Division Staff 1995b	85	30	55	85	30	55	10
DLST/TES (7-position)	Nelson (In press) B	100	50	50	100	50	50	1
DLST/TES (7-position)	Nelson Handler Blalock & Hernández (In press) <sup>5C</sup>	49	25	24	98	50	48	2
DLST/TES (ESS)	Nelson & Handler (In press)	100	50	50	100	50	50	1
DLST/TES (ESS)	Nelson, Handler & Morgan (In press)	49	24	25	49	24	25	1
DLST/TES (ESS)	Nelson (In press) <sup>B</sup>	-	-	-	100	50	50	1
DLST/TES (ESS)	Nelson, Handler, Blalock & Hernández (In press) <sup>5C</sup>	-	-	-	98	50	48	2

Поліграфологічні методики	Наукові дослідження	Total N	N Deceptive	N Truthful	Total Scores	Deceptive Scores	Truthful Scores	Scorers
Federal You-Phase (7-position)	Nelson (2011) <sup>D</sup>	100	50	50	100	50	50	1
Federal You-Phase (7-position)	Nelson, Handler, Blalock & Cushman (in press) <sup>6 E</sup>	-	-	-	220	110	110	10
Federal You-Phase (ESS)	Nelson (2011) <sup>D</sup>	100	50	50	100	50	50	1
Federal You-Phase (ESS)	Nelson, Handler, Blalock & Cushman (in press) <sup>6 E</sup>	-	-	-	220	110	110	10
Federal ZCT (7-position)	Blackwell (1998)	100	65	35	300	195	105	3
Federal ZCT (7-position)	Krapohl & Cushman (2006) <sup>7 F</sup>	100	50	50	1,000	500	500	10
Federal ZCT (7-position)	Honts, Amato & Gordon (2004) as reported in Honts in Grahag (2004)	48	24	24	144	72	72	3
Federal ZCT (7-position evidentiary)	Krapohl & Cushman (2006) <sup>7 F</sup>	-	-	-	1,000	500	500	10
Federal ZCT (7-position evidentiary)	Nelson & Krapohl (2011) <sup>8 G</sup>	60	30	30	60	30	30	6
IZCT (Horizontal)	Shurani & Chavez (2010)	84	44	40	84	44	40	4
IZCT (Horizontal)	Gordon, Mohamed, Faro, Platek, Ahmad & Williams (2005)	11	6	5	11	6	5	1
IZCT (Horizontal)	Shurani (2011)	84	36	48	84	36	48	3
MQTZCT (Matte)	Matte & Reuss (1989) dissertation	122	64	58	122	64	58	2
MQTZCT (Matte)	Shurani, Stein & Brand (2009)	57	28	29	57	28	29	4
MQTZCT (Matte)	Mangan, Armitage & Adams (2008)	140	91	49	140	91	49	1
Utah-RCMP/CPC (Utah)	Honts, Hodes & Raskin (1985)	38	19	19	38	19	19	1
Utah-RCMP/CPC (Utah)	Driscoll, Honts & Jones, (1987)	40	20	20	40	20	20	1
Utah-RCMP/CPC (Utah)	Honts (1996)	32	21	11	32	21	11	1

Поліграфологічні методики	Наукові дослідження	Total N	N Deceptive	N Truthful	Total Scores	Deceptive Scores	Truthful Scores	Scorers
Utah-DLC (Utah)	Honts & Raskin (1988)	25	12	13	25	12	13	1
Utah-DLC (Utah)	Horowitz, Kircher, Honts & Raskin (1997)	30	15	15	30	15	15	1
Utah-DLC (Utah)	Kircher & Raskin (1988)	100	50	50	200	100	100	2
Utah-DLC (Utah)	Honts, Raskin & Kircher (1987)	20	10	10	20	10	10	1
ZCT (ESS)	Nelson, Krapohl & Handler (2008) <sup>7</sup>	-	-	-	700	350	350	7
ZCT (ESS)	Nelson, Blalock, Oelrich & Cushman (2011) <sup>7</sup>	-	-	-	250	150	100	25
ZCT (ESS)	Nelson & Krapohl (2011) <sup>8 G</sup>	-	-	-	60	30	30	6
ZCT (ESS)	Nelson et al (2011)	572	304	268	1,382	741	641	74
ZCT (ESS)	Blalock, Cushman & Nelson (2009) <sup>7</sup>	-	-	-	900	450	450	9
ZCT (ESS)	Handler, Nelson, Goodson & Hicks (2010) <sup>7</sup>	-	-	-	1,900	950	950	19

<sup>18</sup> Оцінювання вибірки на основі одних і тих самих випадків вибірки.

<sup>A.G</sup> Оцінювання вибірки, опубліковані в тому ж науковому дослідженні.

## Критеріальна точність включених наукових досліджень

Поліграфологічні методики	Наукові дослідження	Sens.	Spec.	FN	FP	D-INC	T-INC	Unweighted Accuracy	Unweighted INC
AFMGQT (7-position)	Senter, Waller & Krapohl (2008) <sup>1</sup>	.758	.917	.212	.083	.030	.001	.849	.015
AFMGQT (7-position)	Nelson, Handler, Morgan & O'Burke (In press) <sup>2</sup>	.818	.364	.001	.333	.182	.303	.761	.242
AFMGQT (7-position)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	.780	.420	.040	.200	.140	.420	.814	.280
AFMGQT (ESS)	Nelson, Blalock & Handler (2011) <sup>2</sup>	.831	.616	.010	.175	.158	.208	.883	.183
AFMGQT (ESS)	Nelson & Blalock (In press) <sup>1</sup>	.511	.862	.211	.027	.277	.028	.839	.152
AFMGQT (ESS)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	.806	.639	.067	.131	.127	.229	.876	.178
Backster You-Phase (Backster)	Nelson, Handler, Adams & Backster (In press) <sup>4</sup>	.943	.543	.009	.274	.048	.183	.828	.116
Backster You-Phase (Backster)	Nelson (In press)	.668	.592	.019	.079	.313	.329	.927	.321
CIT	MacLaren 2001	.815	.832	.185	.168	.001	.001	.823	.000
DLST/TES (7-position)	Research Division Staff 1995a	.654	.676	.154	.206	.192	.118	.788	.155
DLST/TES (7-position)	Research Division Staff 1995b	.833	.909	.167	.073	.000	.018	.880	.009
DLST/TES (7-position)	Nelson (In press) B	.910	.677	.037	.184	.053	.139	.874	.096
DLST/TES (7-position)	Nelson Handler Blalock & Hernández (In press) <sup>5C</sup>	.583	.940	.271	.020	.145	.039	.831	.092
DLST/TES (ESS)	Nelson & Handler (In press)	.917	.587	.036	.253	.047	.160	.831	.104
DLST/TES (ESS)	Nelson, Handler & Morgan (In press)	.625	.950	.210	.040	.165	.010	.854	.088
DLST/TES (ESS)	Nelson (In press) <sup>B</sup>	.935	.730	.046	.195	.020	.075	.871	.048
DLST/TES (ESS)	Nelson, Handler, Blalock & Hernández (In press) <sup>5C</sup>	.665	.839	.207	.040	.126	.119	.859	.123



Поліграфологічні методики	Наукові дослідження	Sens.	Spec.	FN	FP	D-INC	T-INC	Unweighted Accuracy	Unweighted INC
Federal You-Phase (7-position)	Nelson (2011) <sup>D</sup>	.833	.417	.010	.138	.157	.444	.870	.301
Federal You-Phase (7-position)	Nelson, Handler, Blalock & Cushman (In press) <sup>E</sup>	.844	.730	.036	.171	.119	.097	.885	.108
Federal You-Phase (ESS)	Nelson (2011) <sup>D</sup>	.813	.729	.050	.126	.090	.102	.897	.096
Federal You-Phase (ESS)	Nelson, Handler, Blalock & Cushman (In press) <sup>E</sup>	.859	.770	.027	.143	.145	.325	.906	.235
Federal ZCT (7-position)	Blackwell (1998)	.923	.448	.015	.295	.062	.257	.793	.159
Federal ZCT (7-position)	Krapohl & Cushman (2006) <sup>F</sup>	.824	.560	.044	.180	.132	.260	.853	.196
Federal ZCT (7-position)	Honts, Amato & Gordon (2004) as reported in Honts in Grahmag (2004)	.917	.917	.001	.083	.083	.000	.958	.042
Federal ZCT (7-pos. evidentiary)	Krapohl & Cushman (2006) <sup>F</sup>	.792	.824	.122	.116	.086	.060	.872	.073
Federal ZCT (7-pos. evidentiary)	Nelson & Krapohl (2011) <sup>G</sup>	.933	.667	.000	.233	.067	.133	.870	.100
IZCT (Horizontal)	Shurani & Chavez (2010)	.955	.900	.023	.001	.023	.100	.988	.061
IZCT (Horizontal)	Gordon, Mohamed, Faro, Platek, Ahmad & Williams (2005)	.999	.800	.001	.001	.001	.200	.999	.100
IZCT (Horizontal)	Shurani (2011)	.999	.999	.001	.001	.001	.001	.999	.000
MQTZCT (Matte)	Matte & Reuss (1989) dissertation	.969	.914	.000	.001	.031	.086	.999	.059
MQTZCT (Matte)	Shurani, Stein & Brand (2009)	.929	1.000	.071	.001	.000	.001	.964	.000
MQTZCT (Matte)	Mangan, Armitage & Adams (2008)	.978	1.000	.001	.001	.022	.001	.999	.011
Utah-RCMP/CPC (Utah)	Honts, Hodes & Raskin (1985)	.895	.421	.001	.211	.105	.368	.833	.237
Utah-RCMP/CPC (Utah)	Driscoll, Honts & Jones, (1987)	.900	.900	.001	.001	.100	.100	.999	.100
Utah-RCMP/CPC (Utah)	Honts (1996)	.714	.818	.048	.001	.238	.182	.969	.210

Поліграфологічні методики	Наукові дослідження	Sens.	Spec.	FN	FP	D-INC	T-INC	Unweighted Accuracy	Unweighted INC
Utah-DLC (Utah)	Honts & Raskin (1988)	.917	.846	.083	.001	.001	.154	.958	.077
Utah-DLC (Utah)	Horowitz, Kircher, Honts & Raskin (1997)	.733	.867	.133	.133	.133	.001	.856	.067
Utah-DLC (Utah)	Kircher & Raskin (1988)	.880	.860	.060	.060	.060	.080	.935	.070
Utah-DLC (Utah)	Honts, Raskin & Kircher (1987)	.800	.700	.001	.200	.200	.100	.889	.150
ZCT (ESS)	Nelson, Krapohl & Hanlder (2008) <sup>7</sup>	.749	.814	.154	.077	.097	.109	.872	.103
ZCT (ESS)	Nelson, Blalock, Oelrich & Cushman (2011) <sup>7</sup>	.793	.930	.073	.001	.133	.070	.958	.102
ZCT (ESS)	Nelson & Krapohl (2011) <sup>8 G</sup>	.833	.633	.001	.133	.167	.233	.913	.200
ZCT (ESS)	Nelson et al (2011)	.863	.789	.047	.093	.103	.107	.921	.105
ZCT (ESS)	Blalock, Cushman & Nelson (2009) <sup>7</sup>	.773	.727	.122	.102	.104	.171	.870	.138
ZCT (ESS)	Handler, Nelson, Goodson & Hicks (2010) <sup>7</sup>	.865	.881	.103	.089	.040	.039	.901	.040

<sup>1-8</sup> Оцінювання вибірки на основі одних і тих самих випадків вибірки.

<sup>A-G</sup> Оцінювання вибірки, опубліковані в тому ж науковому дослідженні.

## Статистика надійності для включених досліджень

Поліграфологічні методики	Наукові дослідження	Fleiss' Карра	Decision Agreement	Correlation
AFMGQT (7-position)	Senter, Waller & Krapohl (2008) <sup>1</sup>	.750	.930	.940
AFMGQT (7-position)	Nelson, Handler, Morgan & O'Burke (In press) <sup>2</sup>	-	1.000	-
AFMGQT (7-position)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	-	-	-
AFMGQT (ESS)	Nelson, Blalock & Handler (2011) <sup>2</sup>	-	1.000	.931
AFMGQT (ESS)	Nelson & Blalock (In press) <sup>1</sup>	-	-	-
AFMGQT (ESS)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	-	-	-
Backster You-Phase (Backster)	Nelson, Handler, Adams & Backster (In press) <sup>4</sup>	-	-	.567
Backster You-Phase (Backster)	Nelson (In press)	-	-	-
CIT	MacLaren 2001	-	-	-
DLST/TES (7-position)	Research Division Staff 1995a	.760	.890	-
DLST/TES (7-position)	Research Division Staff 1995b	-	-	-
DLST/TES (7-position)	Nelson (In press) B	-	-	-
DLST/TES (7-position)	Nelson Handler Blalock & Hernández (In press) <sup>5C</sup>	-	.722	-
DLST/TES (ESS)	Nelson & Handler (In press)	-	-	-
DLST/TES (ESS)	Nelson, Handler & Morgan (In press)	-	.911	-
DLST/TES (ESS)	Nelson (In press) <sup>B</sup>	-	-	-
DLST/TES (ESS)	Nelson, Handler, Blalock & Hernández (In press) <sup>5C</sup>	-	.769	-
Federal You-Phase (7-position)	Nelson (2011) <sup>D</sup>	-	-	-
Federal You-Phase (7-position)	Nelson, Handler, Blalock & Cushman (In press) <sup>6E</sup>	-	.852	-

Поліграфологічні методики	Наукові дослідження	Fleiss' Каппа	Decision Agreement	Correlation
Federal You-Phase (ESS)	Nelson (2011) <sup>D</sup>	-	-	-
Federal You-Phase (ESS)	Nelson, Handler, Blalock & Cushman (In press) <sup>6E</sup>	-	.897	-
Federal ZCT (7-position)	Blackwell (1998)	.570	.800	-
Federal ZCT (7-position)	Krapohl & Cushman (2006) <sup>7F</sup>	-	-	-
Federal ZCT (7-position)	Honts, Amato & Gordon (2004) as reported in Honts inGrahmag (2004)	-	-	-
Federal ZCT (7-position evidentiary)	Krapohl & Cushman (2006) <sup>7F</sup>	-	.870	-
Federal ZCT (7-position evidentiary)	Nelson & Krapohl (2011) <sup>8G</sup>	-	-	-
IZCT (Horizontal)	Shurani & Chavez (2010)	-	-	-
IZCT (Horizontal)	Gordon, Mohamed, Faro, Platek, Ahmad & Williams (2005)	-	-	-
IZCT (Horizontal)	Shurani (2011)	-	-	-
MQTZCT (Matte)	Matte & Reuss (1989) dissertation	-	-	.990
MQTZCT (Matte)	Shurani, Stein & Brand (2009)	-	-	-
MQTZCT (Matte)	Mangan, Armitage & Adams (2008)	-	-	-
Utah-RCMP/CPC (Utah)	Honts, Hodes & Raskin (1985)	.480	.950	.880
Utah-RCMP/CPC (Utah)	Driscoll, Honts & Jones, 1987)	-	-	.860
Utah-RCMP/CPC (Utah)	Honts (1996)	-	.930	.910
Utah-DLC (Utah)	Honts & Raskin (1988)	-	-	.940
Utah-DLC (Utah)	Horowitz, Kircher, Honts & Raskin (1997)	-	-	.920
Utah-DLC (Utah)	Kircher & Raskin (1988)	.730	.990	.970
Utah-DLC (Utah)	Honts, Raskin & Kircher (1987)	.730	.960	-

Поліграфологічні методики	Наукові дослідження	Fleiss' Карра	Decision Agreement	Correlation
ZCT (ESS)	Nelson, Krapohl & Handler (2008) <sup>7</sup>	.610	-	-
ZCT (ESS)	Nelson, Blalock, Oelrich & Cushman (2011) <sup>7</sup>	-	.950	-
ZCT (ESS)	Nelson & Krapohl (2011) <sup>8 G</sup>	-	-	-
ZCT (ESS)	Nelson et al (2011)	-	-	-
ZCT (ESS)	Blalock, Cushman & Nelson (2009) <sup>7</sup>	.560	-	-
ZCT (ESS)	Handler, Nelson, Goodson & Hicks (2010) <sup>7</sup>	.590	-	.840

<sup>18</sup> Оцінювання вибірки на основі одних і тих самих випадків вибірки.  
<sup>A,G</sup> Оцінювання вибірки, опубліковані в тому ж науковому дослідженні.

## Середні значення (Means) та стандартні відхилення (Standard Deviations) критеріїв оцінювання для правди (T) та обману (D)

Поліграфологічні методики	Наукові дослідження	Mean D	StDev D	Mean T	StDev T
AFMGQT (7-position)	Senter, Waller & Krapohl (2008) <sup>1</sup>	-	-	-	-
AFMGQT (7-position)	Nelson, Handler, Morgan & O'Burke (In press) <sup>2</sup>	-2.995*	4.727*	2.365*	3.879*
AFMGQT (7-position)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	-2.827*	4.504*	3.556*	3.766*
AFMGQT (ESS)	Nelson, Blalock & Handler (2011) <sup>2</sup>	-3.850*	4.730*	4.530*	5.180*
AFMGQT (ESS)	Nelson & Blalock (In press) <sup>1</sup>	-2.000*	5.030*	3.420*	3.470*
AFMGQT (ESS)	Nelson, Handler, & Senter (In press) <sup>3A</sup>	-3.031*	4.535*	3.265*	3.661*
Backster You-Phase (Backster)	Nelson, Handler, Adams & Backster (In press) <sup>4</sup>	-19.649	6.482	3.612	10.010
Backster You-Phase (Backster)	Nelson (In press)	-12.460	8.353	6.820	10.572
CIT	MacLaren 2001	-	-	-	-
DLST/TES (7-position)	Research Division Staff 1995a	-	-	-	-
DLST/TES (7-position)	Research Division Staff 1995b	-	-	-	-
DLST/TES (7-position)	Nelson (In press) B	-2.418*	3.818*	2.653*	3.618*
DLST/TES (7-position)	Nelson Handler Blalock & Hernández (In press) <sup>5C</sup>	-1.833*	4.099*	3.670*	3.443*
DLST/TES (ESS)	Nelson & Handler (In press)	-2.442*	3.531*	2.086*	3.460*
DLST/TES (ESS)	Nelson, Handler & Morgan (In press)	-1.271*	3.131*	4.660*	2.299*
DLST/TES (ESS)	Nelson (In press) <sup>B</sup>	-3.031*	5.104*	3.265*	3.935*
DLST/TES (ESS)	Nelson, Handler, Blalock & Hernández (In press) <sup>5C</sup>	-1.781*	3.437*	3.636*	2.917*
Federal You-Phase (7-position)	Nelson (2011) <sup>D</sup>	-6.398	4.914	5.485	5.106
Federal You-Phase (7-position)	Nelson, Handler, Blalock & Cushman (In press) <sup>6E</sup>	-7.991	6.733	6.514	6.680
Federal You-Phase (ESS)	Nelson (2011) <sup>D</sup>	-6.685	6.881	6.735	6.045

Поліграфологічні методики	Наукові дослідження	Mean D	StDev D	Mean T	StDev T
Federal You-Phase (ESS)	Nelson, Handler, Blalock & Cushman (In press) <sup>6 E</sup>	-8.606	5.842	6.018	7.107
Federal ZCT (7-position)	Blackwell (1998)	-10.385	9.510	6.981	7.495
Federal ZCT (7-position)	Krapohl & Cushman (2006) <sup>7 F</sup>	-6.264	10.863	9.776	8.212
Federal ZCT (7-position)	Honts, Amato & Gordon (2004) as reported in Honts in Grahag (2004)	-8.420	6.837	6.640	9.187
Federal ZCT (7-position evidentiary)	Krapohl & Cushman (2006) <sup>7 F</sup>	-6.264	10.863	9.776	8.212
Federal ZCT (7-position evidentiary)	Nelson & Krapohl (2011) <sup>8 G</sup>	-9.600	7.356	6.926	10.709
IZCT (Horizontal)	Shurani & Chavez (2010)	-8.847	15.264	21.181	5.097
IZCT (Horizontal)	Gordon, Mohamed, Faro, Platek, Ahmad & Williams (2005)	-36.000	12.946	8.750	1.635
IZCT (Horizontal)	Shurani (2011)	-19.667	9.607	28.948	5.963
MQTZCT (Matte)	Matte & Reuss (1989) dissertation	-9.148 <sup>+</sup>	2.843 <sup>+</sup>	3.099 <sup>+</sup>	6.002 <sup>+</sup>
MQTZCT (Matte)	Shurani, Stein & Brand (2009)	-6.949 <sup>+</sup>	1.630 <sup>+</sup>	5.388 <sup>+</sup>	1.246 <sup>+</sup>
MQTZCT (Matte)	Mangan, Armitage & Adams (2008)	-10.037 <sup>+</sup>	2.995 <sup>+</sup>	7.190 <sup>+</sup>	3.189 <sup>+</sup>
Utah-RCMP/CPC (Utah)	Honts, Hodes & Raskin (1985)	-11.950	6.520	9.000	10.660
Utah-RCMP/CPC (Utah)	Driscoll, Honts & Jones, (1987)	-10.700	6.000	10.350	6.470
Utah-RCMP/CPC (Utah)	Honts (1996)	-15.000	5.564	8.170	5.270
Utah-DLC (Utah)	Honts & Raskin (1988)	-11.500	5.803	9.000	5.803
Utah-DLC (Utah)	Horowitz, Kircher, Honts & Raskin (1997)	-7.000	13.500	8.500	11.500
Utah-DLC (Utah)	Kircher & Raskin (1988)	-7.710	8.420	10.785	8.671
Utah-DLC (Utah)	Honts, Raskin & Kircher (1987)	-14.000	7.490	9.600	7.490
ZCT (ESS)	Nelson, Krapohl & Hanlder (2008) <sup>7</sup>	-9.606	9.743	9.162	8.564
ZCT (ESS)	Nelson, Blalock, Oelrich & Cushman (2011) <sup>7</sup>	-10.740	8.263	8.690	4.585

Поліграфологічні методики	Наукові дослідження	Mean D	StDev D	Mean T	StDev T
ZCT (ESS)	Nelson & Krapohl (2011) <sup>8 G</sup>	-11.833	7.764	6.000	9.592
ZCT (ESS)	Nelson et al (2011)	-11.354	9.392	7.373	9.270
ZCT (ESS)	Blalock, Cushman & Nelson (2009) <sup>7</sup>	-11.253	9.786	7.191	8.785
ZCT (ESS)	Handler, Nelson, Goodson & Hicks (2010) <sup>7</sup>	-7.953	10.017	11.212	8.619

<sup>1-8</sup> Оцінювання вибірки на основі одних і тих самих випадків вибірки.  
<sup>A-G</sup> Оцінювання вибірки, опубліковані в тому ж науковому дослідженні.  
<sup>\*</sup> Середні значення та стандартні відхилення приводяться як оцінювання по окремо взятому релевантному запитанню (subtotal scores) для окремих запитань.  
<sup>•</sup> Середні значення та стандартні відхилення приводяться як оцінювання по окремо взятому релевантному запитанню (subtotal scores) для окремих чартів тесту.



## AFMGQT / 7-позиційна система оцінювання

Наукові дослідження	Senter, Waller & Krapohl (2008)	Nelson, Handler, Morgan & O'Burke (In press)	Nelson & Handler (In press)
Sample N	69	22	100
N Deceptive	33	11	50
N Truthful	36	11	50
Scorers	12	3	1
D Scores	33	33	50
T Scores	36	33	50
Total Scores	69	66	100
Mean D	-2.000	-2.995	-2.827
StDev D	5.030	4.727	4.504
Mean T	3.420	2.365	3.556
StDev T	3.470	3.879	3.766
Reliability Kappa	.750	-	-
Reliability Agreement	.930	.999	-
Reliability Correlation	.940	-	-
Unweighted Average Accuracy	.849	.761	.814
Unweighted Inconclusives	.015	.242	.280
Sensitivity	.758	.818	.780
Specificity	.917	.364	.420
FN Errors	.212	.000	.040
FP Errors	.083	.333	.200
D-INC	.030	.182	.140
T-INC	.000	.303	.420
PPV	.901	.711	.796
NPV	.812	.999	.913
D Correct	.781	.999	.951
T Correct	.917	.522	.677

## AFMGQT / ESS

Наукові дослідження	Nelson, Blalock Handler (In press)	Nelson & Blalock (In press)	Nelson, Handler & Senter (In press)
Sample N	22	69	100
N Deceptive	11	33	50
N Truthful	11	36	50
Scorers	3	1	1
D Scores	33	33	50
T Scores	33	36	50
Total Scores	66	69	100
Mean D	-3.850	-2.000	-3.031
StDev D	4.730	5.030	4.535
Mean T	4.530	3.420	3.265
StDev T	5.180	3.470	3.661
Reliability Kappa	-	-	-
Reliability Agreement	.999	-	-
Reliability Correlation	.931	-	-
Unweighted Average Accuracy	.883	.839	.876
Unweighted Inconclusives	.183	.152	.178
Sensitivity	.831	.511	.806
Specificity	.616	.862	.639
FN Errors	.010	.211	.067
FP Errors	.175	.027	.131
D-INC	.158	.277	.127
T-INC	.208	.028	.229
PPV	.826	.951	.860
NPV	.984	.803	.905
D Correct	.988	.708	.923
T Correct	.779	.970	.830

## Backster You-Phase

Наукові дослідження	Nelson, Handler, Adams & Backster (In press)	Nelson (In press)
Sample N	22	100
N Deceptive	11	50
N Truthful	11	50
Scorers	7	1
D Scores	77	50
T Scores	77	50
Total Scores	154	100
Mean D	-19.649	-12.460
StDev D	6.482	8.353
Mean T	3.612	6.820
StDev T	10.010	10.572
Reliability Kappa	-	-
Reliability Agreement	-	-
Reliability Correlation	.567	-
Unweighted Average Accuracy	.825	.927
Unweighted Inconclusives	.117	.321
Sensitivity	.948	.668
Specificity	.532	.592
FN Errors	.001	.019
FP Errors	.286	.079
D-INC	.052	.313
T-INC	.182	.329
PPV	.768	.894
NPV	.999	.969
D Correct	.999	.972
T Correct	.650	.882

**Concealed Information Test / Guilty Knowledge Test**

Повідомляється МакЛареном (MacLaren, 2001)

MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *Journal of Applied Psychology*, 86, 674–683.

Результати повідомляються з всіма інформованими учасниками і тільки з тими інформованими учасниками, які також беруть участь в поведінкових діях

	Mean (St. Er.) {95% CI}	
	Informed/guilty and uninformed participants	All informed and uninformed participants
Number of studies	39	50
N Deceptive	666	843
N Truthful	404	404
Total N	1070	1243
Unweighted Accuracy	.823 (.011) {.801 to .846}	.795 (.043) {.711 to .880}
Unweighted Inconclusives	-	-
Sensitivity	.815 (.014) {.789 to .842}	.759 (.053) {.655 to .864}
Specificity	.832 (.019) {.795 to .868}	.832 (.068) {.698 to .965}
FN Errors	.185 (.014) {.158 to .211}	.241 (.053) {.136 to .345}
FP Errors	.168 (.019) {.132 to .205}	.168 (.068) {.035 to .302}
D-INC	-	-
T-INC	-	-
PPV	.889 (.011) {.868 to .909}	.904 (.041) {.824 to .984}
NPV	.732 (.021) {.69 to .774}	.623 (.075) {.477 to .770}
D Correct	.815 (.014) {.789 to .842}	.759 (.053) {.655 to .864}
T Correct	.832 (.019) {.795 to .868}	.832 (.068) {.698 to .965}

## Directed Lie Screening Test (TES) / 7-позиційна система оцінювання

Наукові дослідження	Research DivisionStaff (1995a)	Research DivisionStaff (1995b)	Nelson (In press)	Nelson, Handler, Blalock & Hernández (In press)
Sample N	94	85	100	49
N Deceptive	26	30	50	25
N Truthful	68	55	50	24
Scorers	3	10	1	2
D Scores	26	30	50	50
T Scores	68	55	50	48
Total Scores	94	85	100	98
Mean D	-	-	-2.418	-1.833
StDev D	-	-	3.818	4.099
Mean T	-	-	2.653	3.670
StDev T	-	-	3.618	3.443
Reliability Kappa	.760	-	-	-
Reliability Kappa	.890	-	-	.722
Reliability Agreement	-	-	-	-
Unweighted Average Accuracy	.788	.880	.874	.831
Unweighted Inconclusives	.155	.009	.096	.092
Sensitivity	.654	.833	.910	.583
Specificity	.676	.909	.677	.940
FN Errors	.154	.167	.037	.271
FP Errors	.206	.073	.184	.020
D-INC	.192	.001	.053	.145
T-INC	.118	.018	.139	.039
PPV	.761	.920	.832	.967
NPV	.815	.845	.948	.776
D Correct	.810	.833	.961	.683
T Correct	.767	.926	.786	.979

## Directed Lie Screening Test (TES) / ESS

Наукові дослідження	Nelson & Handler (In press)	Nelson, Handler & Morgan (In press)	Nelson (In press)	Nelson, Handler, Blalock & Hernández (In press)
Sample N	100	49	100	49
N Deceptive	50	24	50	25
N Truthful	50	25	50	24
Scorers	1	1	1	2
D Scores	50	24	50	50
T Scores	50	25	50	48
Total Scores	100	49	100	98
Mean D	-2.442	-1.271	-3.031	-1.781
StDev D	3.531	3.131	5.104	3.437
Mean T	2.086	4.660	3.265	3.636
StDev T	3.460	2.299	3.935	2.917
Reliability Kappa	-	-	-	-
Reliability Kappa	-	.911	-	.769
Reliability Agreement	-	-	-	-
Unweighted Average Accuracy	.831	.854	.871	.859
Unweighted Inconclusives	.104	.088	.048	.123
Sensitivity	.917	.625	.935	.665
Specificity	.587	.950	.730	.839
FN Errors	.036	.210	.046	.207
FP Errors	.253	.040	.195	.040
D-INC	.047	.165	.020	.126
T-INC	.160	.010	.075	.119
PPV	.784	.940	.827	.943
NPV	.942	.819	.941	.802
D Correct	.962	.749	.953	.763
T Correct	.699	.960	.789	.954

## Federal You-Phase / 7-позиційна система оцінювання

Наукові дослідження	Nelson (In press)	Nelson, Handler, Blalock & Cushman (In press)
Sample N	100	22
N Deceptive	50	11
N Truthful	50	11
Scorers	1	10
D Scores	50	110
T Scores	50	110
Total Scores	100	220
Mean D	-6.398	-7.991
StDev D	4.914	6.733
Mean T	5.485	6.514
StDev T	5.106	6.680
Reliability Kappa	-	-
Reliability Kappa	-	.852
Reliability Agreement	-	-
Unweighted Average Accuracy	.870	.885
Unweighted Inconclusives	.301	.108
Sensitivity	.833	.844
Specificity	.417	.730
FN Errors	.010	.036
FP Errors	.138	.171
D-INC	.157	.119
T-INC	.444	.097
PPV	.858	.832
NPV	.977	.953
D Correct	.988	.959
T Correct	.751	.810

## Federal You-Phase / ESS

Наукові дослідження	Nelson (In press)	Nelson, Handler, Blalock & Cushman (In press)
Sample N	100	22
N Deceptive	50	11
N Truthful	50	11
Scorers	1	10
D Scores	50	110
T Scores	50	110
Total Scores	100	220
Mean D	-6.685	-8.606
StDev D	6.881	5.842
Mean T	6.735	6.018
StDev T	6.045	7.107
Reliability Kappa	-	-
Reliability Kappa	-	0.9
Reliability Agreement	-	-
Unweighted Average Accuracy	.897	.906
Unweighted Inconclusives	.096	.235
Sensitivity	.813	.859
Specificity	.729	.770
FN Errors	.050	.027
FP Errors	.126	.143
D-INC	.090	.145
T-INC	.102	.325
PPV	.866	.857
NPV	.936	.966
D Correct	.942	.970
T Correct	.853	.843



## Federal ZCT / 7-позиційна система оцінювання

Наукові дослідження	Blackwell (1998)	Krapohl & Cushman (2006)	Honts, Amato & Gordon (2004) as reported in Grahnag(2004)
Sample N	100	100	48
N Deceptive	65	50	24
N Truthful	35	50	24
Scorers	3	10	3
D Scores	195	500	72
T Scores	105	500	72
Total Scores	300	1,000	144
Mean D	-10.385	-6.264	-8.420
StDev D	9.510	10.863	6.837
Mean T	6.981	9.776	6.640
StDev T	7.495	8.212	9.187
Reliability Kappa	.570	-	-
Reliability Kappa	.800	-	-
Reliability Agreement	-	-	-
Unweighted Average Accuracy	.793	.852	.958
Unweighted Inconclusives	.159	.198	.042
Sensitivity	.923	.824	.917
Specificity	.448	.556	.917
FN Errors	.015	.044	.000
FP Errors	.295	.180	.083
D-INC	.062	.132	.083
T-INC	.257	.264	.001
PPV	.758	.821	.917
NPV	.967	.927	.999
D Correct	.984	.949	.999
T Correct	.603	.755	.917

**Federal ZCT / 7-позиційна система оцінювання з доказовим правилами  
(Evidentiary Rules)**

Наукові дослідження	Krapohl & Cushman (2006)	Nelson & Krapohl (2011)
Sample N	100	60
N Deceptive	50	30
N Truthful	50	30
Scorers	10	6
D Scores	500	30
T Scores	500	30
Total Scores	1,000	60
Mean D	-6.264	-9.600
StDev D	10.863	7.356
Mean T	9.776	6.926
StDev T	8.212	10.709
Reliability Kappa	-	-
Reliability Agreement	0.870	-
Reliability Correlation	-	-
Unweighted Average Accuracy	.872	.870
Unweighted Inconclusives	.073	.100
Sensitivity	.792	.933
Specificity	.824	.667
FN Errors	.122	.001
FP Errors	.116	.233
D-INC	.086	.067
T-INC	.060	.133
PPV	.872	.800
NPV	.871	.999
D Correct	.999	.999
T Correct	0.88	.741

**Integrated Zone Comparison Technique / Горизонтальна система оцінювання  
(Horizontal Scoring System)**

Наукові дослідження	Gordon, Mohamed, Faro, Platek, Ahmad & Williams (2005)	Shurani & Chaves (2010)	Shurani (2011)
Sample N	11	84	84
N Deceptive	6	44	36
N Truthful	5	40	48
Scorers	1	4	1
D Scores	6	44	36
T Scores	5	40	48
Total Scores	11	84	84
Mean D	-36.000	-8.847	-19.667
StDev D	12.946	15.264	9.607
Mean T	8.750	21.181	28.948
StDev T	1.635	5.097	5.963
Reliability Kappa	-	-	-
Reliability Agreement	-	-	-
Reliability Correlation	-	-	-
Unweighted Average Accuracy	.999	.988	.999
Unweighted Inconclusives	.100	.061	.001
Sensitivity	.999	.955	.999
Specificity	.800	.900	.999
FN Errors	.001	.023	.001
FP Errors	.001	.001	.001
D-INC	.001	.023	.001
T-INC	.200	.100	.001
PPV	.999	.999	.999
NPV	.999	.975	.999
D Correct	.999	.977	.999
T Correct	.999	.999	.999

## Matte Quadri-Track Zone Comparison Technique

Наукові дослідження	Matte & Reuss (1989)	Shurani, Stein & Brand (2009)	Mangan, Armitage & Adams (2008)
Sample N	122	57	140
N Deceptive	64	28	91
N Truthful	58	29	49
Scorers	2	4	1
D Scores	64	28	91
T Scores	58	29	49
Total Scores	122	57	140
Mean D	-9.148	-6.949*	-10.037*
StDev D	2.843	1.630*	2.995*
Mean T	3.099	5.388*	7.190*
StDev T	6.002	1.246*	3.189*
Reliability Kappa	-	-	-
Reliability Agreement	-	-	-
Reliability Correlation	.990	-	-
Unweighted Average Accuracy	.999	.964	.999
Unweighted Inconclusives	.059	.001	.011
Sensitivity	.969	.929	.978
Specificity	.914	.999	.999
FN Errors	.001	.071	.001
FP Errors	.001	.001	.001
D-INC	.031	.001	.022
T-INC	.086	.001	.001
PPV	.999	.999	.999
NPV	.999	.933	.999
D Correct	.999	.929	.999
T Correct	.999	.999	.999

## Utah PLC / Числова система оцінювання Utah

Наукові дослідження	Kircher & Raskin (1988)	Honts, Raskin & Kircher (1987)
Sample N	100	20
N Deceptive	50	10
N Truthful	50	10
Scorers	1	1
D Scores	50	10
T Scores	50	10
Total Scores	100	20
Mean D	-7.710	-14.000
StDev D	8.420	7.490
Mean T	10.785	9.600
StDev T	8.671	7.490
Reliability Kappa	.730	.730
Reliability Agreement	.990	.960
Reliability Correlation	.970	-
Unweighted Average Accuracy	.935	.889
Unweighted Inconclusives	.070	.150
Sensitivity	.880	.800
Specificity	.860	.700
FN Errors	.060	.000
FP Errors	.060	.200
D-INC	.060	.200
T-INC	.080	.100
PPV	.936	.800
NPV	.935	.999
D Correct	.936	.999
T Correct	.935	.778

## Utah DLC / Числова система оцінювання Utah

Наукові дослідження	Honts & Raskin (1988)	Horowitz, Kircher, Honts & Raskin (1997)
Sample N	25	30
N Deceptive	12	15
N Truthful	13	15
Scorers	1	1
D Scores	12	15
T Scores	13	15
Total Scores	25	30
Mean D	-11.500	-7.000
StDev D	5.803	13.500
Mean T	9.000	8.500
StDev T	5.803	11.500
Reliability Kappa	-	-
Reliability Agreement	-	-
Reliability Correlation	.940	.920
Unweighted Average Accuracy	.958	.856
Unweighted Inconclusives	.077	.067
Sensitivity	.917	.733
Specificity	.846	.867
FN Errors	.083	.133
FP Errors	.001	.133
D-INC	.001	.133
T-INC	.154	.001
PPV	.999	.846
NPV	.910	.867
D Correct	.917	.846
T Correct	.999	.867

## Utah RCMP Zone / Числова система оцінювання Utah

Наукові дослідження	Honts Hodes & Raskin (1985)	Honts (1996)	Driscoll, Honts & Jones, (1987)
Sample N	38	32	40
N Deceptive	19	21	20
N Truthful	19	11	20
Scorers	1	1	1
D Scores	19	21	20
T Scores	19	11	20
Total Scores	38	32	40
Mean D	-11.950	-15.000	-10.700
StDev D	6.520	5.564	6.000
Mean T	9.000	8.170	10.350
StDev T	10.660	5.270	6.470
Reliability Kappa	.480	-	-
Reliability Agreement	.950	.930	-
Reliability Correlation	.880	.910	.860
Unweighted Average Accuracy	.833	.969	.999
Unweighted Inconclusives	.237	.210	.100
Sensitivity	.895	.714	.900
Specificity	.421	.818	.900
FN Errors	.001	.048	.001
FP Errors	.211	.001	.001
D-INC	.105	.238	.100
T-INC	.368	.182	.100
PPV	.810	.999	.999
NPV	.999	.945	.999
D Correct	.999	.938	.999
T Correct	.667	.999	.999

## Utah PLT DLC Combined / Числова система оцінювання Utah

Методика	Utah PLC	Utah DLC	RCMP
Sample N	120	55	110
N Deceptive	60	27	60
N Truthful	60	28	50
Scorers	2	2	3
D Scores	60	27	60
T Scores	60	28	50
Total Scores	120	55	110
Mean D	-10.855	-9.250	-12.550
StDev D	7.955	9.652	6.028
Mean T	10.193	8.750	9.173
StDev T	8.081	8.652	7.467
Reliability Kappa	.730	-	.480
Reliability Agreement	.975	-	.940
Reliability Correlation	.970	.930	.883
Unweighted Average Accuracy	.927	.902	.939
Unweighted Inconclusives	.083	.073	.185
Sensitivity	.867	.815	.833
Specificity	.833	.857	.700
FN Errors	.050	.111	.017
FP Errors	.083	.071	.080
D-INC	.083	.074	.150
T-INC	.083	.071	.220
PPV	.912	.919	.912
NPV	.943	.885	.977
D Correct	.945	.880	.980
T Correct	.909	.923	.897



## Zone Comparison Techniques / ESS

Наукові дослідження	Nelson, Krapohl & Handler (2008)	Nelson, Blalock, Cushman & Oelrich (2011)	Nelson & Krapohl (2011)	Nelson et al. (2011)	Blalock, Cushman & Nelson (2009)	Handler, Nelson, Goodson & Hicks (2010)
Sample N	100	10	60	562	100	100
N Deceptive	50	6	30	298	50	50
N Truthful	50	4	30	264	50	50
Scorers	7	25	6	74	9	19
D Scores	350	150	30	741	450	950
T Scores	350	100	30	641	450	950
Total Scores	700	250	60	1,382	900	1,900
Mean D	-9.634	-10.740	-11.833	-11.354	-11.253	-7.953
StDev D	8.475	8.263	7.764	9.392	9.786	10.017
Mean T	8.849	8.690	6.000	7.373	7.191	11.212
StDev T	7.457	4.585	9.592	9.270	8.785	8.619
Reliability Kappa	.610	-	-	-	.560	.590
Reliability Agreement	-	.950	-	-	-	-
Reliability Correlation	-	-	-	-	-	.840
Unweighted Average Accuracy	.872	.958	.913	.883	.870	.867
Unweighted Inconclusives	.103	.102	.200	.114	.138	.115
Sensitivity	.749	.793	.833	.804	.773	.666
Specificity	.814	.930	.633	.761	.727	.873
FN Errors	.154	.073	.001	.090	.122	.196
FP Errors	.077	.001	.133	.117	.102	.035
D-INC	.097	.133	.167	.105	.104	.138
T-INC	.109	.070	.233	.122	.171	.092
PPV	.907	.999	.862	.873	.883	.950
NPV	.841	.927	.999	.894	.856	.817
D Correct	.829	.916	.999	.899	.864	.773
T Correct	.914	.999	.826	.867	.877	.961

**Поліграфологічні методики, для яких не вдалося знайти опубліковані дослідження або для яких жодні опубліковані дослідження не могли бути включені в Мета-аналіз.**

- Backster Exploratory
- Backster SKY
- Backster ZCT
- IZCT Screening
- Law Enforcement Pre-Employment Test (LEPET)\*
- Marcy Technique
- Matte Quinque Track
- Matte SGK
- RCMP B Series
- Searching Peak of Tension
- Utah MGQT\*

\* Хоча немає опублікованих досліджень для опису цих методик, формати LEPET і Utah MGQT структурно майже ідентичні AFMGQT. Дані валідності для AFMGQT можна узагальнити для цих методик, якщо їх оцінити за допомогою моделей TDA, описаних в опублікованих дослідженнях.

### Наукові дослідження, які не були включені в Мета-аналіз

Ці дослідження не були виключені з різних причин, включаючи використання приладів, які не відповідають тим, що використовуються в практиці, використання послідовностей тестових запитань або моделей аналізу тестових даних, які не відповідають протоколам польових досліджень, нерепрезентативні вибірки досліджень, або недостатньо статистичних даних для розрахунку профілю критеріальної точності або розподілу вибірки.

Анслі (Ansley, 1989) провів дослідження ефективності скринінгового тесту Relevant-Irrelevant за участю студентів-спеціалістів. Крюсон (Crewson, 2001) включив результати цього дослідження в опитування поліграфологів, медиків і психологів щодо діагностичних і скринінгових тестів. Однак звіт про наукове дослідження недоступний, і його не вдалося включити до Мета-аналізу.

Барланд, Хонтс і Баргер (Barland, Honts & Barger, 1989) вивчали точність багатотемних скринінгових досліджень. Немає даних щодо надійності, а також щодо середніх значень вибірки чи стандартних відхилень. Дані недоступні для подальшого аналізу. Числове оцінювання було виконано відповідно до попередніх протоколів навчання та польової практики Міністерства оборони США та може не відображати поточну практику. Без можливості порівняти ці результати з результатами інших досліджень цю серію досліджень не можна було включити в Мета-аналіз.

Белл, Кірчер, Бернхардт (Bell, Kircher & Bernhardt, 2008) порівняли методи Utah PLC та DLC та дійшли висновку, що суттєвих відмінностей немає. TDA був обмежений автоматизованими методами, тому не був включений у мета-аналіз.

Браунлі, Джонсон та Найл (Brownlie, Johnson & Knill, 1997) завершили дослідження методики RI. Дослідження описано в звіті NRC (2003) і Крюсона (Crewson, 2001). Крюсон (Crewson, 2001) включив результати цього дослідження в опитування поліграфологів, медиків і психологів щодо діагностичних і скринінгових тестів. Однак звіт про дослідження недоступний, і дослідження не вдалося включити до Мета-аналізу.

Корреа і Адамс (Correa & Adams, 1981) повідомили про результати дослідження методики RI. Процедура тестування не відповідає польовій практиці та передбачає використання термисторного давача дихання замість стандартних торакальних та абдомінальних пневмодавачів. Крім того, замість давачів кардіографа використовувався ЕКГ. Результати, як повідомляється, були 100% точними. Не вдалося отримати дані для огляду.

Форман та МакКаулі (Forman & McCauley, 1986) повідомили про результати дослідження методики позитивного контролю (positive-control technique). Повідомлені результати не надають середньостатистичні дані або стандартне відхилення для розподілу балів. Крім того, у дослідженні використовувалася квазічислова система оцінювання (quasi-numerical scoring system), яка не відображає польову практику РСТ, як описано в інших дослідженнях.

Гангалі, Ларі та Бхасеін (Ganguly, Lahri & Bhaseen, 1986) повідомили про результати дослідження, заснованого на методиці Піда (Reid Technique). Однак процедури тестування не відображають польову практику, оскільки під час дослідження оцінювалися лише дані пневмографа та кардіографа.

Гінтон, Дайе, Елаад, Бен-Шахар (Ginton, Daie, Elaad & Ben-Shakhar, 1982) повідомили про результати цікавого польового дослідження, яке включало унікальну

модифікацію методики MGQT, використовуючи запитання загальної правдивості (overall truth question) у позиції 3. Це дослідження не було включено, оскільки ця послідовність запитань не відображає польові практики.

Гордон, Флейшер, Морсі, Хабіб та Салах (Gordon, Fleisher, Morsie, Habib & Salah, 2000) повідомили про результати польового дослідження IZCT, включаючи 309 підтверджених випадків і 1 помилку. До публікації не було включено жодних даних щодо надійності, а також жодного статистичного опису розподілу вибірки оманливих і правдивих оцінок. Автори повідомили, що дані належать розвідувальній службі іноземного уряду, і основний автор повідомив спеціальний комітет (особисте повідомлення 10 червня 2011), що він завершив звіт про дослідження, навіть не побачивши даних.

Хонтс та Амато (Honts & Amato, 1999) повідомили про результати автоматизованої презентації поліграфологічної методики RI. Ця процедура не відображає польову практику, тому інформація не може бути включена до Мета-аналізу.

Хонтс і Ходес (Honts, Hodes, 1983) повідомили про ті ж результати, що й експеримент 1 у дослідженні Хонтса, Ходеса та Раскіна (Honts, Hodes & Raskin, 1985), який не включав стандартну серцево-судинну манжету.

Експеримент 1 Хонтса, Ходеса та Раскіна (Honts, Hodes & Raskin, 1985) включав методику Backster You-Phase, але не включав стандартну кардіо манжету, тому його не можна було включити в Мета-аналіз.

Хонтс і Реві (Honts & Reavy, 2009) використовували Federal ZCT у великомасштабному лабораторному експерименті. Однак це дослідження було розроблено для оцінки та порівняння ефективності PLCs та DLCs. Повідомлені дані не можна було використати для розрахунку розмірного профілю узагальнених критеріїв критеріальної точності.

Хоровіц, Кірчер, Хонтс та Раскін (Horowitz, Kircher, Honts & Raskin, 1997) включили результати дослідження методик RI. Результати RI не можна було включити, оскільки протокол підрахунку балів не відображав польову практику (глобальний аналіз і суб'єктивну оцінку послідовних і значущих реакцій) і передбачав використання семипозиційних числових процедур оцінювання, у яких реакції RQ порівнювалися з реакціями на нейтральні питання.

Хорвац (Horvath, 1988) повідомив про результати дослідження методики Ріда (Reid technique) за участю двох сліпих оцінювачів, які використовували 7-позиційну модель системи оцінювання, описану як подібну до тієї, яку використовували Барланд і Раскін (Barland & Raskin, 1975), з +/- 5 порогоми прийняття рішення. Це не метод підрахунку балів Ріда, який був описаний як трипозиційна модель TDA, яка не використовує фіксованих порогових значень. Вони також включили п'ятий RQ, що не відповідає поточній польовій практиці.

Хорвац і Палмат'єр (Horvath, Palmatier, 2008) повідомили про результати формату MGQT, структурованого як методика Ріда (Reid technique). У дослідженні брали участь два сліпі оцінювачі, які використовували 7-позиційну модель системи оцінювання, описану як подібну Раскіну (Raskin, 1975), з +/- 6 порогоми прийняття рішення. Це не метод підрахунку балів Ріда, який описано як трипозиційна модель TDA, яка не має порогових значень. Вони також включили п'ятий RQ, що не відповідає польовій практиці. Крім того, був лише один оцінювач, тому статистику надійності не можна було обчислити.

Хорвац і Рід (Horvath & Reid, 1971) повідомили про результати дослідження методики Ріда (Reid technique). Дослідження не можна було включити до мета-аналізу з кількох причин, у тому числі через високу вибірковість вибірки та відсутність

чітко структурованої моделі прийняття рішень. З початкових 75 поліграфологічних досліджень 40 було обрано автором для включення у вибірку дослідження, а 35 досліджень було вилучено з вибірки, оскільки автор вважав, що їх було занадто легко оцінити. Факт відбору досліджень з окремих файлів справ, які вони вважали невідповідними для дослідження, ставить під сумнів, чи буде отримана вибірка репрезентативною для реальних випадків. Спеціалістам було заборонено робити невизначені висновки, і вони повинні були надати висновок DI або NDI для кожного випадку. Деякі випадки мали п'ять RQ, що не сумісно з методами польового тестування, які зараз використовуються. Не було надано жодного середнього значення та стандартного відхилення, а також неможливо було надати дані для їх розрахунку. Крім того, не повідомлялося про статистику достовірності.

Хантер і Еш (Hunter & Ash 1973) повідомили про результати дослідження методики Ріда (Reid technique). Методика Ріда не використовує структуровану модель прийняття рішень або фіксовані показники, а покладається на імпресіоністські рішення спеціаліста. Звіт не містить жодної інформації про надійність між оцінювачами та не містить жодних даних розподілу вибірки, які можна використовувати для її порівняння. Крім того, вибірку дослідження було створено за допомогою високовибіркового процесу, що включає перевірені випадки, проведені основним автором. Без даних про надійність і без параметрів розподілу вибірки або доступу до даних дослідження вибірка є неперевіреною репрезентативністю та узагальненістю.

Джейн (Jaune, 1989) повідомив про результати польового дослідження прогностичної цінності поліграфологічних скринінгових тестів. Метою дослідження було порівняти точність скринінгового поліграфа з іншими скринінговими методами перевірки перед працевлаштуванням. На додаток до вимог щодо високовибіркового та не випадкового відбору випадків, критеріальний статус вибірових скринінгових випадків визначався в залежності від результатів наступного діагностичного поліграфологічного дослідження щодо розслідування крадіжки, що скоїв співробітник, яке було незалежним від скринінгового поліграфа.

Незважаючи на новаторську спробу вивчити майбутні результати, пов'язані зі скринінговим поліграфом, результати цього дослідження не підходять для використання в якості дослідження критеріальної точності.

Джейн (Jaune, 1990) повідомив про результати дослідження методики Ріда (Reid technique). Незважаючи на те, що результати були раніше описані Креполом (Krapohl, 2006), це дослідження не могло бути включене в цей мета-аналіз, тому що було використано два різних підходи оцінювання, жоден з яких не відображає польову практику. Один із методів оцінювання передбачав проведення точних лінійних вимірювань тестових даних. Інший, більш поширений, числовий метод оцінювання, був завершений, виключаючи четверте RQ, яке було описане як вторинне RQ, що також могло оцінюватися як CQ. Тесту, для якого характер і призначення стимулу визначаються постфактум, бракує наукової суворості. Сучасна польова практика не підтримує виключення окремих RQ або використання вторинного RQ як CQ. Крім того, у цьому дослідженні не використовувалися правила прийняття рішень або числові порогові значення, а рішення приймалися відповідно до суб'єктивної думки оцінювача.

Крепол (Krapohl, 2005) повідомив про результати семипозиційного доказового оцінювання досліджень Federal ZCT. Дані дослідження включали випадки з чотирьох різних архівних вибірок, для яких розподіл вибірки не можна було ефективно порівняти з розподілом вибірки з інших досліджень.

Крепол (Krapohl, 2010) показав, що семипозиційні бали досліджень You-Phase можна трансформувати в бали ESS. Вибіркові дані були високовибірковими та нерепрезентативними дослідженнями Backster You-Phase, наданими Мейроном, Креполом і Ешкеназі (Meiron, Krapohl & Ashkenazi, 2008).

Матте (Matte, 1990) – це підсумок дисертаційного дослідження, виконаного в установі, яка була схвалена штатом Каліфорнія, але не акредитована установою, визнаною урядом США, або іноземним еквівалентом. Пізніше UMI став частиною ProQuest, яка згодом прийняла політику, згідно з якою лише ті дисертації з регіональних акредитованих університетів будуть перераховані та доступні для громадськості. Дисертація підтримується ProQuest з Матте (Matte) як автором і ProQuest як видавцем. Дані цього дисертаційного дослідження були включені в цей мета-аналіз, оскільки ці матеріали були раніше опубліковані в журналі Polygraph (Matte & Ruess, 1989).

Матте (Matte, 2010) повідомив про результати дослідження правила Бакстера "або-або". Однак правдивих випадків до вибірки не було включено. Це дослідження не можна розглядати як дослідження критеріальної точності і не може бути включене в Мета-аналіз.

Мейрон, Крепол і Ешкеназі (Meiron, Krapohl & Ashkenazi, 2008) повідомили про результати дослідження правила Бакстера "або-або" (Backster Either-Or rule). Дані, як повідомлялося на щорічній конференції АПА 2008 року в Індіанapolisі, включали високовибіркову та невідповідну вибірку, в якій не враховувались результати проблемних досліджень. Результатом цієї форми високовибіркового відбору є те, що дані вибірки систематично позбавлені дисперсії помилок.

Петрік та Айконо (Patrick & Iacono, 1991) повідомили про результати дослідження Federal ZCT під час перегляду CQ між чартами. Прилади реєстрували як провідність шкіри, так і частоту серцевих скорочень, які порівнювалися з рівнями, які були до стимула таким чином, що не відображало польову практику.

Петрік та Айконо (Patrick & Iacono, 1989) у реплікації попереднього дослідження Раскіна і Харе (Raskin & Hare, 1978) повідомили про результати дослідження, яке включало методіку, яка нагадує Federal ZCT, використовуючи систему оцінювання Юта та правило прийняття рішення за загальним балом (grand total). Це наукове дослідження є цікавим, але не в повній мірі відображає польову практику, щоб включити його до поточного мета-аналізу критеріальної точності.

Подлесний і Раскін (Podlesny & Raskin, 1978) зареєстрували вісім різних фізіологічних каналів, використовуючи послідовність тестових запитань, яка нагадує Federal ZCT із запитанням про комплекс провини замість симптоматичного запитання в позиції 8. Числове оцінювання базувалося на методі, описаному Барландом і Раскіном (Barland & Raskin, 1975) і Раскіном і Харе (Raskin & Hare, 1978), який є ранньою версією числової системи оцінювання Utah. Були вивчені комплексні запитання провини та відмінності між виключаючими і невиключаючими CQ. Крім того, CQT порівнювали з GKT. Відсутність стандартних відхилень перешкоджає порівнянню вибіркового розподілу числового оцінювання з результатами інших наукових досліджень. Це дослідження було розроблено для аналізу ряду важливих запитань, але воно не могло вписатися в чітку категорію послідовності запитань тесту та моделі TDA, для якої можна було б знайти дослідження реплікації для порівняння. Як наслідок, це дослідження не можна було включити до мета-аналізу критеріальної точності поліграфологічних методик, які зараз використовуються в польових умовах.

Раскін і Харе (Raskin & Hare, 1978) повідомили про результати важливого дослідження за участю досліджуваних, яких вважали психопатами-злочинцями. Результати цього дослідження не можна було включити до критеріальної точності мета-аналізу польових поліграфологічних методик через використання нестандартного обладнання для тестування, яке не включало стандартний давач серцевої активності артеріального тиску.

Співробітники дослідницького відділу (2001) описали дослідження, яке використовувалося для розробки лабораторного сценарію щодо маніпулювання об'єктами дослідження. Це дослідження не використовувалося як дослідження щодо критеріальної точності.

Ровнер (Rovner, 1986) у цікавому дослідженні щодо контрзаходів поліграфу не повідомляв про рішення, помилки та невизначені результати окремо для правдивих і оманливих випадків. Дані недоступні для подальшого аналізу. Як наслідок, повний розмірний профіль критеріальної точності не можна було розрахувати, і дослідження не можна було включити як дослідження щодо критеріальної точності.

Сентер і Доллінс (Senter & Dollins, 2002) опублікували цікаве та інформативне дослідження щодо правил прийняття рішень, яке не підходить для використання в якості критеріального дослідження.

Сентер (Senter, 2003) опублікував цікаве та інформативне дослідження щодо правил прийняття рішень, яке не підходить для використання в якості критеріального дослідження.

Сентер і Доллінс (Senter & Dollins, 2004) опублікували цікаве та інформативне дослідження щодо правил прийняття рішень, яке не підходить для використання в якості критеріального дослідження.

Сентер і Доллінс (Senter & Dollins, 2008) опублікували цікаве та інформативне дослідження щодо правил прийняття рішень, яке не підходить для використання в якості критеріального дослідження.

Словік і Баклі (Slowik & Buckley, 1975) повідомили про результати дослідження методики Ріда (Reid technique). Вибірку дослідження було відібрано з перевірених випадків, але не описується процес перевірки (верифікації). Методика Ріда не використовує структуровану модель прийняття рішень або фіксовані показники, а покладається на імпресіоністські рішення дослідника. Звіт не містить жодної інформації про надійність між оцінювачами та не містить жодних параметрів розподілу вибірки, які можна використовувати для порівняння розподілу вибірки. Без статистики розподілу вибірки або доступу до даних дослідження вибірка має неперевірену репрезентативність і невідому можливість узагальнення.

Вен Херк (Van Herk, 1990) повідомив про результати пілотного дослідження та, можливо, першу публікацію про 3-позиційну TDA. Одна третина випадків була не підтвердженою, і, отже, це дослідження не могло бути включене в мета-аналіз.

Вікландер і Хантер (Wicklander & Hunter, 1975) повідомили про результати дослідження методики Ріда. Зразок дослідження було відібрано з перевірених випадків, але не описується процес перевірки (верифікації). Методика Ріда не використовує структуровану модель прийняття рішень або фіксовані показники, а покладається на імпресіоністські рішення спеціаліста. Звіт не містить жодної інформації про надійність між оцінювачами та не містить жодних даних розподілу вибірки, які можна використовувати для їх порівняння. Вибірка має неперевірену репрезентативність та невідомий порядок узагальнення без статистики розподілу або доступу до даних дослідження.

**Методики, щодо яких існує лише одне дослідження,  
яке відповідає якісним і кількісним критеріям для включення  
в Мета-аналіз**

Стандарти практики АПА вимагають мінімум два опубліковані дослідження щодо певної методики.

**Методика Арзера (Arther Technique)**

Хорвац (Horvath, 1977) повідомив про результати польового дослідження з використанням модифікації Арзером методики Ріда.

**Методика Ріда (Reid Technique)**

Хорвац (Horvath, 1988) повідомив про результати лабораторного дослідження з використанням методики Ріда.

*Раніше ці дослідження були об'єднані Креполом (Krapohl, 2006). Проте подальший огляд свідчить про те, що методики Ріда та Арзера є досить несхожими, тому результати цих досліджень не можна вважати такими, що повторюють один одного.*

Методика Ріда відрізняється від інших форматів CQT кількома важливими моментами. Іноді використовується п'ятий RQ, тоді як усі інші поліграфологічні методики обмежені чотирма RQ. Крім того, методика Ріда не використовує фіксованих числових порогових значень або структурованої моделі прийняття рішень. Натомість спеціаліст приймає рішення імпресіоністично, використовуючи інформацію з даних тестування, попереднього інтерв'ю, спостережень за поведінкою та інформації з матеріалів справи. Включення клінічних вражень у процес прийняття рішення суттєво обмежує можливість підтвердити його так, як це було зроблено з іншими методиками. Хоча методику Ріда можна вважати джерелом багатьох важливих інновацій і вона справді є джерелом, з якого розвинулися всі інші CQT, наразі її не викладають у жодній поліграфологічній школі, акредитованій АПА. Кількість практиків, які використовують цю методику, значно зменшилася після закриття Поліграфологічної школи Ріда. Більшість інших методик у мета-аналізі зараз викладаються та/або практикуються у значному масштабі. Огляд дослідження на підтримку методики виявив неможливість задовольнити вимоги критеріїв відбору дослідження, включаючи статистику надійності між оцінювачами та нормативні параметри, за допомогою яких можна розрахувати можливість узагальнення даних вибірки. Результати, описані в опублікованій літературі, і дані, доступні комітету, не дозволяють статистичну обробку, застосовану до всіх інших методів. Незважаючи на ці обмеження, середній рівень точності досліджень за методикою Ріда істотно не відрізнявся від результатів цього мета-аналізу.



**AFMGQT / 3-позиційна система оцінювання**

Два придатних для використання дослідження AFMGQT із 3-позиційною системою оцінювання дали сукупний незважений середній рівень точності .816 (.059) з невизначеним показником .443 (.044). Середньозважене значення правильних рішень для 3-позиційної моделі з AFMGQT становило .989 (.016) для критеріальних оманливих випадків і .643 (.116) для критеріальних правдивих випадків. Середньозважені невизначені показники становили .237 (.060) для критеріальних оманливих випадків і .648 (.067) для критеріальних випадків правди

Наукові дослідження	Nelson & Handler (In press)	Handler & Nelson (In press)
Sample N	100	22
N Deceptive	50	11
N Truthful	50	11
Scorers		3
D Scores	50	33
T Scores	50	33
Total Scores	100	66
Mean D	-1.886	-1.903
StDev D	3.161	2.986
Mean T	2.427	1.424
StDev T	2.557	2.722
Reliability Kappa	-	-
Reliability Agreement	-	.945
Reliability Correlation	-	-
Unweighted Accuracy	.869	.740
Unweighted Inconclusives	.457	.421
Sensitivity	.737	.780
Specificity	.260	.180
FN Errors	.007	.010
FP Errors	.088	.186
D-INC	.256	.209
T-INC	.658	.633
PPV	.894	.807
NPV	.974	.947
D Correct	.991	.987
T Correct	.748	.492

**Арму MGQT / 7-позиційна система оцінювання**

Два придатних для використання дослідження армійського MGQT дали незважений рівень точності .694 (.043) з невизначеним показником .133 (.038). Середньозважене значення правильних рішень для армійського MGQT з використанням 7-позиційної моделі TDA становило .999 (.050) для критеріальних оманливих випадків і .039 (.085) для критеріальних правдивих випадків. Середньозважені невизначені показники становили .043 (.034) для критеріальних оманливих випадків і .224 (.065) для критеріальних правдивих випадків.

Наукові дослідження	Krapohl & Norris(2000)	Blackwell (1999)
Sample N	32	100
N Deceptive	16	80
N Truthful	16	20
Scorers	3	3
D Scores	16	240
T Scores	16	60
Total Scores	32	300
Mean D	-	-
StDev D	-	-
Mean T	-	-
StDev T	-	-
Reliability Kappa	-	-
Reliability Agreement	-	-
Reliability Correlation	.750	.907
Unweighted Accuracy	.833	.660
Unweighted Inconclusives	.219	.125
Sensitivity	.813	.967
Specificity	.500	.250
FN Errors	.001	.001
FP Errors	.250	.533
D-INC	.188	.033
T-INC	.250	.217
PPV	.765	.644
NPV	.999	.999
D Correct	.999	.999
T Correct	.667	.319

**Directed-Lie Screening Test / 3-позиційна система оцінювання**

Два корисних дослідження DLST/TES з 3-позиційною системою оцінювання дали сукупний незважений середній рівень точності .869 (.037) з невизначеним показником .228 (.043). Середньозважене значення правильних рішень для 3-позиційної моделі DLST становило .827 (.060) для критеріальних оманливих випадків і .912 (.043) для критеріальних правдивих випадків. Середньозважені невизначені показники становили .427 (.063) для критеріальних оманливих випадків і .210 (.060) для критеріальних правдивих випадків.

Наукові дослідження	Nelson (In press)	Nelson, Handler, Blalock & Hernández (In press)
Sample N	100	49
N Deceptive	50	25
N Truthful	50	24
Scorers	1	2
D Scores	50	50
T Scores	50	48
Total Scores	100	98
Mean D	-1.585	-1.458
StDev D	2.382	2.784
Mean T	1.719	2.470
StDev T	2.253	1.853
Reliability Kappa	-	-
Reliability Agreement	-	0.762
Reliability Correlation	-	-
Unweighted Accuracy	.893	.817
Unweighted Inconclusives	.208	.248
Sensitivity	.829	.415
Specificity	.595	.848
FN Errors	.033	.228
FP Errors	.127	.009
D-INC	.138	.355
T-INC	.277	.141
PPV	.867	.979
NPV	.947	.788
D Correct	.962	.645
T Correct	.824	.989

**Federal You-Phase / 3-позиційна система оцінювання**

Два придатних для використання дослідження методики Federal You-Phase з 3-позиційною системою оцінювання дали сукупний незважений середній рівень точності .881 (.041) з невизначеним показником .282 (.046). Середньозважене значення правильних рішень для трипозиційної моделі з Federal You-Phase становило .977 (.023) для критеріальних оманливих випадків і .786 (.078) для критеріальних правдивих випадків. Середньозважені невизначені показники становили .181 (.055) для критеріальних оманливих випадків і .348 (.072) для критеріальних правдивих випадків.

Наукові дослідження	Nelson (In press)	Nelson, Handler, Blalock & Cushman (In press)
Sample N	100	22
N Deceptive	50	11
N Truthful	50	11
Scorers	1	10
D Scores	50	110
T Scores	50	110
Total Scores	100	220
Mean D	-4.720	-5.394
StDev D	3.345	4.219
Mean T	3.901	3.982
StDev T	3.804	4.432
Reliability Kappa	-	-
Reliability Agreement	-	0.877
Reliability Correlation	-	-
Unweighted Accuracy	.889	.878
Unweighted Inconclusives	.387	.235
Sensitivity	.740	.826
Specificity	.380	.530
FN Errors	.001	.027
FP Errors	.107	.143
D-INC	.260	.145
T-INC	.513	.325
PPV	.874	.852
NPV	.997	.952
D Correct	.999	.968
T Correct	.780	.788

**Federal ZCT / 3-позиційна система оцінювання (+/-6)**

Два придатних для використання дослідження Federal ZCT з 3-позиційною системою оцінювання з використанням традиційних порогових значень (+/-6 і -3) дали сукупний незважений середній рівень точності .883 (.048) з невизначеним показником .318 (.043). Середньозважене значення правильних рішень для трипозиційної моделі Federal ZCT із традиційними пороговими значеннями становило .990 (.016) для критеріальних оманливих випадків і .675 (.095) для критеріальних правдивих випадків. Середньозважені невизначені показники становили .158 (.050) для критеріальних оманливих випадків і .477 (.069) для критеріальних правдивих випадків.

Наукові дослідження	Capps & Ansley (1992)	Blackwell (1998)
Sample N	100	100
N Deceptive	52	65
N Truthful	48	35
Scorers	1	3
D Scores	52	195
T Scores	48	105
Total Scores	100	300
Mean D	-9.640	-5.938
StDev D	5.146	5.503
Mean T	4.780	4.867
StDev T	5.461	5.580
Reliability Kappa	-	.360
Reliability Agreement	-	.660
Reliability Correlation	-	-
Unweighted Average Accuracy	.977	.779
Unweighted Inconclusives	.386	.293
Sensitivity	.769	.851
Specificity	.438	.314
FN Errors	.001	.010
FP Errors	.021	.238
D-INC	.231	.138
T-INC	.542	.448
PPV	.974	.781
NPV	.999	.968
D Correct	.999	.988
T Correct	.955	.569

**Federal ZCT / 3-позиційна система оцінювання (+/-4)**

Два придатних для використання дослідження федерального ZCT із трипозиційним TDA з використанням покращених показників порогових значень (+/-4 та -3) дали сукупний незважений середній рівень точності .939 (.028) з невизначеним показником .269 (.044). Середньозважене значення правильних рішень для федеральної трипозиційної моделі Federal ZCT з покращеними показниками порогових значень становило .932 (.042) для критеріальних оманливих випадків і .946 (.035) для критеріальних правдивих випадків. Середньозважені невизначені показники становили .314 (.066) для критеріальних оманливих випадків і .225 (.059) для критеріальних правдивих випадків.

Наукові дослідження	Krapohl (1998)	Harwell (2000)
Sample N	100	88
N Deceptive	50	60
N Truthful	50	28
Scorers	5	3
D Scores	250	180
T Scores	250	84
Total Scores	500	264
Mean D	-7.700	-6.194
StDev D	5.876	5.576
Mean T	6.300	5.113
StDev T	5.317	5.521
Reliability Kappa	-	-
Reliability Agreement	-	.990
Reliability Correlation	.900	-
Unweighted Accuracy	.929	.937
Unweighted Inconclusives	.264	.296
Sensitivity	.604	.689
Specificity	.768	.631
FN Errors	.068	.017
FP Errors	.032	.071
D-INC	.328	.294
T-INC	.200	.298
PPV	.950	.906
NPV	.919	.974
D Correct	.899	.976
T Correct	.960	.898

**Positive Control Technique / 7-позиційна система оцінювання\***

Два дослідження Методики позитивного контролю (Positive Control Technique) показали сукупний незважений середній рівень точності .820 (.043) з невизначеним показником .292 (.045). Середньозважене значення правильних рішень для методики позитивного контролю становило .679 (.080) для критеріальних оманливих випадків і .962 (.032) для критеріальних правдивих випадків. Середньозважені невизначені показники становили .333 (.066) для критеріальних оманливих випадків і .250 (.063) для критеріальних правдивих випадків.

Наукові дослідження	Driscoll, Honts & Jones (1987)	Forman & McCauley (1986)
Sample N	40	38
N Deceptive	20	22
N Truthful	20	16
Scorers	1	1
D Scores	20	22
T Scores	20	16
Total Scores	40	38
Mean D	-2.000	-
StDev D	3.800	-
Mean T	6.600	-
StDev T	5.700	-
Reliability Kappa	-	-
Reliability Agreement	-	.800
Reliability Correlation	.840	.800
Unweighted Accuracy	.889	.777
Unweighted Inconclusives	.450	.131
Sensitivity	.350	.545
Specificity	.650	.750
FN Errors	.100	.318
FP Errors	.001	.063
D-INC	.550	.136
T-INC	.350	.125
PPV	.999	.897
NPV	.867	.702
D Correct	.778	.632
T Correct	.999	.923

\* Driscoll, Honts & Jones (1987) використовували 7-позиційну числову систему оцінювання. Forman & McCauley використовували нечисловий підхід оцінювання тестових даних (TDA), і він включений тут лише з метою порівняння.

### Методика Relevant-Irrelevant

Існує одне корисне опубліковане дослідження методики Relevant-Irrelevant (Krapohl, Senter & Stern, 2005), результатом якого є незважений рівень точності .732 (.044). Крепол (Krapohl, 2006) раніше повідомляв про точність методики RI на рівні .83 з нульовими невизначеними результатами, включаючи результати Корреа і Адамса (Correa & Adams 1981), які повідомили про рівень точності 100% у лабораторному дослідженні, в якому використовували методологію, яка не відображає практику.

Наукові дослідження	Krapohl, Senter & Stern (2005)
Sample N	100
N Deceptive	59
N Truthful	41
Scorers	1
D Scores	59
T Scores	41
Total Scores	100
Mean D	-
StDev D	-
Mean T	-
StDev T	-
Reliability Kappa	-
Reliability Agreement	.700
Reliability Correlation	-
Unweighted Accuracy	.732
Unweighted Inconclusives	.001
Sensitivity	.831
Specificity	.634
FN Errors	.169
FP Errors	.366
D-INC	.001
T-INC	.001
PPV	.694
NPV	.789
D Correct	.831
T Correct	.634



### Zone Comparison Techniques / Rank Order Scoring System

Два придатних для використання дослідження ZCT з системою оцінювання тестових даних Rank Order дали сукупний незважений середній рівень точності .886 (.036) з невизначеним показником .213 (.042). Середньозважене значення правильних рішень для методики порівняння зон із системою оцінювання тестових даних Rank Order становило .885 (.050) для критеріальних випадків обману і .887 (.053) для критеріальних випадків правди. Середньозважені невизначені показники становили .020 (.061) для критеріальних оманливих випадків і .225 (.057) для критеріальних правдивих випадків.

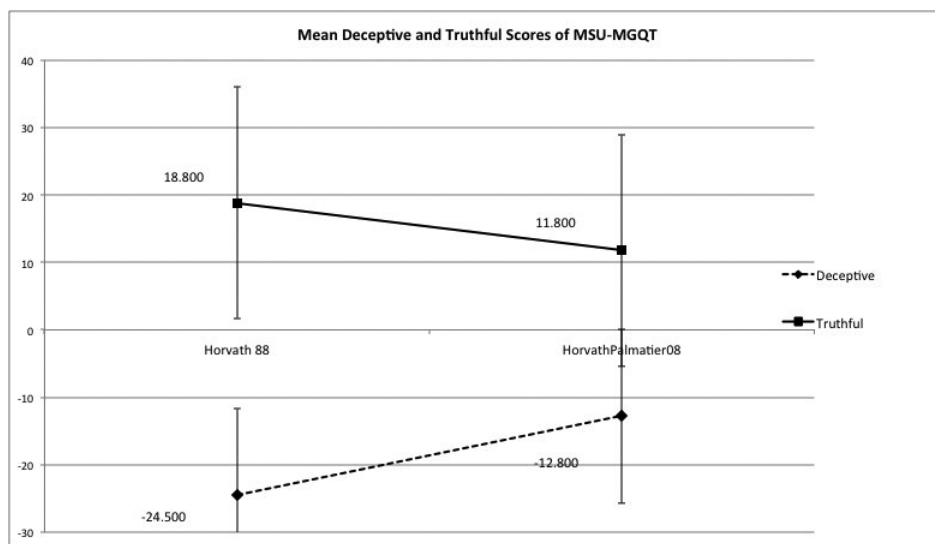
Наукові дослідження	Krapohl, Dutton & Ryan (2001)	Honts & Driscoll (1987)
Sample N	100	60
N Deceptive	50	30
N Truthful	50	30
Scorers	3	2
D Scores	50	30
T Scores	50	30
Total Scores	100	60
Mean D	-24.650	-20.900
StDev D	18.970	8.660
Mean T	7.400	13.400
StDev T	17.060	8.660
Reliability Kappa	-	-
Reliability Agreement	-	-
Reliability Correlation	-	.930
Unweighted Accuracy	.879	.906
Unweighted Inconclusives	.150	.317
Sensitivity	.720	.600
Specificity	.720	.633
FN Errors	.120	.033
FP Errors	.080	.100
D-INC	.100	.367
T-INC	.200	.267
PPV	.900	.857
NPV	.857	.950
D Correct	.857	.947
T Correct	.900	.864

## Додаток до мета-аналітичного огляду 2011 року – MSU-MGQT Реймонд Нельсон (Raymond Nelson)

Цей варіант модифікованої методики загальних запитань був вперше описаний Хорвацем (Horvath, 1988) і повторений Хорвацем і Палматьєром (Horvath & Palmatier, 2008). Обидва дослідження були розроблені для вивчення ефектів, пов'язаних із типами запитань порівняння. Обидва дослідження були проведені в Мічиганському Університеті, і тут використовувалась назва «Модифікована методика загального запитання Університету штату Мічиган» (Michigan State University Modified General Question Technique, MSU-MGQT). Функціонування та виконання MSU-MGQT певною мірою не відрізняються від інших форматів MGQT, і подробиці можна знайти в цитованих раніше дослідженнях.

Хорвац (Horvath, 1988) описав лабораторне дослідження, яке включало послідовність тестових запитань, подібну до тієї, яку описав Рід (Reid, 1947), але включало п'ять релевантних запитань, яке використовувалося для опису статусу винного учасника. Було проведено сліпе оцінювання двома оцінювачами, які використовували 7-позиційну модель оцінювання, описану як подібну до тієї, яку використовували Барленд і Раскін (Barland & Raskin, 1975), із порогом прийняття рішень за загальною сумою балів, встановленим на рівні  $\pm 5$ . Незважаючи на точність рішення сліпих числових балів становила .871 з невизначеним показником .025.

Рисунок 1. Середні значення та стандартні відхилення для вибірки правдивих і оманливих випадків у двох наукових дослідженнях щодо MSU-MGQT.



Хорвац і Палматьєр (Horvath & Palmatier, 2008) повідомили про результати іншого дослідження типів запитань порівняння, які також включали формат MSU-MGQT, але також включали п'ять релевантних запитань, яке використовувалося для опису статусу винного учасника. Оцінювання були виконано з використанням 7-позиційної моделі оцінювання, описаної як подібної до тієї, яку використовували Барленд і Раскін (Barland & Raskin, 1975), із порогом прийняття рішень за загальною сумою балів, встановленим на рівні  $\pm 6$ . Незважаючи на точність рішення сліпих числових балів становила .882 з невизначеним показником .167.

На рисунку 1 показано графік середнього значення та стандартного відхилення балів розподілів вибірки включених досліджень MSU-MGQT. Двофакторний ANOVA показав, що взаємодія розподілу вибірки та статусу критерію була значною [ $F(1,92) = 8,7$ , ( $p = .004$ )]. Джерело взаємодії принаймні частково пояснюється тим фактом, що оцінки як для оманливих, так і для правдивих учасників були значно вищими в дослідженні Хорваца (Horvath, 1988), і обидва були ближче до нуля в дослідженні Хорваца і Палмат'єра (Horvath & Palmatier, 2008). Однофакторний ANOVA не показав суттєвих відмінностей в оцінках двох наукових досліджень ні для оманливих випадків [ $F(1,46) = 0,231$ , ( $p = .633$ )], ні для правдивих випадків [ $F(1,46) = 0.147$ , ( $p = .703$ )].

У таблиці 1 наведено підсумок двох об'єднаних наукових досліджень. Таблиця 2 показує профіль і статистичні довірчі інтервали для показників критеріальної точності. Таблиця 3 показує узагальнення окремих досліджень.

<b>Таблиця 1. Узагальнення наукових досліджень щодо MSU-MGQT</b>	
Number of studies	2
Total N	50
N Deceptive	25
N Truthful	25
Number of Examiners/Scorers	3
Total Scores	100
D Scores	50
T Scores	50
Mean D	-18.650
StDev D	-
Mean T	15.300
StDev T	-
Reliability Kappa	-
Reliability Agreement	0.95
Reliability Correlation	0.92

**Таблиця 2. Критеріальна точність  
і довірчі інтервали наукових досліджень щодо MSU-MGQT**

Unweighted Average Accuracy	.877 (.049) {.781 to .972}
Unweighted Inconclusives	.110 (.043) {.025 to .195}
Sensitivity	.820 (.068) {.686 to .954}
Specificity	.740 (.072) {.599 to .881}
FN Errors	.120 (.064) {.001 to .246}
FP Errors	.100 (.060) {.001 to .218}
D-INC	.060 (.047) {.001 to .153}
T-INC	.160 (.073) {.017 to .303}
PPV	.891 (.065) {.765 to .999}
NPV	.860 (.075) {.713 to .999}
D Correct	.872 (.068) {.739 to .999}
T Correct	.881 (.072) {.740 to .999}

**Таблиця 3. Узагальнення окремих наукових досліджень,  
які використовували MSU-MGQT**

Наукові дослідження	Horvath (1988)	Horvath & Palmatier (2008)
Sample N	20	30
N Deceptive	10	15
N Truthful	10	15
Scorers	2	1
D Scores	20	30
T Scores	20	30
Total Scores	40	60
Mean D	-24.500	-12.800
StDev D	-	17.200
Mean T	18.800	11.800
StDev T	-	12.900

Наукові дослідження	Horvath (1988)	Horvath & Palmatier (2008)
Reliability Kappa	-	
Reliability Agreement	0.95	
Reliability Correlation	0.920	
Unweighted Average Accuracy	0.871	0.882
Unweighted Inconclusives	0.025	0.167
Sensitivity	0.900	0.767
Specificity	0.800	0.700
FN Errors	0.100	0.133
FP Errors	0.150	0.067
D-INC	0.000	0.100
T-INC	0.050	0.233
PPV	0.857	0.920
NPV	0.889	0.840
D Correct	0.900	0.852
T Correct	0.842	0.913

Сукупний рівень точності рішення в дослідженнях MSU-MGQT, зважений за розміром вибірки та кількістю оцінювачів, склав .877 із сукупним коефіцієнтом невизначених показників .11. Надійність для досліджень SU-MGQT, виражена як коефіцієнт узгодженості для категоріальних рішень, була описана Хорватцем (Horvath, 1988) на рівні .95, з коефіцієнтом кореляції .92 для числових оцінок двох оцінювачів.

### **Список використаних джерел**

- Barland, G. H. & Raskin, D.C. (1975). Psychopathy and detection of deception in criminal suspects. *Psychophysiology*, 12, 224.
- Horvath, F. S. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. *Journal of Police Science and Administration*, 16, 198–209.
- Horvath, F. & Palmatier, J. (2008). Effect of two types of control questions and two question formats on the outcomes of polygraph examinations. *Journal of Forensic Sciences*, 53(4), 1–11.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37, 542-547. Reprinted in *Polygraph* 11, 17–21.