



الخصائص السيكومترية لأساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري بالجامعة الأسمرية الإسلامية فتحي عمر ضو بلعيد

¹قسم التربية الخاصة، كلية العلوم الإنسانية - بنات، الجامعة الأسمرية الإسلامية، زليتن، ليبيا.

dawfathi80@asmarya.edu.ly

تاريخ النشر: 2026-06-16

تاريخ قبول البحث: 2025-01-15

تاريخ استلام البحث: 2025-01-01

المخلص

هدف البحث إلى الكشف عن الخصائص السيكومترية لأساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري بالجامعة الأسمرية الإسلامية، ووظف المنهج الوصفي التحليلي؛ لتحليل نتائج طلاب المرحلة السريرية التي تضم السنتين: الرابعة والخامسة، ويدرس الطلاب خلالها ثماني مقررات، وقد تكونت عينة البحث من جميع طلاب المرحلة البالغ عددهم 184 طالباً وطالبة خلال خمسة أعوام جامعية من 2017 إلى 2023م، وأظهرت أبرز النتائج أن أساليب التقييم الثلاثة (الشفوي، العملي السريري، التحريري) تسفر عن نتائج ذات توزيع اعتدالي يفيد باتساق أساليب التقييم مع الواقع الفعلي الطبيعي للطلاب. كما أن أداء طلاب كل سنة دراسية يتسم بمعامل ثبات مقبول في الإطار العام لدرجات المقررات؛ فقد بلغ معامل ألفا كرونباخ 77. بالسنة الرابعة، و 87. بالسنة الخامسة، ونظراً لتباين الأوزان النسبية لأساليب التقييم؛ حُسبت معاملات الثبات في كل مقرر على حدة باستخدام معادلة راجو؛ فتبين أن معاملات الثبات منخفضة؛ حيث كانت من 44. إلى 64. وفُسر ذلك بتباين الأهداف التي يقيسها كل أسلوب تقييم. وللكشف عن مؤشرات صدق تفسير نتائج أساليب التقييم تسنى للباحث التحقق من ثلاثة مؤشرات هي: مؤشر التمييز، ومؤشر الصعوبة، ومؤشر عدالة الاختبار؛ حيث أسفرت المؤشرات الثلاثة عن صدق تفسير نتائج معظم أساليب التقييم، فيما عدا بعض الحالات الفردية. ومن ذلك خلص الباحث إلى أن أساليب التقييم ذات خصائص سيكومترية جيدة في حدود البحث والمؤشرات التي تحقق منها، باستثناء بعض الحالات التي تستوجب تفسيراً علمياً من قبل القائمين عليها، ثم اتخاذ الإجراءات اللازمة لتحسين خصائصها السيكومترية، وأوصى الباحث بالتحقق من مؤشر صدق المحتوى لأساليب التقييم قبل توظيفها في اتخاذ القرار بشأن مدى تحقيق الطلاب لأهداف المقررات.

الكلمات المفتاحية: الثبات والصدق، الخصائص السيكومترية، الطب البشري، المرحلة السريرية، أساليب التقييم.

1. المقدمة:

يمتاز التعليم الطبي بممارسة الطلاب للمعرفة والمهارات في بيئة واقعية، من خلال مرحلة سريرية لها طبيعتها الخاصة في إعداد الطلاب لأداء المهام الطبية تحت إشراف أساتذتهم، ونظراً لاتساع مخرجات التعلم المستهدفة من برامج التعليم الطبي؛ فإن عملية تقييم الطلاب تحظى باهتمام بالغ من قبل

القائمين على البرامج الطبية، وقد خلص مؤتمر Ottawa للتعليم الطبي المنعقد عام 2010م إلى بيان توافقي حدد سبعة معايير أساسية للتقييم الجيد، تنص على الخصائص السيكمترية لأساليب التقييم، متمثلة في مؤشرات ثباتها وصدقها، كما تضمن المعيار الثالث من معايير ضمان جودة التعليم الطبي الصادرة عن الاتحاد العالمي للتعليم الطبي (World Federation for Medical Education [WFME], 2012) مؤشراً نصّه: "يجب على كلية الطب تحديد المبادئ والأساليب والممارسات المستخدمة لتقييم طلابها ... وينبغي أن توثق التحقق من ثبات أساليب التقييم وصدقها" (p.25).

وتُمثل مؤشرات الثبات والصدق محور جودة أساليب القياس والتقييم في معايير الاختبارات التربوية والنفسية الصادرة عن الجمعية الأمريكية للبحوث التربوية [AERA] American Educational Research Association (2014) "التي تعتمد أكثر من 50 جمعية ومنظمة مهنية طبية مثل: ABIM, NBME, Prometric" (Song et al., 2025, P.264). كما أنها من أبرز مؤشرات جودة أدوات القياس بمعايير منظمة COSMIN التي تُعد مركز الخبرة الدولي في أدوات تقييم نتائج البحوث الطبية والممارسات السريرية.

ويُشير الثبات Reliability إلى مقدار الخطأ في الدرجات المُلاحظة، أو هو "نسبة التباين الكلي في القياسات الناتج عن اختلافات حقيقية، ويهدف إلى تقييم ما إذا كانت أداة القياس قادرة على التمييز بين الأفراد ذوي الأداء المختلف أم لا" (Mokkink & Eekhout, 2026, P.2). ويُعد الثبات من الخصائص السيكمترية التي تُحدّد مقداراً كمياً لمصادر التباين في أداء الأفراد، سواء المرتبطة بوقت إجراء التقييم، أو نوع التقييم، أو المرتبطة بالقائمين على التقييم وتصحيح الامتحانات، أو المرتبطة باتساق أداء الفرد في أساليب التقييم، كما يُعد الثبات إطاراً أساسياً لمؤشرات الصدق؛ فإذا لم تتسم النتائج بالاتساق والقبالية للتكرار فإنها تفقد مصداقيتها؛ لذلك يُلاحظ اهتمام معظم الدراسات التي استهدفت تقييم الاختبارات في المجال الطبي بتسليط الضوء على خاصية الثبات.

فقد استهدفت دراسة Al Alak et al. (2017) تقييم نتائج الامتحانات النهائية (التحريرية والشفوية والسريرية) لطلاب السنة النهائية بكلية الطب البشري بجامعة كربلاء العراقية، وتكونت عينة الدراسة من 122 طالباً وطالبة ممن تخرجوا في الكلية خلال عامي (2015، 2016)، وأظهرت نتائج الدراسة أن ثبات مجموع درجات الامتحانات باستخدام معادلة ألفا كرونباخ امتد ما بين 0.86 إلى 0.88. كما أسفرت النتائج عن وجود علاقة ارتباطية موجبة بين المعدل التراكمي لسنوات الدراسة بالكلية والأداء في الامتحان السريري، ولم تظهر أي فروق دالة إحصائية بين الجنسين في الأداء على الامتحانات.

كما أجرى Lister et al. (2026) دراسة استهدفت التحقق من ثبات وصدق أساليب تقييم أداء طلاب مرحلة التدريب السريري في الطب الباطني، حيث تكونت عينة الدراسة من 398 طالباً من طلاب السنة الثالثة بكلية الطب البشري بجامعة مينيسوتا ما بين عامي 2023-2025م، وتم التحقق من ثبات أساليب التقييم وفق مؤشر نظرية إمكانية التعميم، كما أستخدم أسلوب تحليل الانحدار لإيجاد مؤشر صدق البناء، وأظهرت النتائج أن أساليب التقييم المستخدمة تتسم بمؤشرات ثبات وصدق مقبولة.

ويُعرف الصدق validity بأنه: "مدى دعم الأدلة والنظريات لتفسير نتائج الاختبارات في استخدامها المقترح، وهو الاعتبار الأساسي في تطوير الاختبارات وتقييمها ... ولمؤشرات الصدق خمسة مصادر هي: محتوى أداة القياس، وعمليات الاستجابة، والبنية الداخلية، والعلاقة مع متغيرات أخرى، والآثار الناتجة عن استخدام الأداة" (AERA et al., 2014, p.14). ويتم التحقق من صدق أدوات القياس بجمع المؤشرات التي تتكامل فيما بينها لدعم صدق الأداة في قياس ما أعدت لقياسه، ويعتمد أسلوب اختيار المؤشرات الداعمة للصدق على طبيعة السمة المقاسة، والإجراء المتبع في قياسها؛ ففي حالات التحقق من مؤشرات صدق الاختبارات النهائية من خلال نتائج الطلاب؛ غالباً ما يُعتمد على مصدري (البنية الداخلية، والعلاقة مع متغيرات أخرى).

ففي دراسة Latif & Alsunni (2023) التي استهدفت المقارنة بين أساليب التقييم المستخدمة في سنوات الدراسة الجامعية بكلية الطب البشري بجامعة الإمام عبد الرحمن بن فيصل السعودية، تمت مقارنة مؤشرات الصعوبة والتميز لنتائج أعمال السنة والامتحانات النهائية التحريرية والعملية لطلاب السنة الثانية والثالثة بالعام الجامعي 2020-2021م لعينة عينة بلغت 485 طالباً وطالبة، وأظهرت نتائج مؤشر الصعوبة أن تقييم أعمال السنة كان سهلاً، والامتحانات التحريرية كان متوسطاً، في حين كان التقييم العملي OSCE صعباً، أما مؤشرات التمييز بالسنتين الثانية والثالثة فقد أظهرت جميع النتائج أنها ضعيفة فيما عدا الامتحانات التحريرية كانت مؤشرات تمييزها متوسطة.

كما هدفت دراسة Simaremare (2021) إلى تحليل نتائج اختبار Progress test لطلاب مرحلتي الطب الأساسي والسريري بكلية الطب بجامعة HKBP Nommensen الأندونيسية، وتكونت عينة الدراسة من جميع الطلاب الدارسين بالمرحلتين خلال الفصل الثاني من العام الجامعي 2018/2019 والبالغ عددهم 215 طالباً، ووظفت الدراسة المنهج الوصفي لتحليل نتائج بنود الاختبار من حيث مؤشري الصعوبة والتمييز؛ فأظهرت النتائج أن أداء الطلاب على الاختبار يزداد مع تقدم الطلاب في سنوات

الدراسة بالكلية، وأن صعوبة بنود الاختبار في المستوى المتوسط، وأغلب مؤشرات التمييز منخفضة لفئتي الطب الأساسي والسريري.

ورغم تعقيد نظام تقييم الطلاب بالتعليم الطبي؛ إلا أن ذلك لا يُسوغ إغفال ضمان جودته؛ ففي ذات الإطار تُقر منظمة WFME (2023) بأن: "ضمان جودة نظام التقييم في التعليم الطبي يُعد تحدياً خاصاً، لكن لا بد من مواجهته" (p.23). "فلتقييم طالب الطب تقييماً شاملاً يجب استخدام أساليب متنوعة لتقييم الكفاءات الأساسية المختلفة: المعرفة، والمهارات السريرية، والاتجاهات" (Canhão, 2025, P.44). لذلك تنتوع أساليب تقييم طلاب المرحلة السريرية ضمن إطار تكاملي لا يقتصر على إصدار الأحكام التقييمية، بل يمتد أثره إلى تحفيز عملية التعلم، وصقل مهارات الطلاب، وتعزيز قدراتهم على اتخاذ القرارات السريرية، "فإذا هدف التدخل التعليمي إلى زيادة المعرفة، فإن التقييم التحريري سيحفز الطلاب على الدراسة النظرية، أما إن كان الهدف هو تنمية المهارات السريرية؛ فإن التقييم السريري يحفز الطلاب على التفاعل مع المرضى" (Norcini, 2007, P.240).

ويُصنّف تقييم طلاب المرحلة السريرية -حسب أهدافه- إلى نوعين أساسيين أولهما: التقييم التكويني (أو التقييم من أجل التعلم) الذي يستهدف تحديد نقاط القوة ومواطن الضعف في العملية التعليمية؛ حيث يقدم التغذية الراجعة للأساتذة حول استراتيجية التدريس المتبعة لتحسينها، وتقديم التغذية الراجعة للطلاب حول استراتيجية التعلم التي يتبعونها لتعزيزها، وتحسين التعلم وتحفيزه واستمراره، والآخر هو التقييم النهائي (تقييم التعلم) الذي يستهدف تقييم مدى تحقيق الطلاب لمخرجات التعلم المستهدفة؛ وبناء على نتائجه يتم اتخاذ القرارات، وإصدار الأحكام النهائية بشأن مستوى تحصيل الطلاب من المعارف والمهارات المستهدفة.

وتبرز أهمية التقييم النهائي بالمرحلة السريرية استناداً إلى أهمية القرارات التي تُبنى على نتائجه، "فالغرض من هذا التقييم هو تحديد ما إذا كان الطالب قد تعلم ما تم تدريسه، ويسهم في اتخاذ قرارات تُشير إلى ما إذا كان الطالب مؤهلاً للمضي قدماً في التدريب أو الممارسة" (Dent et al., 2021, P.675). لذلك لا يعتمد التقييم النهائي على أسلوب واحد بل تنتوع أساليب التقييم المستخدمة بمعظم المؤسسات الطبية لتشمل: امتحانات تحريرية وشفوية وتقييم الأداء العملي والسريري المباشر، ولا يُكتفى بأي من هذه الأساليب دون البقية، "إلا أنه غالباً ما تُجرى هذه التقييمات بشكل منفصل ودون تنسيق" (Dent et al., 2021, P.677) مما دعا العديد من الباحثين إلى التصدي للكشف عن طبيعة هذه الأساليب ومدى كفاية خصائصها السيكمترية. فقد أجرى Song et al. (2025) دراسة اتّبعَت منهج ما وراء التحليل لمراجعة

أساليب التقييم في التعليم الطبي الجامعي؛ للكشف عن السمات الرئيسة لأساليب التقييم المستخدمة وخصائصها السيكمترية، حيث شمل التحليل عدد 214 دراسة نُشرت بين عامي 2004-2021، وأظهرت النتائج أنه لا يوجد أسلوب تقييم محدد كافٍ للتمييز بين طلاب الطب، كما أوصت الدراسة بضرورة تدريب القائمين على التقييم بكليات الطب البشري على أساليب التحقق من الخصائص السيكمترية للأساليب التي يستخدمونها لتقييم طلابهم.

وللتحقق من جودة أساليب التقييم المتنوعة التي توظفها مؤسسات التعليم الطبي، وإيفائها بمعايير التقييم الجيد؛ برز اهتمام الباحثين بإجراء دراسات مستفيضة للتحقق من جودة أساليب التقييم المستخدمة من جوانب متعددة؛ فقد أجرى Memon & Shaikh (2020) دراسة استهدفت مقارنة أداء طلاب كلية الطب البشري بجامعة LUMHS الباكستانية في الامتحان النهائي (التحريري والسريري) لمقرر طب الأطفال، وتكونت العينة من 160 طالباً خلال 2016/2017 ممن تقدموا للامتحان النهائي التحريري والسريري، وخلصت الدراسة إلى أن أداء الطلاب في الامتحان السريري كان أفضل من أدائهم في الامتحان التحريري، وأظهرت أن أداء الإناث أفضل من أداء الذكور في الامتحانين.

كما هدفت دراسة Buzaid et al. (2023) إلى مقارنة أداء طلاب السنة النهائية بكلية الطب البشري بجامعة بنغازي الليبية في الامتحانات التحريرية والسريرية والشفوية لمقرر الطب الباطني، فتكونت عينة الدراسة من جميع طلاب السنة النهائية بالعام الجامعي 2019، 2020م وعددهم 679 طالباً وطالبة، وخلصت النتائج إلى أن الأداء العام للطلاب في الامتحانات السريرية والشفوية بلغ 73.9 % و 67.5 % على التوالي وهو أعلى من أدائهم في الامتحانات التحريرية بفارق دال إحصائياً عند مستوى 0.001. ولم تظهر النتائج فروقاً دالة إحصائياً بين أداء الطلاب والطالبات في أي من أنواع الامتحانات الثلاثة، كما أسفرت نتائج الدراسة عن وجود علاقة ارتباطية موجبة امتدت من 63. إلى 76. وهي ارتباطات دالة إحصائياً عند مستوى 0.01 بين الامتحانات الثلاثة.

مما سبق استعراضه من دراسات سابقة يتضح تشابه أساليب التقييم المستخدمة بالمرحلة السريرية، ومعظمها في إطار (التحريري، والشفوي، والعمل السريري) ولم تُشر إلى التحقق من التوزيع الاعتدالي لدرجات الطلاب في أساليب التقييم المختلفة، ورغم أهمية التحقق من ثبات أساليب القياس، قبل التحقق من مؤشرات الصدق؛ حيث يؤكد Kaplan, et al. (2017) "إن محاولة تحديد صدق اختبار ما ستكون عديمة الجدوى إذا لم يكن الاختبار ثابتاً؛ فلا ينبغي أن يكون ارتباط الاختبار بأي متغير آخر

أعلى من ارتباطه بنفسه" (p.155). إلا أن أغلب الدراسات السابقة اهتمت بدراسة مؤشرات الصدق دون الإشارة إلى ثبات أساليب التقييم؛ فيما عدا دراستي (Al Alak et al., 2017 & Lister et al., 2026) فقد حُسب معامل ثبات أساليب التقييم في الدراستين، واتفقت نتائجهما في أن أساليب التقييم اتسمت بمعاملات ثبات مقبولة، وتجدر الإشارة إلى أن حساب معاملات الثبات بالدراستين تم لمجموع نتائج الطلاب في أساليب التقييم، ولم يُستخرج لكل أسلوب تقييم في كل مقرر على حدة، وتركز التحقق من مؤشرات الصدق وأدلته في حساب معاملات الصعوبة، ومعاملات التمييز، والعلاقة الارتباطية بين أساليب التقييم، وبينها وبين المعدل التراكمي للطلاب، بالإضافة إلى الفروق بين الجنسين، والفروق بين أداء الطلاب في أساليب التقييم الثلاثة، ولم تتفق الدراسات في نتائج مؤشرات الصدق؛ مما يعزز الحاجة إلى إجراء دراسات مستفيضة لبحث جودة أساليب التقييم المستخدمة في التعليم الطبي.

1.1 مشكلة البحث:

جاء إحساس الباحث بالمشكلة خلال تكليفه من قبل المركز الوطني لضمان جودة واعتماد المؤسسات التعليمية والتدريبية ضمن فرق تدقيق على بعض كليات الطب البشري؛ حيث لاحظ أن دليل معايير اعتماد الكليات الطبية الصادر عن المركز عام (2021) قد أفرد معايير متكاملة لعملية تقييم الطلاب، بل ينص المؤشر العاشر من المعيار الثالث على: "يجب على البرنامج تقييم وتوثيق مصداقية وصحة طرق التقييم" (ص.18). إلا أن محاولات الكليات التي تسنى للباحث الاطلاع على آلياتها المعتمدة للإيفاء بمتطلبات هذا المؤشر يعثرها قصور ملحوظ في تفسير هذا المؤشر فضلاً عن سلامة الإجراءات المتبعة لتحقيقه؛ ونظراً لأهمية القرارات التي تُبنى على نتائج أساليب التقييم بالكليات الطبية في تخريج أطباء مؤهلين لممارسة مهنة دقيقة للغاية، تنبّه الباحث إلى ضرورة التحقق من جودة أساليب التقييم المستخدمة من خلال الإجابة عن تساؤل نصّه: ما الخصائص السيكومترية لأساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري بالجامعة الأسمرية الإسلامية؟ وتتفرع عنه الأسئلة الآتية:

- هل تتوزع درجات الطلاب في امتحانات المرحلة السريرية بكلية الطب البشري توزيعاً اعتدالياً؟
- ما معاملات ثبات أساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري؟
- ما مؤشرات صدق أساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري؟

2.1 أهداف البحث

- التحقق من اعتدالية توزيع درجات الطلاب في امتحانات المرحلة السريرية، وتفسيرها.

- حساب معاملات ثبات نتائج أساليب تقييم طلاب المرحلة السريرية وتفسيرها.
- وصف مؤشرات صدق أساليب تقييم طلاب المرحلة السريرية، وتفسيرها.

3.1. أهمية البحث

يستمد هذا البحث أهميته من أهمية ضمان جودة التعليم الطبي، ودوره البارز في تقديم الخدمات الصحية للمجتمع، بالإضافة إلى أهمية القرارات التي تُبنى على نتائج أساليب تقييم أداء طلاب كليات الطب البشري؛ مما قد ينبه القائمين على نظام التعليم الطبي إلى خصائص أساليب التقييم المستخدمة، وكيفية التحقق الدوري من كفايتها، ثم البناء عليها لتحسين ضوابط إعداد الامتحانات بأنواعها، وتطوير بنوك الأسئلة بالكليات، كما قد يفتح آفاقاً بحثية جديدة في مجال التحقق من جودة أساليب التقييم، لدعم المكتبة الليبية وإثرائها مواكبة للتطور العلمي.

4.1. حدود البحث

يقتصر تعميم نتائج البحث على أساليب التقييم المتبعة في تقييم طلاب المرحلة السريرية بكلية الطب البشري بالجامعة الأسمرية الإسلامية الكائنة بمدينة زليتن الليبية، ما بين عامي 2017-2023م.

5.1. التحديد الإجرائي لمصطلحات البحث

الخصائص السيكمترية: هي الأسس الكمية والنوعية اللازمة للتأكد من صلاحية المقاييس النفسية والتربوية وفق نظرية القياس التقليدية. وتعرّف إجرائياً بأنها معاملات ثبات أساليب التقييم، ومؤشرات صدقها، بعد التحقق من التوزيع الاعتدالي لدرجات الطلاب.

أساليب التقييم: هي الطرق المتبعة في تقييم طلاب المرحلة السريرية بنهاية السنة الدراسية. وتعرّف إجرائياً بأنها: الدرجات الخام لطلاب المرحلة السريرية في الامتحانات النهائية بفروعها الثلاثة (الشفوية، العملية السريرية، التحريرية) في مقررات السنتين الرابعة والخامسة بكلية الطب البشري.

طلاب المرحلة السريرية: تحددهم وزارة التعليم العالي والبحث العلمي-ليبيا (2022) بالطلاب المقيدين بالمرحلة الثالثة "السريرية" التي تضم السنتين الرابعة والخامسة بكلية الطب البشري (ص.16).

كلية الطب البشري: هي إحدى كليات الجامعة الأسمرية الإسلامية، أنشئت عام 2013م، وتقع بجوار المركز الطبي بمدينة زليتن الليبية.

2. إجراءات البحث

لتحقيق أهداف البحث تم توظيف المنهج الوصفي التحليلي، لفرز وتحليل نتائج الامتحانات النهائية لطلاب المرحلة السريرية بكلية الطب البشري، حيث شمل مجتمع البحث جميع الطلاب المقيدون بالمرحلة السريرية التي تتمثل في السنتين (الرابعة والخامسة) للدفعات الأربع التي أنهت المرحلة السريرية بالكلية منذ إنشائها حتى تاريخ جمع البيانات، وتمثلت عينة البحث في كامل مجتمعه بعد استبعاد الطلاب غير المنتظمين في أداء الامتحانات النهائية لمقررات المرحلة السريرية سواءً بالانقطاع، أو الغياب أو غيرهما، ويبين الجدول الآتي الأعداد التفصيلية لمجتمع البحث وعينته:

الجدول 1

توزيع أعداد مجتمع البحث وعينته

العام الجامعي	السنة الرابعة		السنة الخامسة		توزيع الذكور والإناث بعينة البحث		
	المجتمع	العينة	المجتمع	العينة	السنة الرابعة	السنة الخامسة	
					إناث	ذكور	
2018/2017	53	53	--	--	34	19	--
2019/2018	35	35	50	53	24	11	31
2020/2019	51	50	35	40	33	17	24
2021/2020	46	46	51	56	34	12	35
2023/2022	--	--	40	47	--	--	28
المجموع	185	184	196	176	125	59	118

يتضح من الجدول 1 أن مجتمع البحث شمل جميع طلاب المرحلة السريرية خلال خمسة أعوام جامعية ما بين 2017-2023م وتجدر الإشارة هنا إلى أن الكلية لم تجر أي امتحانات نهائية لطلاب المرحلة السريرية خلال العام الجامعي 2022/2021م نظراً لخصوصية تنظيم مدة الدراسة بعد ظروف جائحة كوفيد 19، كما يتبين أن عينة السنة الرابعة شملت جميع الطلاب تقريباً، أما عينة السنة الخامسة فقد مثلت نحو 90% من مجتمع البحث؛ ويعود ذلك لاستبعاد بعض الحالات منها على سبيل المثال: عدم استكمال الطالب لامتحانات (الشفوية، والعملية، والتحريرية) لجميع المقررات المعتمدة بالسنة الخامسة، أو أداء الطالب لامتحانات المقرر خلال عامين جامعيين بسبب الرسوب أو الغياب. ونظراً لأن طلاب السنة الخامسة لأي عام جامعي هم نفس طلاب السنة الرابعة في العام الجامعي السابق له تقريباً؛ لذلك تكونت عينة البحث من 184 طالباً وطالبة، بلغ عدد الذكور منهم نسبة 32% تقريباً.

ونظراً لأن طبيعة البحث تستهدف أساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري من خلال النتائج الفعلية للامتحانات التي تم إجراؤها مسبقاً (أي بأثر رجعي)؛ لذلك قابل الباحث السيد عميد

الكلية لطلب الإذن في الحصول على النتائج التفصيلية للفئة المستهدفة، حيث رحب العميد بموضوع البحث وأهميته مُبدئاً استعداداه التام لتقديم التسهيلات الممكنة، ثم أصدر تعليماته إلى القائمين على قسم الدراسة والامتحانات* بتزويد الباحث بالبيانات المطلوبة؛ مما مكن الباحث من الحصول على البيانات من الكشوف الرسمية بالكلية، ثم عمل الباحث على تنويبها وتنظيمها في ملف Excel بما يخدم إجراءات تحليلها إحصائياً باستخدام برنامج SPSS-27 لتحقيق أهداف البحث.

وقد تمثلت بيانات البحث في الدرجات الخام لطلاب المرحلة السريرية في الامتحانات النهائية بفروعها الثلاثة (الشفوية، العملية، التحريرية) في مقررات السنة الرابعة ومقررات السنة الخامسة، ويبين الجدول الآتي أسماء المقررات وحروف رموزها المختصرة بكشوف الدراسة والامتحانات بالكلية:

الجدول 2

بيان بأسماء المقررات الدراسية المعتمدة بالمرحلة السريرية ورموزها

السنة	الرابعة	الخامسة
اسم المقرر	طب الأسرة والمجتمع والنساء والتوليد	طب الجراحة العامة وفروعها
الرمز المختصر للمقرر	COM GYN OPH RDA FOR MED SUR PED	

يتبين من الجدول 2 أن العدد الكلي لمقررات المرحلة السريرية بكلية الطب البشري هو ثمانية مقررات، منها خمسة يدرسها الطلاب بالسنة الرابعة، وثلاثة مقررات أخرى يدرسها الطلاب بالسنة الخامسة، وينبه الباحث إلى أن الرموز الواردة بالجدول هي اختصارات معتمدة للإشارة إلى أسماء المقررات في كشوف نتائج الطلاب بقسم الدراسة والامتحانات، وقد اعتمدها البحث للإشارة إلى المقرر المتعين خلال عرض نتائج البحث ومناقشتها.

3. نتائج البحث ومناقشتها وتفسيرها

للإجابة عن التساؤل الرئيس للبحث الذي نصه: ما الخصائص السيكومترية لأساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري بالجامعة الأسمرية الإسلامية؟ تمت الإجابة عن التساؤلات التي تفرعت عنه كما يلي:

* الشكر والتقدير لكل من: د. أسامة الصاري، د. خالد البكوش، أ. عطية حشاد، أ. يامن الجفائري

1.3. نتائج التساؤل الفرعي الأول الذي نصّه: "هل تتوزع درجات الطلاب في امتحانات المرحلة السريرية بكلية الطب البشري توزيعاً اعتدالياً؟"

للإجابة عن هذا التساؤل تم التحقق من اعتدالية توزيع درجات الطلاب في الامتحانات النهائية بفروعها الثلاثة (الشفوية، العملية، التحريرية) لكل مقرر من مقررات المرحلة السريرية باستخدام طريقة حساب معاملات الالتواء والتفرطح؛ فعلى الرغم من عدم وجود طريقة مثالية للتحقق من اعتدالية التوزيع، إلا أن هذه الطريقة تتفادى حالات عدم التوافق التي قد تحدث بين نتائج المعاينة التقديرية لمنحنى التوزيع ونتائج اختبارات التوزيع الاعتدالي مثل: Kolmogorov-Smirnov & Shapiro-Wilk التي تعتمد أسلوب الدلالة الإحصائية، كما أن طريقة حساب قيم الالتواء والتفرطح ملائمة للعينات الكبيرة والصغيرة (Kim, 2013, 52).

وقبل الشروع في استعراض قيم معاملات الالتواء (Sk) والتفرطح (Ku) تجدر الإشارة إلى أن القيمة المثالية للتوزيع الطبيعي تتحقق عندما تكون قيمة كل منهما تساوي صفر، وهي حالة نادرة في العينات الممثلة لمجتمعها؛ لذلك فإن قيمهما ممتازة ما بين -1 و+1، ومقبولة ما بين -2 و+2 أما إذا خرجت عن حدود هذا المدى في أي من الاتجاهين؛ فهي غير طبيعية (Hair et al., 2022, p.66).

الجدول 3

قيم معاملات الالتواء والتفرطح لتوزيع درجات الطلاب

المقرر	COM n=184		GYN n=184		OPH n=184		RDA n=184		FOR n=184		MED n=174		SUR n=174		PED n=174	
	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk	Ku	Sk
الشفوي	1.69	-0.12	0.40	0.92	-0.93	0.62	-0.72	0.78	-0.02	0.68	-0.58	0.15	-0.08	0.64	-1.04	1.65
العملي	0.78	-0.28	0.07	0.62	-0.33	0.21	-0.61	0.07	-0.62	0.56	0.02	0.33	-0.44	0.78	-0.28	0.46
التحريري	0.38	-0.35	0.06	1.12	-0.50	0.15	-0.33	0.38	-0.51	0.06	0.26	-0.29	0.66	1.46	-0.29	0.33

يتضح من الجدول 3 أن درجات طلاب المرحلة السريرية تتوزع توزيعاً اعتدالياً في جميع المقررات باختلاف أساليب التقييم المستخدمة، حيث امتدت قيم الالتواء والتفرطح ما بين -1.12 و1.65 وهي جميعاً ضمن المدى المقبول للتوزيع الاعتدالي أي ما بين -2 و+2، وهو ما يُفسر بسلامة أساليب التقييم وملاءمتها لطبيعة الطلاب وشموليتها للمهارات المستهدفة بمختلف مستوياتها؛ ففي ذات السياق قال Zuo (2021): "إن التوزيع الطبيعي لدرجات الطلاب يعد مؤشراً على أن نتائج الاختبارات متسقة مع الواقع الفعلي للطلاب، وأن أسئلة الاختبار شاملة ودرجة صعوبتها معتدلة" (P.475).

2.3. نتائج التساؤل الفرعي الثاني الذي نصّه: "ما معاملات ثبات أساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري؟"

للإجابة عن هذا السؤال حُسب ثبات أداء الطلاب في امتحانات كل سنة على حدة باستخدام معادلة ألفا كرونباخ Cronbach's Alpha كما حُسب الخطأ المعياري للقياس لأهميته في ذات السياق؛ "إذا كان الثبات يشير إلى قدرة الأداة على التمييز بين الأفراد، وهو ما يعتمد على كل من التباين بين الأفراد ودقة النتيجة؛ فإن الخطأ المعياري يشير إلى دقة القياس، أي مدى تشابه النتائج عبر القياسات المتكررة لدى الفرد الواحد. (Mokkink & Eekhout, 2026, p.1). ويُحسب بناتج ضرب الانحراف المعياري في الجذر التربيعي لنواتج طرح قيمة معامل الثبات من الواحد الصحيح.

الجدول 4

قيم معامل الثبات بكل سنة دراسة

السنة الدراسية	الرابعة	الخامسة
معامل الثبات	.77	.87
الانحراف المعياري	.11	.09
الخطأ المعياري للقياس	.05	.03

يتضح من الجدول 4 أن قيم معامل ثبات أساليب تقييم الطلاب ضمن المدى المقبول بالسنة الرابعة، وضمن مدى الممتاز بالسنة الخامسة، وذلك حسب التصنيف الذي نشره مكتب التقييم التربوي بجامعة واشنطن، الذي يُصنف قيم معامل ثبات الاختبارات التحصيلية إلى ست مستويات كما يلي: (Office of Educational Assessment, n.d.)

الجدول 5

تصنيف قيم معامل ثبات الاختبارات التحصيلية

قيمة معامل الثبات	التصنيف
أقل من 5.	منخفض ومشكوك فيه، ولا ينبغي أن يسهم في درجة المقرر
من 5. إلى أقل من 6.	منخفض ويجب مراجعة الاختبار
من 6. إلى أقل من 7.	منخفض نوعاً ما، وينبغي دعمه بإجراءات تقييم أخرى
من 7. إلى أقل من 8.	مقبول على مستوى اختبارات الفصل الدراسي
من 8. إلى أقل من 9.	ممتاز على مستوى اختبارات الفصل الدراسي
من 9. فأكثر	ممتاز على مستوى أفضل الاختبارات المعيارية

وقد تبين من الجدول 5 أن معامل الثبات المحسوب في حالتي السنة الرابعة والسنة الخامسة يتحدد تعميمه على مستوى الفصل الدراسي؛ أي لا يُعمم على استخدام ذات الاختبارات بمؤسسات أخرى. ومن جانب آخر أظهرت النتائج بالجدول 4 أن قيم الخطأ المعياري للقياس مقبولة؛ فهي لم تتجاوز نصف الانحراف المعياري، "ويُعد الخطأ المعياري للقياس أهم مؤشر لثبات تفسير الدرجات؛ فهو يُستخدم لتحديد

نطاق حول الدرجة الملاحظة، وكلما اتسع هذا النطاق، انخفض ثبات الدرجة" (Kaplan, 2017, p.124). وتشير هذه النتائج إلى ثبات أداء طلاب المرحلة السريرية في الامتحانات النهائية، وهي نتيجة تتفق مع نتائج دراستي (Al Alak et al., 2017; Lister et al., 2026) وللكشف عن معاملات ثبات أداء الطلاب في أساليب التقييم الثلاثة لكل مقرر على حدة؛ أستخدمت معادلة راجو Raju التي تلائم تقدير ثبات أساليب التقييم ذات الأشكال المختلفة (congeneric tests) من حيث طبيعتها وأوزانها النسبية، وتتمثل معادلة راجو في الصيغة الآتية: (العمرى، 2022، ص.88).

$$\text{معامل الثبات} = \frac{1}{1 - \left(\frac{\text{مجموع مربعات الأوزان النسبية للمكونات}}{\text{مجموع تباينات المكونات}} - 1 \right)}$$

وبحساب معامل الثبات لأداء الطلاب في امتحانات كل مقرر على حدة كانت النتائج كما يلي:

الجدول 5

قيم معامل الثبات لأساليب التقييم (شفوي، عملي، تحريري) في كل مقرر على حدة

المقرر	COM n=86	GYN n=184	OPH n=184	RDA n=184	FOR n=184	MED n=176	SUR n=176	PED n=176
Raju	.57	.44*	.64	.61	.57	.58	.63	.6

ملحوظة: *معامل ثبات منخفض

يُلاحظ من الجدول 5 أن معظم قيم معاملات ثبات أساليب التقييم في مستوى المتوسط؛ فقد امتدت ما بين (.57 - .64). وهي تُصنف على أنها قيم معاملات ثبات متوسطة، فيما عدا ثبات أساليب التقييم في مقرر طب النساء والتوليد (GYN) التي تبين النتائج أنها منخفضة (.44)، وبذلك يمكن استخلاص نتيجة مفادها أن اتساق أداء الطلاب في المقررات المختلفة أعلى من اتساق أدائهم في أساليب التقييم المتنوعة في ذات المقرر، وتُفسر هذه النتيجة بأن تباين أساليب التقييم يؤثر سلباً على ثبات التقييم؛ فقد خلص Spielvogel, et al. (2012) إلى أن "الاختلافات في أساليب القياس تؤدي إلى تقليل الثبات" (p.245). ويُفسر بتباين السمات التي يستهدف قياسها كل أسلوب، وهذا مؤشر داعم لاستخدام أساليب تقييم متنوعة في المرحلة السريرية بكليات الطب البشري.

3.3. نتائج التساؤل الفرعي الثالث الذي نصّه: "ما مؤشرات صدق أساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري؟"

للإجابة عن هذا التساؤل تبرز أهمية التحقق من مؤشر صدق المحتوى لأساليب التقييم المستخدمة؛ إلا أنه تعذر على الباحث التحقق من صدق المحتوى نظراً لعوامل متعددة من أبرزها تعدد

السنوات الدراسية التي شملها البحث، وصعوبة الحصول على جميع الأسئلة المستخدمة في التقييم، وخاصة في التقييم الشفوي والتقييم العملي، وما يتعلق بطبيعتها؛ لذلك سعى الباحث إلى التحقق من مؤشرات الصدق باتباع ثلاث طرق كما يلي:

1.3.3. **مؤشر الصدق التمييزي:** إن تقييم طلاب المرحلة السريرية يستهدف قياس كفايتهم في جوانب متعددة؛ لذلك لا يُكتفى بأسلوب تقييم واحد بل تتعدد أساليب التقييم لتشمل القدرات والمهارات المستهدفة من دراسة كل مقرر، ولتحقيق صدق بناء أساليب التقييم المستخدمة يجب أن يُظهر كل منها تمايزاً ملحوظاً في القدرات والمهارات التي يستهدف قياسها؛ وفي هذه الحال يبرز مؤشر الصدق التمييزي Discriminant Validity "كأحد مؤشرات التحقق من صدق البناء الذي يُصنفه العديد من علماء النفس اليوم بأنه المؤشر الرئيس للصدق" (Kaplan, 2017, p.154).

ويُعرف الصدق التمييزي بأنه "درجة ارتباط مقياسين مصممين لقياس سمات متشابهة لكنها متميزة من حيث المفهوم، وغالباً ما يُعد الارتباط المنخفض إلى المتوسط دليلاً على صحة التمييز" (Roemer, et al., 2021). وقد تم التحقق من معاملات الارتباط بين أساليب التقييم في كل مقرر على حدة؛ فكانت النتائج كما هي موضحة بالجدول الآتي:

الجدول 6

قيم معامل الارتباط بين أساليب التقييم

المقرر	COM	GYN	OPH	RDA	FOR	MED	SUR	PED
	n=88	n=184	n=184	n=184	n=184	n=176	n=176	n=176
التقييم	شفوي عملي	شفوي عملي	شفوي عملي	شفوي عملي	شفوي عملي	شفوي عملي	شفوي عملي	شفوي عملي
شفوي	.40	.27	.18	.19	.45	.33	.55	.29
تحرير	.32	.24	.31	.09*	.26	.43	.47	.39

ملحوظة: * غير دال إحصائياً

يتضح من الجدول 6 أن قيم معامل الارتباط بين أساليب التقييم المستخدمة في معظم المقررات امتدت ما بين (.18) و (.57). وهي قيم موجبة ودالة إحصائياً عند مستوى (.05). مما يُفيد بوجود علاقة ارتباطية موجبة بين أساليب التقييم الثلاثة بكل مقرر، كما أن حجم التأثير Effect Size امتد ما بين متوسط وقوي، ويُفسر ذلك بارتباط المعارف والمهارات المقاسة بكل مقرر، وتتفق هذه النتيجة مع نتائج دراسة Buzaid et al. (2023) وفي ذات الوقت فإن قيم معاملات الارتباط بالبحث الحالي منخفضة بما يكفي لتحقيق مؤشر الصدق التمييزي فهي لم تتجاوز (.57). وتتفق هذه النتيجة مع نتائج دراسة Simaremare (2021) وفي هذا الصدد يرى Panzeri, et al. (2024) أن الارتباط بين أساليب التقييم يفيد

بقياها لخصائص مرتبطة ببعضها، لكن ينبغي أن تكون قيم معاملات الارتباط منخفضة بما يكفي للدلالة على أن أساليب التقييم لا تقيس الشيء نفسه؛ فالارتباط المرتفع جداً قد يُعد تكراراً لقياس ذات المهارات.

ومن ذلك يُستنتج تحقق مؤشر الصدق التمييزي في أساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري، ويُستثنى من هذه النتيجة ثلاث حالات منها اثنتان تتعلق بمقرر طب النساء والتوليد (GYN) حيث يُلاحظ انخفاض شديد في ارتباط الامتحان التحريري بالشفوي والعملي، وتمثلت الحالة الثالثة في انخفاض ارتباط الامتحان التحريري بالشفوي في مقرر الأشعة التشخيصية والتداخلية (RDA) ولتحديد الأسلوب الذي تسبب في انخفاض الارتباط قام الباحث بحساب معامل ارتباط كل أسلوب تقييم مع مجموع الأسلوبين الآخرين فكانت النتائج كما هي موضحة بالجدول الآتي:

الجدول 7

قيم معاملات ارتباط كل أسلوب تقييم مع الدرجة الكلية (مع استبعاد درجة ذات الأسلوب)

المقرر	COM n=88	GYN n=184	OPH n=184	RDA n=184	FOR n=184	MED n=121	SUR n=121	PED n=176
الشفوي	.51	.19	.27	.17	.44	.44	.58	.43
العملي	.40	.25	.36	.45	.57	.54	.61	.41
التحريري	.38	.10*	.50	.36	.44	.59	.52	.50

ملحوظة: * غير دال إحصائياً

يتبين من الجدول 7 أن حالة انخفاض قيم معاملات الارتباط في مقرر طب النساء والتوليد (GYN) تعود إلى الامتحان التحريري؛ وحالة انخفاض معامل الارتباط في أساليب تقييم مقرر الأشعة التشخيصية التداخلية (RDA) تعود إلى الامتحان الشفوي؛ ففي الحالتين يُعد حجم تأثير الارتباط ضعيف، ولصعوبة وضع تفسير محدد لهذه النتيجة؛ فإنها قد تشير إلى احتمال وجود مشكلة ما تتعلق بالخطوات المتبعة في إعداد أسئلة الامتحانين - عبر السنوات التي يشملها البحث الحالي - وتحتاج إلى مراجعة تقييمية ملحة من قبل معدي الامتحانين مع متخصصين في القياس والتقييم.

2.3.3. مؤشر الصعوبة Index of Difficulty: لا يُعد مؤشر صعوبة الاختبار دليلاً مباشراً للصدق؛ فهو لا يتعدى كونه متوسط أداء العينة على الاختبار؛ فقد عرّفه كل من Ebel & Frisbie (1991) بأنه: "النسبة المئوية لأداء العينة على الاختبار" (p.228). وقد يكشف مؤشر الصعوبة عن قدرات أفراد العينة ومهاراتهم وخبراتهم بصورة أكثر منطقية من كشفه عن خصائص الاختبار نفسه.

وعلى الرغم من الجدل القائم حول علاقة مؤشر الصعوبة بصدق الاختبار؛ فإن العديد من الدراسات السابقة مثل: (Latif & Alsunni, 2023; Simaremare, 2021) وظّفت مؤشر الصعوبة في التحقق من صدق أساليب تقييم طلاب المرحلة السريرية؛ لاعتبارات قد تبرز في كشف عن درجة ملائمة الاختبار لقدرات الطلاب المستهدفين به، فإذا كان الاختبار سهلاً جداً، أو صعباً جداً فهو لا يميز بين الطلاب، وبالتالي لا مصداقية لتفسير نتائجه؛ أما إذا كان مؤشر الصعوبة ضمن المدى المقبول؛ فهو يميز بين الطلاب المستهدفين؛ مما يشير إلى صدق بناء الاختبار في تفسير نتائج المجيبين؛ واستناداً إلى ذلك قام الباحث بحساب مؤشر صعوبة أساليب التقييم بقسمة مجموع نسب أداء المجيبين في أسلوب التقييم على عدد المجيبين؛ فكانت النتائج كما هي موضحة بالجدول الآتي:

الجدول 8

قيم مؤشر الصعوبة لأساليب التقييم بكل مقرر

المقرر	COM n=184	GYN n=178	OPH n=178	RDA n=178	FOR n=178	MED n=176	SUR n=176	PED n=176
الشفوي	.88**	.77*	.78*	.73*	.78*	.78*	.73*	.85**
العملي	.86**	.76*	.78*	.8*	.84**	.73*	.72*	.75*
التحريري	.82**	.70	.77*	.74*	.77*	.62	.65	.63
المتوسط	.85**	.74*	.77*	.76*	.8*	.71*	.7	.74*

ملحوظات. *سهل، **سهل جداً. وينبه الباحث هنا إلى أن عدد عينة الامتحان العملي بمقرر COM يساوي (86) فرداً

يلحظ من الجدول 8 أن قيم مؤشر الصعوبة امتدت ما بين (.62 - .86) بمتوسط عام بلغ (.76). وبالرجوع إلى مستويات الصعوبة ومقارنتها بالمدى المثالي لمؤشر الصعوبة الملائم لطبيعة أساليب التقييم المستخدمة وهو ما بين (.3 - .7) (Kaplan, 2017, p.174). فإن معظم أساليب تقييم الطلاب في مقررات المرحلة السريرية أعلى من المدى المثالي؛ مما يُستخلص منه أنها تتسم بالسهولة؛ وقد تُفسر هذه النتيجة بالمستوى المرتفع لقدرات طلاب المرحلة السريرية التي تضم طلاب المرحلة الثالثة بكلية الطب البشري، ومن المتوقع أن يكون أداء طلاب هذه المرحلة مرتفعاً، كما تبين من الجدول 8 أن أسهل أساليب التقييم هو الامتحان الشفوي يليه العملي وأصعبها الامتحان التحريري؛ حيث بلغ متوسط كل منها على التوالي (.79 ، .77 ، .71). وتختلف هذه النتيجة مع ما خلصت إليه دراسة Latif & Alsunni (2023) التي استهدفت تقييم أساليب التقييم بالمرحلة الثانية "الأساسية" فقد أظهرت نتائجها أن الامتحان التحريري أسهل من الامتحان العملي؛ وقد يعود ذلك إلى اختلاف المرحلة المستهدفة، ومن اللافت أيضاً أن أساليب تقييم الطلاب بمقرر طب الأسرة والمجتمع COM بمختلف أنواعها كانت سهلة جداً فقد امتدت ما بين (.82 ، .88). وقد يعود ذلك إلى طبيعة المقرر.

ورغم أن النتائج تشير -إحصائياً- إلى سهولة أساليب التقييم المستخدمة في تقييم طلاب المرحلة السريرية بكلية الطب البشري، وأنها خارج المدى المثالي لمؤشر الصعوبة؛ فإنها تشير من جانب آخر إلى كفاية قدرات طلاب المرحلة السريرية بالكلية، أو مدى صعوبة أساليب التقييم بالنسبة للطلاب المتقدمين لها، "فإن صعوبة الأسئلة تعتمد على الخبرات الطلاب وخلفيتهم العملية والشخصية، وتفاعلهم مع الأنشطة المقررة، واستخدامهم للموارد، وقدرتهم على التعلم الذاتي ... وينبغي أن تخدم صعوبة الاختبار غرض الحفاظ على صدقه، لا أن تكون غاية في حد ذاتها" (Burr, et al., 2026).

لذلك لا ينبغي الاكتفاء بمؤشر الصعوبة لتقييم أسلوب تقييم في مؤسسة تعليمية أو حتى نظام تعليمي؛ فقد يقيس المؤشر كفاية النظام التعليمي ونجاعته، أكثر من قياسه لسلامة أسلوب التقييم وكفايته. فالمعيار هو مدى قياس أساليب التقييم لمخرجات التعلم المستهدفة من دراسة كل مقرر، فإن أثبتت كفايتها في قياس جميع المخرجات؛ تُعد نتائج مؤشر الصعوبة دليلاً على صدق أساليب التقييم في ملائمتها لقدرات الطلاب.

3.3.3. مؤشر العدالة Index of Fairness:

تُعد العدالة (عدم التحيز) في التقييم من المعايير التي خلص إليها مؤتمر Ottawa للتعليم الطبي المنعقد عام 2010م لمواصفات الاختبار التقييمي الجيد، كما أنها من مؤشرات صدق أدوات القياس؛ حيث "يتطلب صدق الاستدلالات القائمة على الدرجات أن تعمل بنود الاختبار بنفس الكفاءة على جميع المستهدفين" (Dale, et al., 2025, p.2). ومن خصائص الاختبار التحصيلي الجيد أن يميز بين الطلاب وفقاً لمستوى تحصيلهم ولا شيء غير ذلك؛ "فيجب أن يُمثل الاختبار الصادق السمات أو القدرات المحددة التي يستهدف قياسها بدقة، وألا يقيس صفات أو مهارات أخرى خارج النطاق المقصود" (Stokes, 2024, P.3). ويُعد متغير الجنس (ذكر، أنثى) من أهم الصفات التي لا ينبغي أن تؤثر في أداء الطلاب على أساليب التقييم؛ لذلك فإن العديد من الدراسات التي استهدفت الكشف عن خصائص أساليب التقييم بكليات الطب البشري مثل: (Memon & Shaikh, 2020; Buzaid et al., 2023) اهتمت بدراسة أثر متغير الجنس على الأداء في الامتحانات. وقد استخدم الباحث اختبار (t . test) للكشف عن الفروق بين أداء الذكور والإناث؛ فكانت النتائج كما يلي:

الجدول 9

قيم (*t. test*) لدلالة الفروق بين متوسطي أداء عيني (الذكور، والإناث) على أساليب التقييم

المقرر	COM n=122	GYN n=122	OPH n=122	RDA n=122	FOR n=122	MED n=120	SUR n=120	PED n=120
الشفوي	-4.25**	1.44	0.58	1.34-	1.26-	1.93	1.42	0.85
العملي	-1.46	1.10	1.24	1.37	**2.66	1.13	0.94	-0.46
التحريري	0.58	0.09	1.43	0.61	0.4	**3.13	1.49	1.06
المجموع	-0.38	1.07	1.35	0.28	0.43	*2.47	1.80	0.51

ملحوظات: *دال عند مستوى (0.05)، ** دال عند مستوى (0.01). (-) الفروق لصالح الإناث. وينبه الباحث إلى أن عدد عينة الامتحان العملي بمقرر COM تكونت من (86) طالباً وطالبة.

يُلاحظ من الجدول 9 أن معظم قيم (*t. test*) لم تُظهر فروقاً دالة إحصائية بين الذكور والإناث؛ مما يُفسر بأن معظم أساليب التقييم المستخدمة بالمرحلة السريرية غير متحيزة لأي من الجنسين، وتتفق هذه النتيجة مع دراسة Buzaid et al. (2023) وتختلف مع دراسة Memon & Shaikh (2020) فيما عدا ثلاث حالات أظهرت نتائجها وجود فروق دالة إحصائية عند مستوى (0.01). هي: امتحان الشفوي في مقرر COM لصالح الإناث، وكانت لصالح الذكور في كل من: الامتحان العملي في مقرر FOR والامتحان التحريري في مقرر MED.

وقد حُسب حجم التأثير Effect Size للفروق الدالة إحصائية في الحالات الثلاث باستخدام قيمة مربع إيتا بالمعادلة $\eta^2 = t^2 \div (t^2 + df)$ حيث إن: (*t*) هي قيمة اختبار "*t-test*" و(*df*) هي درجة الحرية؛ فكانت قيم η^2 تساوي (0.13، 0.06، 0.08) على التوالي؛ وبذلك فإن حجم تأثيرها متوسط؛ لأنها ضمن المدى (0.06، 0.14) وهو ما يُفسر بأن الفروق ذات الدلالة الإحصائية بين الجنسين في الحالات الثلاث "قد تكون لها أهمية يُسترشد بها على صعيد البحث العلمي ويمكن أن تعمم، لكن إسهامها في بناء تراكم معرفي يُستند عليه في اتخاذ قرارات عملية يبقى محدوداً؛ لأن تعميم مثل هذه النتائج يكون على ذات المدة الزمنية وليس طويل الأمد" (بلعيد، 2021، ص.227).

4.3. الاستنتاج

إن أساليب تقييم طلاب المرحلة السريرية بكلية الطب البشري تسفر عن نتائج ذات توزيع اعتدالي يفيد بأن أساليب التقييم متسقة مع الواقع الفعلي للطبيعي للطلاب، كما أن أداء طلاب كل سنة دراسية يتسم بثبات مرتفع في الإطار العام لجميع المقررات، إلا أن ثبات أساليب التقييم في كل مقرر على حدة أظهرت النتائج أنه محدود، ويُفسر في ضوء تباين الأهداف التي يقيسها كل أسلوب، أما صدق تفسير نتائج أساليب فقد تحقق بمستوى مقبول من خلال مؤشرات التمييز والصعوبة وعدم التحيز.

ويُستخلص من ذلك أن أساليب التقييم المستخدمة لتقييم طلاب المرحلة السريرية بكلية الطب البشري ذات خصائص سيكومترية جيدة في حدود البحث والمؤشرات التي تحقق منها؛ باستثناء بعض الحالات الخاصة التي تمت الإشارة إليها خلال استعراض النتائج.

5.3. التوصيات: تعذر على الباحث التحقق من مؤشر صدق المحتوى لأساليب التقييم؛ لذلك يوصي الباحث بأن تحتفظ الكليات الطبية بنسخ من جداول المواصفات التي تم إعدادها لضبط محتوى كل اختبار، وأن تُجرى دراسات دورية مستمرة للتحقق من الخصائص السيكومترية لأساليب التقييم.

المراجع

العمرى، حسان غازي. (2018). المقارنة بين ثلاث طرائق في تقدير ثبات الاختبارات المركبة التي تتضمن نوعية من الفقرات (ألفا، ألفا الطبقي، راجو). *مجلة اتحاد الجامعات العربية للتربية وعلم*

النفس، 16 (2)، 85-102. <https://doi:10.35201/0246-016-002-004>

المركز الوطني لضمان جودة واعتماد المؤسسات التعليمية والتدريبية. (2021). *دليل اعتماد التعليم*

الطبي "الاصدار الثاني" [دليل]. (تاريخ الاطلاع: 12 أبريل 2026) <https://qaa.ly>

بلعيد، فتحي عمر ضو. (2021). الدلالة العملية لنتائج البحوث التربوية والنفسية المنشورة بمجلات الجامعة الأسمرية الإسلامية. *مجلة الجامعة الأسمرية: العلوم الشرعية والإنسانية*، 34(3)،

812-835. <https://doi.org/10.59743/jau.v34i3.374>

وزارة التعليم العالي والبحث العلمي-ليبيا. (2022). *دليل الدراسة بكليات الطب البشري بالجامعات الليبية*

[لليل]. (تاريخ الاطلاع: 8 مايو 2026). https://mhesr.gov.ly/?page_id=649

Al Alak, M., Al Hakeem, A. F., & Al Mousawi, A. (2017). Academic performance of final year medical students at kerbala medical college. *Journal of Contemporary Medical Sciences*, 3(9), 170-177.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Burr, S. A., Zahra, D., & Robinson, I. M. (2026). The Difficulty with Setting a Standard Based on Difficulty: The Role of Validity in Determining Assessment Standards. *Education Sciences*, 16(3), Article 488. <https://doi.org/10.3390/educsci16030488>

Buzaid, N., Lawgaly, S. A., Alawgali, S. M., Albash, A., & Alfakhri, M. (2023). Medical Students' Performances Using Different Assessment Methods during the Final Examination in Internal Medicine at the University of Benghazi, Libya. *Libyan International Medical University Journal*, 8(2), 70-75. <https://doi.org/10.1055/s-0043-1776309>

- Canhão, P. (2025). *Assessment methods in Undergraduate Medical Education: brief considerations. Jornal da Sociedade das Ciências Médicas de Lisboa*, 169(3), 44–51. <https://doi.org/10.57849/ulisboa.fm.jsclm.0000031.2025>
- Dale, E. D., Abulela, M. A. A., Jia, H., & Violato, C. (2025). Are medical school preclinical tests biased for sex and race? A differential item functioning analysis. *BMC medical education*, 25(1), 146. <https://doi.org/10.1186/s12909-024-06540-6>
- Dent, J., Harden, R. M., & Hunt, D. (2021). *A practical guide for medical teachers (6th ed.)*. Elsevier.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement (5th ed.)*. Prentice-Hall.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) (3 ed.)*. Thousand Oaks, CA: Sage.
- Kaplan, R. M., & Saccuzzo, D. P. (2017). *Psychological testing: Principles, applications, and issues (9th ed.)*. Cengage Learning. ISBN 9781337517065
- Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative dentistry & endodontics*, 38(1), 52-54. <https://doi.org/10.5395/rde.2013.38.1.52>
- Latif, R., & Alsunni, A. A. (2023). A comparison of the assessments used in campus-based years at the College of Medicine, Imam Abdulrahman Bin Faisal University, Saudi Arabia. *Postgraduate Medical Journal*, 99(1175), 1020-1026. <https://doi.org/10.1093/postmj/qgad005>
- Lister, J., Jaffer, S., Poepelman, R. S., Lane, K., Jankowski, K., Angell, K., & Violato, C. (2026). Implementation of Core Entrustable Professional Activities in the Internal Medicine Clerkship: A Psychometric Study. *Journal of general internal medicine*, 10.1007/s11606-026-10210-2. Advance online publication. <https://doi.org/10.1007/s11606-026-10210-2>
- Memon, S., & Shaikh, S. U. (2020). Comparison of performance on written and OSCE assessment during end semester pediatric examination. *Pakistan journal of medical sciences*, 36(4), 711–716. <https://doi.org/10.12669/pjms.36.4.2026>
- Mokkink, L. B., & Eekhout, I. (2026). The measurement properties reliability and measurement error explained - a COSMIN perspective. *Journal of clinical epidemiology*, 190, 112058. <https://doi.org/10.1016/j.jclinepi.2025.112058>
- Norcini, J.J. & McKinley, D.W. (2007). Assessment methods in medical education. *Teaching and Teacher Education: An International Journal of Research and Studies*, 23(3), 239-250. Elsevier Ltd
- Office of Educational Assessment. (n.d.). *Understanding item analyses*. University of Washington. Retrieved May 23, 2026, from

- Panzeri, A., Castelnovo, G., & Spoto, A. (2024). Assessing Discriminant Validity through Structural Equation Modeling: The Case of Eating Compulsivity. *Nutrients*, 16(4), 550. <https://doi.org/10.3390/nu16040550>
- Roemer, E., Schuberth, F., & Henseler, J. (2021). HTMT2—an improved criterion for assessing discriminant validity in structural equation modeling. *Industrial Management & Data Systems*, 121(12), 2637–2650. doi.org/10.1108/IMDS-02-2021-0082
- Simaremare, A. P. R. (2021). Analysis of Progress Test Results in Medical Faculty Students. *Jurnal Pendidikan Kedokteran Indonesia: The Indonesian Journal of Medical Education*, 10(1), 63-74. <https://doi.org/10.1016/j.jtumed.2013.12.002>
- Song, X., Cleaves, E., Gluzman, E., Kotlyar, B., Russo, R. A., Schilling, D. C., Tsao, C. P., & West, J. C. (2025). A Scoping Review of Assessments in Undergraduate Medical Education: Implications for Residency Programs and Medical Schools. *Academic psychiatry : the journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, 49(3), 263–273. <https://doi.org/10.1007/s40596-025-02136-4>
- Spielvogel, R., Stednick, Z., Beckett, L., & Latimore, D. (2012). Sources of variability in medical student evaluations on the internal medicine clinical rotation. *International Journal of Medical Education*, 3, 245–251. [doi.org](https://doi.org/10.1186/1752-4505-3-245)
- Stokes, R. (2024, November 15). Reliability and Validity in Psychometric Assessments for Hiring Decisions (White Paper). TTI Success Insights.
- World Federation for Medical Education. (2012). *Basic medical education: WFME global standards for quality improvement (2012 revision)*. <https://wfme.org>
- Zuo, J. (2021). Standard normal distribution test of college students' test score distribution. In *International Conference on Economics. 2021 5th International Conference on Economics, Management Engineering and Education Technology (ICEMEET 2021)*, Canada.

Psychometric Properties of Assessment Methods for Clinical Stage Students at the Faculty of Human Medicine, Asmarya Islamic University

Fathi Omar Daw Belaid

Department of Special Education, Faculty of Human Sciences - Girls, Asmarya Islamic University, Zliten, Libya.

dawfathi80@asmarya.edu.ly

Abstract

This research aimed to detect the psychometric properties of the assessment methods used for clinical stage students at the Faculty of Human Medicine, Asmarya Islamic University. A descriptive-analytical approach was employed to analyze the results of students in the clinical stage, which encompasses the fourth and fifth years, during which students study eight courses. The research

sample consisted of the entire student population of the clinical stage, totaling 184 male and female students across five academic years between 2017 and 2023.

The most prominent results revealed that the three assessment methods (oral, clinical/practical, and written) yield normally distributed results, indicating the consistency of these assessment methods with the actual, natural performance of the students. Furthermore, the students' performance in each academic year exhibited an acceptable reliability coefficient within the general framework of all courses; the Cronbach's alpha value reached .77 for the fourth year and .87 for the fifth year. Due to the varying relative weights of the assessment methods, reliability coefficients were calculated for each course individually using Raju's formula. These results showed that the reliability coefficients fell below the acceptable level, ranging from .44 to .64, which was attributed to the diverse objectives measured by each distinct assessment method.

To detect the validity indicators for interpreting the assessment results, the researcher verified three indicators: discrimination, difficulty, and test fairness. These indicators demonstrated the validity of interpretation for most assessment methods, except for a few individual cases. Based on these findings, the researcher concluded that the assessment methods possess good psychometric properties within the scope of this research and the verified indicators, with the exception of certain cases that require scientific explanation from the administrators and necessary measures to improve their psychometric properties. The researcher recommended the crucial need to verify the content validity indicator of assessment methods before employing them in decision-making regarding the extent to which students achieve course objectives.

Keywords: *Reliability and Validity, Psychometric Properties, Human Medicine, Clinical Stage, Assessment Methods.*