

## Building a Rubric for Peer Review of Manuscripts in Psychological and Behavioral Sciences

Haidar Zaza

College of Educational Sciences, The University of Jordan, Amman, Jordan.

Received: 11/4/2020  
Revised: 27/6/2020  
Accepted: 14/7/2020  
Published: 1/6/2021

Citation: Zaza, H. (2021). Building a Rubric for Peer Review of Manuscripts in Psychological and Behavioral Sciences. *Dirasat: Educational Sciences*, 48(2), 127-143. Retrieved from <https://dsr.ju.edu.jo/djournals/index.php/Edu/article/view/2714>

### Abstract

The purpose of this research is to build a rubric that can be used in peer review of research manuscripts in educational and psychological sciences. This rubric is built according to the Helvoort model for performance tasks, and its validity in the construction phase was confirmed by opinions of referees and its reliability was calculated by interrater reliability. The results of the construction stages showed the consensus of studies, methodologies in educational and psychological sciences and review forms used in journals on (6) criteria, namely: introduction, methodology, results, discussion, conclusions and recommendations, references and documentation. The results of the evaluation of rubric by judgments showed the comprehensiveness of its criteria and its relevance to the assessment of manuscripts, assisting the reviewers in focusing on research components, shortening review time, and achieving interrater reliability. It makes the beneficiaries (editor, reviewer, and author) trust its effectiveness and use it in review. Although the rubric is based on best practices in reviewing research manuscripts in educational and psychological sciences, it may not be generalized to other fields. It also provides a general and important vision that should be discussed further in the future.

**Keywords:** Peer review, refereeing, research manuscripts, rubric, educational & psychological sciences.

### بناء سلم تقدير لفظي "روبك" لمراجعة الأقران للمخطوط البحثي في العلوم التربوية والنفسية

حيدر زازا

الجامعة الاردنية، الأردن.

### ملخص

الغرض من هذا البحث هو بناء "روبك" يُستخدم في مراجعة الأقران للمخطوط البحثي في العلوم التربوية والنفسية. تم بناء هذا "الروبك" وفق نموذج هيلفورت Helvoort للمهام الأدائية، وجرى التأكد من صدقه وثباته في مرحلة البناء من خلال آراء المحكمين وثبات المقيمين. أظهرت نتائج مراحل البناء اجماع الدراسات والمنهجيات في العلوم التربوية والنفسية ونماذج المراجعة المستخدمة في المجلات على (6) محكات، وهي: المقدمة، المنهجية، النتائج، المناقشة، الاستنتاجات والتوصيات، المراجع وتوثيقها، التي تندرج تحتها (16) عنصراً فرعياً. كما أظهرت نتائج تقييم "الروبك" من قبل المحكمين شمولية محكاته وصلتها بتقييم المخطوط، ومساعدته المقيمين في التركيز على مكونات البحث واختصاره زمن المراجعة، إضافة إلى ثبات بين المقيمين مقداره (0.96)؛ مما يجعل المستفيدين (المحرر، المراجع، المؤلف) يتقنون بفاعليته ويستخدمونه في المراجعة. وبالرغم من أن الروبك يستند إلى أفضل الممارسات في مراجعة المخطوطات البحثية في العلوم التربوية والنفسية، إلا أنه قد لا يكون قابلاً للتعميم على الحقول الأخرى، كما أنه يُقدم تصوراً عاماً ومهماً ينبغي مناقشته على نحو أوسع في المستقبل. أنجز هذا البحث في إجازة التفرغ العلمي في العام 2018/2019.

الكلمات الدالة: مراجعة الأقران، التحكم، المخطوطات البحثية، سلم تقدير لفظي "روبك"، العلوم التربوية والنفسية.



© 2021 DSR Publishers/ The University of Jordan.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license <https://creativecommons.org/licenses/by-nc/4.0/>

## المقدمة

تُحَفِّزُ معظم الجامعات الباحثين فيها على النشر في المجلات المصنفة (المفهرسة) Indexed journals في قواعد البيانات العالمية، مثل: شبكة العلوم Web of science، وسكوبس Scopus وغيرهما، وذلك لتحقيق أهداف عدة، ومن أهمها إحراز تقدم على معيار جودة البحث ومؤشراته الذي هو أحد معايير التصنيف الأكاديمي للجامعات: مثل تصنيف شنغهاي (Academic Ranking of World Universities [ARWU], 2018) وتصنيف مركز دراسات العلوم والتكنولوجيا (Centre for Science and Technology Studies [CWTS], 2018) وغيرهما. فهذه المجلات وإن كانت تختلف في شروطها وقواعدها، إلا أنها تشترك في ما بينها في مجموعة شروط من أهمها: ارتباط موضوع المخطوط البحثي Research Manuscripts المقدم للنشر بنطاق المجلة Journal scope، وتحقيقه معايير التحرير Editorial criteria الخاصة بالمجلة، والتأكد من مطابقة المخطوط الميثاق الأخلاقي للبحوث الذي تعتمده المجلة، إضافة إلى تحقيقه عددا من المحكات التي تُشكّل في الوقت ذاته مكونات أساس يجب توفرها في البحوث بمواصفات محددة تختلف باختلاف الحقل (A guide to becoming a peer reviewer - Editor Resources, 2019).

ففي العلوم التربوية والنفسية، يتفق العديد من الباحثين (Albarracín, 2015; Albert, Laberge, & McGuire, 2012; Bhattacharjee, 2012; Kerlinger, 1986; Teddlie & Tashakkori, 2009) حول المكونات الأساس التي يجب أن تتضمنها البحوث، وهذه المكونات هي: مشكلة البحث وأهميتها، المصطلحات، عرض الدراسات المرتبطة بالبحث، اختيار أو بناء الأدوات، جمع البيانات وتحليلها، مناقشة النتائج، التضمينات العملية وحدودها، اقتراح الدراسات التي تحتاج إلى مزيد من البحث، وأخيراً التوثيق.

كما تقدم المجلات المصنفة في العلوم التربوية والنفسية للمؤلفين إرشادات تتفق مع تعليمات الجمعية الأميركية لعلم النفس American Psychological Association (APA) لإعداد المخطوط البحثي، التي تقترح أن يتكون تقرير البحث باختلاف التصاميم والمناهج البحثية المستخدمة من العناصر التالية: العنوان، الملخص، متن التقرير (المقدمة، الأساليب والطرق، النتائج، المناقشة)، قائمة المراجع، الملحق (Publication Manual of the American Psychological Association, 2018). ويكون اتباع هذه الإرشادات متطلباً سابقاً من أجل تحويل المخطوط من المحرر إلى المراجع. وفي أثناء المراجعة يُستخدم جزء من هذه التعليمات كمحكات لمراجعة المخطوط والتصديق على صحة الإجراءات التي تحدد مدى معقولية النتائج، تبعاً لسياسة المراجعة التي تتبعها المجلة، التي تتباين من مجلة لأخرى (Chubin & Hackett, 1989). حيث يجري التأكد من تحقيق المخطوط للمحكات التي تُشكّل في الوقت ذاته مكونات أساس من خلال إخضاع المخطوط لمراجعة الأقران، إذ تزود هذه المجلات المراجعين بتعليمات توضح مسؤوليتهم وما يتوجب عليهم فعله، كما تزود بعض المراجعين بنموذج مراجعة أو نموذج تعليقات Reviewer's Comments form وتسمح أخرى للمراجعين برفع توصيات وملفات نقد إضافية إلى المحرر.

واليوم، وبالرغم من أن مراجعة الأقران باتت جزءاً أساسياً في المشهدين العملي والبحثي (Devine & Frass, 2016; Marsh, Jayasinghe, & Bond, 2008). وأمرًا ضروريًا لوجود المحتوى البحثي (McNair, Le Phuong, Cseri, & Szekely, 2019). إلا أن العديد من المهتمين والمراقبين يعربون عن أسفهم حيال النتائج الاحتمالية أو المعيبة التي لا تزال تصل إلى صفحات المجلات التي راجعها الأقران (Baldwin, 2017). في المقابل، يرى ريني Rennie (2003) بأن الأكاديميين والباحثين يتقبلون هذه العملية على الرغم من جوانب الضعف التي تبدو فيها. ولعل من أكثر الإشكالات التي تعترض مراجعة الأقران، تلك المتعلقة بالنماذج المستخدمة في المراجعة، كما أكدت ذلك الدراسة واسعة النطاق التي أجرتها شركة تايلور وفرانيسيس Taylor & Francis في العام 2015 وشارك فيها ما يزيد عن 7400 مؤلف ومراجع ومحرر من جميع أنحاء العالم، ومن جميع التخصصات الأكاديمية لتعرّف آرائهم حول الجوانب المختلفة لعملية مراجعة الأقران، التي أكدت نتائجها ضرورة تقييم النماذج المطورة على نحو علني، وضرورة بذل مزيدٍ من الجهد لضمان تعزيز مراجعة الأقران. والاهتمام بحفظها في المستقبل (Devine & Frass, 2016). وهذا ما أكدته لاحقاً هورباخ وهالفمان (Horbach and Halffman, 2018) أن هناك حاجة واضحة لإجراء تحليل منهجي لنماذج مراجعة الأقران وما يرتبط بها.

تعدّ مراجعة الأقران Peer review وتُعرف بنصيحة الأقران Peer advice، تقييم الأقران Peer evaluation، حكم الأقران Peer judgment، مراقبة الجودة Quality control، رقابة الأقران Peer censorship، مراجعة الجدارة Merit review، والتحكيم Refereeing الآلية الرئيسية التي تستخدمها المجلات لتقييم جودة المؤلفات العلمية (Chubin & Hackett, 1989). وبحسب نغوين وآخرون (Nguyen et al., 2015)، فإن المجلات التي تتم مراجعتها من قبل الأقران ستبقى الوسيلة الأساس التي ندقق بها في البحث العلمي، وننقل من خلالها اكتشافات جديدة إلى زملائنا العلماء والمجتمع ككل. ويؤيد ذلك نتائج دراسة ألبرت وآخرين (Albert et al., 2012)، التي شملت مقابلة 94 عالماً من علماء الطب الحيوي والعلماء السريريين، والعلماء في الحقول الاجتماعية، الذين يعملون في الجامعات الكندية الذين أكدوا أن مراجعة الأقران أفضل آلية لمراقبة الجودة لتقييم البحث العلمي، وأن مراجعة الأقران هي السمة الرئيسية لشرعية البحوث في الساحة الأكاديمية.

وحول أهمية المراجعة، فهي تمنح المؤلف فرصة فريدة لتعزيز المخطوط بمساعدة خبير في مجال اختصاص المخطوط. فقد أشارت دراسة أُجريت في جامعة تنسي (University of Tennessee and CIBER Research Ltd, 2013) إلى أن 90٪ من الباحثين يشعرون بأنّ مراجعة

الأقران حسّنت من نوعية أوراقهم المنشورة.

هناك أشكال مختلفة من مراجعة الأقران بالرغم من أنّ أساسها واحد، وهو أنّ يضع خبراء التخصص تعليقات على الورقة للمساعدة في تحسينها. كما تتعدد أنواعها، فبحسب هويتها Identifiability تنقسم إلى ثلاثة أنواع: المراجعة فردية التعمية Single-blind review، المراجعة مزدوجة التعمية Double blind review والمراجعة المفتوحة Open review (A guide to becoming a peer reviewer - Editor Resources, 2019). وهناك من يصنفها إلى: المغلقة، التي يُدرك فيها المراجع هوية المؤلف، ولكن لا يعرف المؤلف من راجع مخطوطه. ومزدوجة التعمية، وفيها لا يعرف المؤلف ولا المراجع هويّات بعضهم بعضًا. والمراجعة المفتوحة، حيث يكون المراجع على معرفة بهوية المؤلف، وتكشف للمؤلف هوية المراجع. وفي بعض الحالات تنشر المجلات أيضًا تقارير المراجعين جنبًا إلى جنب مع المخطوط المنشور (Springer -International Publisher Science, Technology, Medicine, 2018).

ترتب على اختلاف سياسات مراجعة البحوث اختلاف عملية المراجعة من مجلة لأخرى، وبالرغم من ذلك، فإن معظمها تتفق على مراحل أساس. وحاليًا أصبحت هذه المراحل تتم عبر الإنترنت Online review. حيث تبدأ باتصال المجلة بالمراجع المحتمل ودعوته لتنفيذ المراجعة، ثم يُعطى المراجع عنوان المخطوط البحثي والملخص والإطار الزمني، وفي ضوء ذلك يوافق المراجع على مراجعة المخطوط أو يرفض، بعد ذلك يتلقى المراجع المخطوط ثم ينفذ المراجعة ثم يرفعها للمحررين، وبعد ذلك يفحص المحررون المراجعات وطلب الحصول على مراجعات إضافية - إذا لزم الأمر. ويتم اتخاذ القرار، ثم يرسل القرار إلى المؤلف (الباحث) مع تعليقات المراجعين، وأخيرًا يتم شكر المراجعين وقد يتم إعلامهم بالقرار، وقد يتلقون نسخة من التعليقات المرسلة إلى المؤلف (Marsh et al., 2008; Rockwell, 2005). وكما هو متعارف عليه، فمن غير المعتاد قبول ورقة للنشر دون تنقيحها مرة واحدة على الأقل في ضوء تعليقات المراجعين والمحررين (Authorservices.wiley.com, 2018). فما يريده المحرر من المراجع، هو أنّ يكون لديه خبرة في مجال واحد أو أكثر، ويتصف بالموضوعية ويراعي عدم تضارب المصالح، القدرة على التحكيم الجيد والتفكير بوضوح ومنطقية، القدرة على كتابة نقد جيد (دقيق ومقروء ومفيد للمحررين وللمؤلفين)، يمكن الاعتماد عليه في إعادة المراجعة، ولديه القدرة على القيام بالمراجعة في الإطار الزمني المحدد (Rockwell, 2005).

وحول مضمون محكات المراجعة، فهناك مناحٍ متعددة يتبناها المراجعون في مراجعاتهم، لكن أحدث الدراسات أظهرت أنّ المراجعين المحترفين يعتمدون إجراء واحدًا من ثلاثة إجراءات عامة (المنحى القائم على المحتوى، المنحى القائم على العملية، المنحى الذي يجمع الاثنين معًا) وذلك عند مراجعة المخطوط (Leung, Law, Kucukusta, & Guillet, 2014). وتكاد تتفق معظم المجلات في الحقول المختلفة على خطوات مراجعة الأقران، حيث يُطلب من المراجع بعد الانتهاء من قراءة المخطوط البحثي تعبئة نموذج للمراجعة Review form (غالبًا قائمة شطب Checklist) يتضمن فقرات تختلف من قائمة لأخرى، تسأل حول مناسبة المخطوط للمجلة، دلالة القضية المطروحة، الاهتمام بالقراءة، جودة البحث (التصميم والتحليل)، مناسبة البيانات، تغطية العنوان، علاقته بمراجعة الأدب، المساهمة في المعرفة الجديدة، سهولة الجداول والأشكال (Provenzale & Stanley, 2005). وفي المقابل، تتضمن بعض النماذج فقرات تقييم مثل: التقييم العام للورقة، المساهمة في المعرفة الحالية، التنظيم والمقرئية، المنهجية، الأدلة الداعمة للخلاصة، مناسبة مراجعة الأدب، جوانب القوة وجوانب الضعف والاقتراحات. وفي بعض المجلات لا تقبل هيئة التحرير تحكيم البحوث دون ملء التقرير التفصيلي، أو نموذج تعليقات المراجع مبيّنًا فيه أوجه القصور ومقترحات لتحسين البحث، وأن يتضمن التقرير معلومات عن محاور التحكيم.

وبالنسبة للمجلات التي تعتمد نموذج مراجعة على شكل قائمة شطب أو مقاييس التدرج Rating scale، فغالبًا ما تنص تعليمات المراجعة على تقييم المخطوط باستخدام تدرج ثنائي أو ثلاثي أو رباعي أو خماسي أو أكثر، مثل: ضعيف جدًا (1)، ضعيف (2)، متوسط (3)، قوي (4)، وقوي جدًا (5). ثم تُجمع الدرجة على مقياس فنوي يتراوح بين 0 و100، وبعضها يتطلب الإجابة عن أسئلة باستخدام: نعم، لا، لا ينطبق. وفي ضوء الدرجة الكلية يتم إعطاء إحدى التوصيات، وهي: (1) القبول Accept، وتعني بأن المخطوط بصيغته الحالية مقبول للنشر، ولن يكون هناك سوى تعديلات ثانوية طفيفة؛ (2) المراجعة البسيطة Minor revision، وهنا يتطلب المخطوط تعديلات طفيفة لأنه غير مقبول في صيغته الحالية؛ (3) المراجعة الرئيسية Major revision، وهنا يتطلب المخطوط إجراء تصويب شامل، وهو غير مقبول في حالته الحالية، (4) الرفض Reject، أي أنّ المخطوط لا يحقق المعايير ليتم نشره في هذه المجلة (Apa.org, 2019; Rosenfeld, 2010). وفي التقييم الختامي Overall Evaluation يُطلب من المحكم وضع تعليقات مكتوبة حول المخطوط إن لزم الأمر، أو تعليقات حرفية Verbatim Comments من أجل التحسين والمراجعة. وتعمل بعض المجلات على توفير مساحة للمراجع في نموذج المراجعة لكتابة تعليقات محددة للمؤلف والمحرر، وفيها يقدم معظم المراجعين تقييمًا مفصلاً للمخطوط، وغالبًا ما يشيرون إلى أرقام السطور والفقرات في المتن من أجل تقديم تصويبات أو تعليقات دقيقة.

ويبين استعراض المواقع الإلكترونية للمجلات - التي شكلت عينة المجلات في الدراسة الحالية - في العلوم التربوية والنفسية، أنّ قرار إرسال الأوراق للمراجعة الرسمية يعتمد على مدى ارتباط موضوعها بنطاق المجلة، وفي تحقيقها لمعايير التحرير الخاصة بها. إذ يتحول جزءٌ من هذه المعايير

ليصبح محكًا في نموذج المراجعة الذي تعتمده المجلة، وقد تؤثر معايير التحرير في الحكم العام وفي الانطباعات الشخصية التي يتركها المراجع في متن المخطوط عند مراجعته.

في المقابل، قد تبتعد آليات مراجعة المخطوط والنماذج المستخدمة فيها عن فحص المكونات الأساس، ظنًا بأن التعليقات التي يتركها المراجع في متن المخطوط وتقديره لدرجة معينة وفقًا لنموذج المراجعة يفي بالمطلوب. ففي تحليل أجراه بروج وكاليه (Bordage and Caellegh, 2001) لمحتوى تعليقات المراجعين على 151 مخطوطًا بحثيًا قُدمت إلى مجلتي مؤتمر البحث في التعليم الطبي للعامين 1997 و1998، تبين أن المراجعين حددوا عيوبًا قاتلة من المرجح أن تكون تبريرًا للرفض مثل: (تجاهل الأدبيات، ضعف تصميم الدراسة، اختيار الأدوات غير الملائمة، الكتابة الركيكة للمخطوط). بل تجاوز ذلك أن بعض الباحثين مثل ماككر، لو، وير، هيان وهسو (McKercher, Law, Weber, Haiyan, and Hsu, 2007) حددوا نسب أكثر المكونات الأساس انتهاكًا، التي كانت سببًا في رفض المخطوطات، حيث أظهرت نتائج تحليل محتوى 373 تقريرًا عن المخطوطات المقدمة إلى 35 مجلة ضيافة وسياحة، أن أكثر المكونات التي وجد فيها المراجعون أخطاء في المخطوطات كانت المنهجية وبنسبة 74% من الأوراق، والفشل في توضيح الأهمية بفاعلية وبنسبة 60%، وضعف أسلوب الكتابة وبنسبة 58%، والمراجعة الضعيفة للأدب وبنسبة 50%، علمًا بأن بعض الأوراق قد تحوي نوعين أو أكثر من الأخطاء.

ولما كانت عملية مراجعة الأقران ونماذجها يشوبها القصور الذي تم استعراضه آنفًا؛ فإن ذلك يعزز ضرورة التفكير في بناء نموذج مراجعة يعالج قصور محكات تقييم المخطوط، ويجعلها أكثر حساسية في الكشف عن انتهاكات المؤلفين للمكونات الأساس للبحوث. وهنا، يأتي دور سلم التقدير اللفظي "روبك" المقترح في التركيز على المكونات الأساس للبحوث من خلال تقديمها كمحكات يسهل على المراجع فهمها وتصنيف المخطوط وفقًا لتدرج قائم على المنطق وابتعد عن الانطباعات الشخصية. إذ بيّنت نتائج دراسة كاياينار (Kayapinar, 2014) بأن أدوات التقييم المتنوعة (مثل: قوائم الشطب، سلالم التقدير) لا تخلو من الذاتية، حتى وإن تم تقدير الدرجات عليها من قبل أكثر من مقيم (مُقدّر).

إن استخدام "روبك" للأغراض البحثية واعد للغاية، كما يذكر كلمنت وتشافوت وفيليب وامبروس (Clement, Chauvot, Philipp, and Ambrose, 2003) بالرغم من أن الدراسات التي بحثت استخدامه في تقييم المخطوط البحثي تكاد تكون نادرة نسبيًا كما تبين من مراجعة الأدب. وبما أن لاستخدام "روبك" في الأوضاع التعليمية المختلفة مزايا عديدة، فإن هذا يؤسس لفرض وجود أثر إيجابي -أيضًا- عند استخدامه في تقييم المخطوط البحثي. وفحص هذا الفرض يتطلب تطوير سلم التقدير اللفظي أولًا، ثم فحص أثره. وهذا -فعلاً- ما جاءت من أجله الدراسة الحالية لتحقيق متطلبات الفرض الأول، وهو بناء سلم التقدير اللفظي لتقييم المخطوط ليبقى الباب مفتوحًا لفحص فاعليته مستقبلاً. إذ يتوقع أن يسهم سلم التقدير اللفظي المقترح في زيادة شفافية عملية مراجعة الأقران التي تُعد مؤثرًا على جودتها (Wichert, 2016). لذلك، فإن هذه الدراسة محاولة من الباحث لبناء سلم تقدير لفظي "روبك" يمكن استخدامه في الحصول على تقديرين رقمي ونوعي لكل محك عند مراجعة المخطوط البحثي.

مثل أي أداة تقييم أخرى، يُعد سلم التقدير اللفظي "روبك" مفيدًا لأغراض معينة. فالغرض الرئيس منه هو تقييم نوعين من الأداء: العمليات (Processes) والمنتجات (Products)، إذ ينضوي تحت المنتجات: تقييم المقالات المكتوبة، الموضوعات (Themes)، التقارير، المصطلحات (Brookhart, 2013).

هناك نوعان من "روبك"، الروبك الكلي (Holistic rubrics) والروبك التحليلي (Analytic rubrics)، وكلاهما من مقاييس التدرج التي تنضوي بدورها تحت أدوات تقييم الأداء (Scoring instruments for performance assessments) إلى جانب قوائم الشطب (Mertler, 2001). يُعد سلم التقدير اللفظي "روبك" أداة رصينة تُستخدم في الموضوعات التعليمية المتنوعة مثل تقييم الطلبة وأداء المعلمين ومراجعة المناهج الدراسية (Martens, 2018). وفي تقييم البرامج (Reddy & Andrade, 2010). وهو واسع الاستخدام في المدارس وفي جميع المراحل، كما يستخدمه طلاب البكالوريوس والدراسات العليا في موضوعات عديدة ومنها تقييم رسائل الماجستير (Andrade, 2005; Prins, De Kleijn, & Van Tartwijk, 2017). وتدعم نتائج دراسة بروخارت وتشن (Brookhart & Chen, 2015) استخدامه في جميع المستويات التعليمية باستثناء الطفولة المبكرة.

وكما أن سلم التقدير اللفظي معروف في الموضوعات التعليمية المختلفة، فإنه أيضًا يساعد في قياس المهارات العليا أو تقييم المهام المعقدة، وفي توضيح الأهداف الغامضة، ويساعد الطلاب على فهم توقعات المعلم وعلى التنظيم الذاتي (Self-regulation) والتعلم الإيجابي، كما أنه يعزز التعليم ويحسنه (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). ويساهم في زيادة تحصيل الطلبة وتحسين التدريس (Reddy & Andrade, 2010). ويساعد المعلمين في التدريس، وفي تنسيق التدريس والتقييم، كما يساعد الطلاب على التعلم. وكذلك يعزز التقييم الذاتي وفهم محكات التقييم (Brookhart & Chen, 2015; Cockett & Jackson, 2018; Panadero & Jonsson, 2013).

ويمكن لسلم التقدير اللفظي أن يُحسن من أداء الطلاب على نحو أفضل وأن يحسن التغذية الراجعة لهم، كما يحسن من التغذية الراجعة

لأعضاء هيئة التدريس والموظفين من خلال إيجاد أنماط التحصيل أو الخطأ لدى الطلبة، وتوفير معلومات تشخيصية حول نقاط القوة والضعف لديهم (Suskie, 2009). كما يمكن لسلم التقدير اللفظي الذي يُشارك في بنائه الطالب Co-creating rubrics من زيادة فرصة الطالب في أن يكون له دور في معايير تقييمه والتدخل فيها (Fraile, Panadero, & Pardo, 2017). بل يتعدى الأمر ذلك، فاستخدام سلم التقدير اللفظي من قبل الطلاب والمعلمين على حدٍ سواء يؤسس لما يسميه كوكس وموريسون وبراثويت (Cox, Morrison, and Brathwaite (2015) بآلية التغذية الأمامية feed forward mechanism وهذا يعني أنّ سلم التقدير اللفظي يجب أن يسمح للطالب في التفكير في عمله والتركيز في كيفية تحسينه في المستقبل، فهو ليس مجرد استخدام الطالب لسلم التقدير كشكل من أشكال التغذية الراجعة ولكن يمنحه فرصة لاستخدام التغذية الراجعة ذاتها. وفي هذا الصدد، تذكر بروخارت (Brookhart (2013 أنّ سلم التقدير اللفظي مهم لأنه يوضح للطلاب الجودة التي يجب أن يتمتع به عملهم، التي غالبًا ما يعبر عنها بفهم الطلاب لهدف التعلم ومحكات النجاح. ويمكن لسلم التقدير اللفظي أن يساعد في تعزيز الحكم المهني Professional judgement التي غالبًا ما يعبر عنها بمصدرًا تعليميًا لتحديد ودعم عمليات التفاوض والاتفاق التي تميز ثقافة التقييم (Menéndez-Varela & Gregori, 2018). في حين يرى ديكنسون وأدمز (Dickinson and Adams (2017 أنّ سلم التقدير اللفظي يُمثل أحد الحلول الممكنة لتحديد الجدارة والأهمية للتقييم، وهو وسيلة المقيمين الذين يسعون إلى نقل التقييم من مجرد وصف التقييم (مثل البرنامج أو المشروع أو السياسة التي سيتم تقييمها) إلى تحديد جودة التقييم ونجاحه.

وحول الخصائص السيكمومترية لسلم التقدير اللفظي، فقد أكدت دراسات عديدة صدقه وثباته في الكثير من عمليات البحث والتقييم، ولعل أكثر طرق الثبات استخدامًا حسب نتائج مراجعة الدراسات التي أجرتها بروخارت وتشن (Brookhart and Chen (2015 هي الاتفاق بين المقيمين Agreements among raters، أما أكثر طرق الصدق استخدامًا فهي صدق البناء (المفهوم) Construct validity. وقد أشارت دراسة لونغرين، أدوي وسفانستروم (Lönngren, Adawi, and Svanström (2019 إلى أنّ سلم التقدير اللفظي يمكن أن يجعل التقييم ثابتًا وصادقًا إذا تم تدريب المقيمين بعناية. وفي مراجعة شملت 75 دراسة تجريبية لسلم التقدير أجراها جونسون وسيفنجي (Jonsson and Svingby (2007 كانت إحدى استنتاجاتها أنه يمكن تعزيز ثبات التصحيح لتقييم الأداء من خلال استخدام سلم التقدير اللفظي خاصة إذا كان تحليليا ومحدد الموضوع ويتضمن أمثلة وتم تدريب المقيم عليه. في المقابل أشارت نتائج دراسة رضائي ولوفورن (Rezaei and Lovorn (2010 إلى أن استخدام سلم التقدير اللفظي قد لا يحسن من الثبات أو الصدق إذا لم يكن المقيمون مدربين جيدًا على كيفية تصميمه وتوظيفه بفاعلية. وفي ما يتعلق بثبات النوعين (الكلي والتحليلي) فقد أكد بعض الباحثين مثل جوننج (Gunning (2006 أن سلم التقدير اللفظي التحليلي أكثر ثباتًا من سلم التقدير اللفظي الكلي.

وهكذا، نرى أنّ سلم التقدير اللفظي "الروبرك" ومن خلال الموضوعات المتنوعة التي يُستخدم فيها، إضافةً إلى دلالات صدقه وثباته، وخفضه للذاتية عندما يقدّر الدرجات عليه مقيمان اثنان أو أكثر، وبحكم نجاحه في مساعدة الطلاب والمعلمين على تحديد "الجودة"، واستخدامه من قبل الطلاب على نحو منظم للحكم على عملهم وقبولهم مزيدًا من المسؤولية عن المنتج النهائي، وتقليله من الوقت الذي يمضيه المعلمون في تصنيف عمل الطلاب والتسهيل عليهم في الشرح حول حصولهم على تقدير معين عن العمل الذي أجروه، وما الذي يمكنهم فعله لتحسينه (Andrade, 2005). إن ذلك كله يجعل سلم التقدير اللفظي "الروبرك" يتفوق على الأدوات الأخرى ويبرر استخدامه في أوضاع جديدة، وهو تقييم المخطوط خاصة وأن العديد من الدراسات مثل (Cicchetti, 1991; Wood, Roberts, & Howell, 2004) تؤكد انخفاض ثبات مراجعة الأقران على نحو عام، وأن أحد أسباب هذا الانخفاض هو عدم ثبات الأدوات المستخدمة في المراجعة (Laming, 1991). على ضوء ما سبق، يتبين أنّ سلم التقدير اللفظي المناسب لهدف الدراسة الحالية هو سلم التقدير اللفظي التحليلي، وذلك لأن الهدف منه هو تقدير درجات للمحكات التي يجب أن تتوفر في المخطوط عند مراجعته. وهذا ما أكدته بروخارت وتشن (Brookhart & Chen (2015 أن سلم التقدير اللفظي "الروبرك" يقدم معلومات ذات جودة مناسبة إذا تم استيفاء شروط معينة وأبرزها وجود محكات واضحة ومركزة.

#### مشكلة الدراسة وأسئلتها

تهدف هذه الدراسة إلى بناء سلم تقدير لفظي "روبرك" يُستخدم في مراجعة الأقران للمخطوط البحثي وفق نموذج هيلفورت (Helvoort (2010 الذي يمثل الصورة المختصرة لخطوات بناء سلم التقدير اللفظي التحليلي والكلي التي وضعها ميرتلر (Mertler (2001 بوصفه أحد النماذج المستخدمة على نطاق واسع في بناء "الروبرك". إذ تمثلت مشكلة هذه الدراسة في الحاجة إلى أداة تقدير لتقييم المخطوط البحثي في العلوم التربوية والنفسية. حيث تُظهر الأدبيات العالمية أن قواعد النشر في الدوريات التربوية حديثة العهد نسبيًا إذا ما قورنت بتلك المتعلقة بالدوريات العلمية (Graue and Galego, 2006).

ويُظهر فحص عدد من نماذج المراجعة المستخدمة في مجالات العلوم التربوية والنفسية استحوادها على قوائم الشطب، إضافةً إلى خلوها من

بعض المحكات التي تُشكل في الوقت ذاته مكونات أساس يتفق الباحثون على وجودها وفق مواصفات محددة في بحوث هذا الحقل، وهذا يجعل فرضية خلو المخطوط من بعض المكونات أو انتهاك مواصفاتها من قبل المؤلفين بدون اكتشافها في أثناء المراجعة متوقعة الحدوث؛ وذلك بسبب أن نموذج المراجعة المستخدم لم يكن حساسًا كفايةً للكشف عن هذا الانتهاك، وقد يتجاهل المراجع هذه المكونات ومواصفاتها، أو قد لا ينتبه إليها في أثناء المراجعة في حال أجريت دون استخدام نموذج المراجعة، وقد يكتفي المراجع بتقدير درجة معينة تبين مدى تحقيق الورقة لمحكات النموذج؛ ويرتب على ذلك أن قرارات مثل قبول الورقة أو رفضها أو إعادة مراجعتها قد تصبح موضع شك، وكذلك الحال لقرار تصنيفها؛ مما يؤثر سلبًا في مستوى جودة البحوث التي تسعى الجامعات إلى تحقيقها كأحد معايير التصنيف الأكاديمي. وهذا كله يزيد من أهمية أداة المراجعة التي يستخدمها المراجع للتحقق من توفر المكونات في المخطوط وفق مواصفات محددة في حقل التخصص.

من هنا، فقد تمثلت مشكلة الدراسة في سعيها إلى بناء سلم تقدير لفظي (روبرك) لمراجعة الأقران للمخطوط البحثي في العلوم التربوية والنفسية واستقصاء خصائصه السيكمترية. ولتحقيق ذلك، فقد طرحت هذه الدراسة السؤالين التاليين:

1. ما إجراءات بناء سلم تقدير لفظي "روبرك" لمراجعة الأقران للمخطوط البحثي في العلوم التربوية والنفسية؟
2. ما دلالات الصدق والثبات التي يمكن توفيرها لسلم تقدير لفظي "روبرك" لمراجعة الأقران للمخطوط البحثي في العلوم التربوية والنفسية؟

### أهمية الدراسة

تأتي هذه الدراسة كمحاولة لوضع اللبنة الأساس لدراسة بعض الجوانب التي لم تنل قسطها من الاهتمام وتحليلها، سعيًا إلى تحقيق إضافة نوعية في جودة البحوث، واستكمالًا لجهود الباحثين في تحقيق فهم أعمق لعملية المراجعة من خلال فحص نماذجها المستخدمة في مجالات العلوم التربوية والنفسية المُصنَّفة على وجه الخصوص. إذ يستند بناء سلم التقدير اللفظي إلى أفضل الممارسات لاستخدامه في مراجعة الأقران للمخطوط البحثي والكشف عن انتهاكات المؤلفين للمكونات الأساس ومواصفاتها والتقليل من ذاتية المراجع؛ مما يسهم في تحسين جودة المخطوط. وتتبلور الأهمية النظرية للدراسة الحالية في توفير إطار نظري يتناول سلم التقدير اللفظي من حيث خطوات البناء وثبات درجاته وصدقه. كما تتبلور الأهمية العملية لهذه الدراسة بالنقاط الآتية:

- أهمية المرحلة التي يتناولها هذا البحث في عملية مراجعة الأقران التي تتوسط بين المحرر والمراجع، تلك المرحلة التي يتم فيها اتخاذ قرارات مهمة حول قبول المخطوط أو رفضه والحكم على جودته.
- توفير أداة لتقييم المخطوط تتضمن محكات تُشكل في الوقت ذاته مكونات أساس للبحوث يتفق عليها الخبراء والباحثون وكذلك الدراسات المتخصصة في حقل العلوم التربوية والنفسية.
- لفت نظر المؤلفين والمراجعين والمحررين إلى أهمية سلم التقدير اللفظي في الكشف عن الانتهاكات التي قد يتضمنها المخطوط - في المكونات الأساس - التي قد لا تتمكن النماذج المستخدمة حاليًا من اكتشافها.
- رفع مستوى جودة المخطوط المقدم للنشر؛ وهذا يتطلب من المحرر والمراجع والمؤلف تبني سلم تقدير لفظي يُسهّل من إجراءات التقييم ويزيد في موضوعيته ودقته، ويتمتع في الوقت ذاته بدلالات صدق وثبات جيدة.

### أهداف الدراسة

1. بناء سلم تقدير لفظي "روبرك" لمراجعة الأقران للمخطوط في العلوم التربوية والنفسية استنادًا إلى نموذج هيلفورت (2010) Helvoort الذي يمثل الصورة المختصرة لخطوات بناء "الروبرك" التحليلي والكمي التي وضعها ميرتلر (2001) Mertler.
2. تعرّف دلالات صدق وثبات "الروبرك" لمراجعة الأقران للمخطوط في العلوم التربوية والنفسية.

### مصطلحات الدراسة وتعريفاتها الإجرائية

1. سلم التقدير اللفظي "روبرك": يُقصد به "دليل التصحيح المستخدم لتقييم جودة الاستجابات التي يتم إنشاؤها من قبل الطلاب" (Popham, 1997). ويعرفه الباحث إجرائيًا بأنه مصفوفة ثنائية البعد صُممت لتقييم أداء المؤلف (الباحث) لمهمة أدائية (Performance task) لا يمكن تصحيحها آليًا، وهذه المهمة هي كتابة المخطوط البحثي المقدم للنشر في مجالات العلوم التربوية والنفسية. بحيث يتضمن العمود الأيمن من المصفوفة المحكات المراد تقييم المخطوط في ضوءها، ويوفر الصف العلوي تدريبًا لفظيًا/عدديًا لمستويات الأداء، ويقدم تقاطعها وصفًا لمستويات الأداء المتوقعة على كل محك أو عناصره، ويتم تقدير كل مستوى على نحو منفصل ثم تجمع نتائج المستويات لتكوين نتيجة كلية للأداء.
2. المخطوط البحثي: يعرفه قاموس أكسفورد بأنه نسخة من كتاب، قطعة موسيقية وما إلى ذلك قبل طباعتها (Manuscript, 2019). ويعرفه الباحث إجرائيًا، بأنه الورقة البحثية/المقال التي أعدها مؤلف (باحث) أو مجموعة من المؤلفين ولم يتم نشره من قبل، وتم طباعته من

قبلهم وحسب تعليمات المجلة أو الدورية التي يتوقع أن يتم نشر البحث فيها. ويُشكّل الورقة الأولى للمقالة العلمية أو الأكاديمية قبل مراجعتها من قبل المحرر أو المراجعين، وغالبًا ما يتم تقديم المخطوط إلكترونيًا إلى المجلات أو الدوريات من أجل إخضاعه للمراجعة (التحكيم) ثم إصدار قرار حول قبوله أو رفضه أو إعادة تقييمه مرة أخرى.

## الطريقة والإجراءات

في ما يلي وصف لمنهجية الدراسة، والإجراءات التي تم استخدامها لبناء أدوات الدراسة "الروبك" وطرق التحقق من صدقها وثباتها.

### المنهجية

توصل الباحث بعد تحديد مشكلة الدراسة والاطلاع على الأدب السابق ومراجعة العديد من المناهج البحثية إلى أنّ المنهج الملائم للدراسة الحالية هو المنهج الوصفي التحليلي Descriptive Analytics، لأنه يمد الباحث ببيانات ومعلومات تُسهّم على نحو كبير في وصف طبيعة البيانات التي يمكن جمعها باستخدام أداة الدراسة "الروبك" واتخاذ القرارات في ضوءه. فالغرض الرئيس من هذا البحث هو تطوير "روبك" تحليلي يتم اختباره لتقييم المخطوط المقدم للنشر. إذ تشير مراجعة الأدبيات - تمّ الإشارة إليها سابقًا - إلى أنّ سلم التقدير اللفظي هو أداة تقييم وتقدير جيدة في الأوضاع التعليمية المختلفة على نحو عام. ولقد تم تقديم "الروبك" التحليلي على "الروبك" الكلي في هذا البحث لأنه يوفر تغذية راجعة مركزة للمراجع والمحرر وكذلك للمؤلف. وحيث إنّ الهدف الرئيس لهذا البحث هو بناء "روبك" لمراجعة الأقران للمخطوط فقد تم استعراض ذلك بمزيد من التفاصيل في أداة الدراسة.

### أداة الدراسة

بما أنّ هدف الدراسة الرئيس هو بناء سلم تقدير لفظي "روبك" لتقييم المخطوط البحثي في العلوم التربوية والنفسية، فقد تناول الباحث في هذا الجزء بناء هذا "الروبك" وفقًا للخطوات المتبعة لبناء مثل هذا النوع من أدوات التقدير. وحيث إنّ "الروبك" الذي تم تصميمه في هذه الدراسة من المفترض أن يكون قابلاً للتطبيق في مراجعة المخطوط، وبما أنّ نموذج "ميرتلر" (Mertler 2001) كان مخصصًا لبناء "روبك" لتصحيح مهام تعليمية محددة. لهذه الأسباب، فقد فضّل الباحث استخدام نموذج "هيلفورت" (Helvoort 2010) الذي يمثل صورة مختصرة من نموذج "ميرتلر" لغير المهمات التعليمية.

وفقًا لإجراءات "هيلفورت"، فقد تطلبت الخطوة الأولى تحديد محكات التقييم، ثم وصف السلوك المهني Professional behaviour الذي يمثل أعلى مستوى أداء في التدريج والسلوك غير المرضية Insufficient behaviour التي تمثل أدناه، وذلك لكل محك ضمن المرحلة الأولى (مرحلة العصف الذهني) Brainstorming phase. وفي المرحلة الثانية (مرحلة التفاصيل) Elaboration phase جرى اختيار عدد من مستويات الأداء المطلوب استخدامها في "الروبك" وهي خطوة إضافية أضافها "هيلفورت" Helvoort، ثم تبعها وصف مستويات الأداء المتوسطة لكل محك. وتطلبت المرحلة الأخيرة (مرحلة الاختبار) Testing phase جمع عينات من المخطوط البحثي التي تغطي مستويات الأداء المختلفة، ونظرًا إلى أن سلم التقدير اللفظي في الدراسة الحالية لم يكن متعلقًا بمهمة تعليمية متوفرة ليتم تقييمها. لذا، لم يتمكن الباحث من جمع عينات من المخطوط البحثي لتقييمها باستخدام "الروبك" المقترح؛ مما اضطر الباحث إلى تقييم الصورة الأولية لسلم التقدير من خلال تقديرات (7) سبعة محكمين جرى اختيارهم من طلبة برنامج الدكتوراه في تخصص القياس والتقويم في الجامعة الأردنية، ومثل هذا الإجراء الخطوة الأخيرة من عملية بناء "الروبك". وفي ما يلي تفاصيل إجراءات بناء سلم التقدير اللفظي "الروبك" المستخدم في هذه الدراسة.

### 1. مرحلة العصف الذهني

في هذه المرحلة استخدمت ثلاث طرق لتحديد المحكات الرئيسية وفي وصف مستويات الأداء المتطرفة "أقل من المتوقع" و"يتجاوز المتوقع" في "الروبك"، وهذه الطرق هي:

- مراجعة الدراسات التي بحثت أدوات المراجعة، ومراجعة منهجيات البحث في العلوم التربوية والنفسية.
  - تحليل مكونات البحوث في عينته من المجلات في العلوم التربوية والنفسية.
  - مراجعة المحكات مع زملاء ممن درسوا مواد البحث في العلوم التربوية وتصميم البحث Research design لمدة تجاوزت عشر سنوات.
- أما نقطة الانطلاق في تعرف المحكات فقد تمثلت في مراجعة الدراسات التي بحثت أدوات مراجعة البحوث على نحو عام (Berk, Campbell, & Pang, 2016). Hirshleifer, 2017; Bordage & Caelleigh, 2001; Raamkumar, Foo, & Pang, 2016). والنفسية (Gastel & Day, 2016 ; Kerlinger, 1986; Teddlie & Tashakkori, 2009) والاطلاع على دراسات ما وراء التحليل التي تجريها بعض المجلات مثل "مجلة النشرة النفسية" Psychological Bulletin (Albarracín et al., 2018). وقد تمخض عن هذه المراجعة (10) عشرة مكونات - الأكثر تكرارًا- تتضمنها الدراسات التي بحثت أدوات المراجعة، إضافة إلى منهجيات البحث في العلوم التربوية والنفسية، التي تحددت في:

أهمية مشكلة البحث، تحديد مشكلة البحث وصياغتها، تعريف المصطلحات المستخدمة، عرض الدراسات المرتبطة بالبحث، اختيار أو بناء الأدوات، جمع البيانات وتحليلها، مناقشة النتائج، التضمينات العملية وحدودها، اقتراح الدراسات والمشكلات التي تحتاج إلى مزيد من البحث، وأخيرًا التوثيق. ولتحليل المكونات الأساس للبحوث، فقد جرى اختيار عشر مجلات في العلوم التربوية والنفسية من أصل 59 مجلة (أي ما نسبته 66%) من محافظة الجمعية الأميركية لعلم النفس للمجلات التي تم فحص متطلباتها ومعاييرها في النشر (2019، "Journal Impact Factors") والمنشورة من قبل APA حسب تقرير استشهد المجلات Citation Journal للعام 2019 والصادر عن Clarivate Analytics والمستندة إلى بيانات 2017 التي معامل تأثيرها<sup>1</sup> Impact factor أكثر من 4. وقد جاء اختيار المجلات المنشورة من قبل APA بسبب أن 25% من هذه المجلات تقع ضمن فئة واحدة على الأقل من أفضل عشر فئات موضوعات واردة في تقارير الاستشهاد بالمجلات (Journal Journal Citation Reports subject category "Journal Journal Citation Reports subject category", 2019). ثم جرى اختيار بحث واحد -عشوائيًا- من ضمن البحوث المنشورة في الموقع الإلكتروني للمجلات العشر ضمن عينة المقالات المجانية Sample articles التي توفرها هذه المجلات، ثم رُصدت المكونات المتضمنة في كل منها، التي تم حصرها في ثلاثة عشر (13) مكونًا، وهي: الملخص، المقدمة، الطريقة، الإجراءات، المقاييس، التحليل، النتائج، التوجهات المستقبلية، التضمينات النظرية والعملية، الحدود، الاستنتاج، المراجع، الملاحق. والجدول (1) يوضح قائمة المجلات التي لها معامل تأثير أكثر من 4 للعام 2019.

الجدول 1: عينة المجلات التي لها معامل تأثير أكثر من 4 للعام 2019

Journal Title	2017 Impact Factor	Journal Citation Reports Subject Category
<a href="#">Psychological Bulletin</a>	13.25	• Psychology, Multidisciplinary
<a href="#">Psychological Review</a>	7.23	• Psychology, Multidisciplinary Psychology*
<a href="#">Psychological Methods</a>	6.485	• Psychology, Multidisciplinary
<a href="#">Journal of Personality and Social Psychology</a>	5.733	• Psychology, Social
<a href="#">American Psychologist</a>	4.856	• Psychology, Multidisciplinary
<a href="#">Journal of Applied Psychology</a>	4.643	• Psychology, Applied Management • Psychology, Multidisciplinary
<a href="#">Journal of Abnormal Psychology</a>	4.642	• Psychology, Clinical • Psychiatry
<a href="#">Journal of Consulting and Clinical Psychology</a>	4.536	• Psychology, Clinical
<a href="#">Journal of Educational Psychology</a>	4.433	• Psychology, Educational
<a href="#">Journal of Experimental Psychology: General</a>	4.107	• Psychology, Experimental

\* الفئات التي تحمل علامة النجمة (\*) موجودة في فهرس الاستشهاد للعلوم (SCIE).

جميع الفئات الأخرى موجودة في مؤشر الاستشهاد بالعلوم الاجتماعية (SSCI).

وفي البيئة العربية، تم فحص المحكات الواردة في عينة من نماذج التحكيم الخاصة بثلاث مجلات في العلوم التربوية والنفسية الصادرة عن الجامعات الأردنية (الأردنية، اليرموك، مؤتة) ومجلة واحدة صادرة عن الاتحاد العام للجامعات العربية (مجلة اتحاد الجامعات العربية). وقد تم تلخيصها بأربعة عشر (14) محكًا، وهي: توافق العنوان مع المحتوى، أصالة البحث وما يضيفه إلى المعرفة في مجاله، أهمية الموضوع ومدى حداثة، جودة الإطار النظري وارتباط الدراسات السابقة بموضوع البحث، أسئلة وفرضيات البحث، سلامة منهجية البحث ودقة إجراءات جمع البيانات، أساليب التحليل وملاءمتها لطبيعة البيانات، عرض النتائج، دقة مناقشة النتائج في ضوء الإطار النظري والدراسات السابقة، التوصيات، المراجع: كفايتها وحدائتها وصحة توثيقها، سلامة لغة البحث وأسلوب العرض ودقتها، مدى مناسبة الطباعة وخلوها من الأخطاء، درجة الالتزام بالأمانة العلمية، الالتزام بالتوثيق في المتن وقائمة المراجع حسب دليل APA.

تم فحص تقاطع المكونات الأساس -أنفة الذكر- مع المحكات الواردة في نماذج المراجعة المستخدمة في المجلات العربية في العلوم التربوية

<sup>1</sup> The 2017 Impact Factor = A/B

A = the number of times articles which published in 2015–2016 were cited in indexed journals during 2017.

B = the total number of research and review articles from the journal published in 2015–2016.



والنفسية، وتمخض عن فحص التقاطعات السابقة المحكات الأكثر تكرارًا والمتشابهة التي تمثل في الوقت ذاته المكونات الأساس التي يجب أن تتحقق في المخطوط. وأمكن جمعها في خمسة (5) مكونات (محكات) رئيسية، وهي: المشكلة، مواد البحث، المنهجية، النتائج، المناقشة والاستنتاج. أما العناصر التي تنضوي تحت هذه المحكات، فقد جاءت على شكل بنود فرعية، وهي ذاتها مكونات يُشترط توفرها لقبول المخطوط وتحويله من المحرر إلى المراجعين. وهذه العناصر تتوافق مع ما اتفق عليه الباحثون في العلوم التربوية والنفسية ومع الدراسات التي تناولت المكونات الأساس في البحوث مثل دراسة بوردرج وكيلي (2001) Bordage & Caelleigh ودراسة رامكومار وآخرين (2016) Raamkumar et al. وتضمنت هذه المحكات الخمسة ثمانية عشر (18) عنصرًا فرعيًا، هي: صياغة المشكلة، تحديد المشكلة، الدراسات السابقة، التوثيق، الأدوات والمواد والأجهزة، الإطار النظري، أسلوب المعالجة، الدراسات الاستطلاعية، التصميم، المعاينة، تطبيق الأدوات، نوع البيانات، تحليل البيانات، النتائج، المناقشة، الاستنتاج، حدود البحث، تطبيقات البحث، اقتراح بحوث مستقبلية. والجدول (2) يوضح المحكات/المكونات التي اتفقت عليها نماذج المراجعة في المجالات التي اتفق حولها الباحثون.

الجدول 2. المحكات/المكونات المتفق عليها بين الباحثين التي تُجمع عليها نماذج المراجعة

المحكات/المكونات	العناصر**
1. المشكلة	1. صياغة المشكلة 2. تحديد المشكلة
2. مواد البحث	3. الإطار النظري 4. الدراسات السابقة 5. الأدوات/المواد/الأجهزة
3. المنهجية / الطريقة	6. أسلوب المعالجة 7. الدراسة الاستطلاعية (التجريب) 8. التصميم 9. المعاينة 10. تطبيق الأدوات 11. نوع البيانات 12. تحليل البيانات
4. النتائج	13. النتائج 14. المناقشة 15. الاستنتاج
5. المناقشة والاستنتاج	16. حدود البحث 17. تطبيقات البحث 18. اقتراح بحوث مستقبلية

\*\* ترتيب العناصر ضمن المحك/المكون لا تعكس نمط محدد متفق عليه في نماذج المراجعة وبين الباحثين، ولا يعني أيضًا ضرورة أن تظهر بالترتيب ذاته في المخطوط، وقد تأتي هذه العناصر أو المحكات منفردة أو مجتمعة في متن المخطوط

أما بالنسبة لأوصاف مستويات الأداء، فقد استُمدت في البداية من خبرة الباحث في كتابة البحوث وفي مراجعة المخطوطات البحثية لباحثين آخرين، وكذلك خبرته في تدريس منهجيات البحوث، إضافة إلى التفكير في أداء "المؤلف الوهمي" على "الروبك" التقدير اللفظي. إذ زوّدت هذه الخبرة المتنوعة الباحث بالكثير من الأمثلة على السلوك التي تتراوح بين "أقل من المتوقع" والسلوك "تتجاوز المتوقع". وقد كان مراجعات الزملاء المتخصصين - ممن درسوا مواد البحث في العلوم التربوية وتصميم البحث - الأثر الأكبر في تعديل أوصاف المستويات المختلفة وفي الحكم على وضوحها، كما أنّ نتائج المراجعات مع الزملاء أدت إلى إعادة صياغة المحكات والأوصاف المهنية للسلوك "أقل من المتوقع" وكذلك أدت إلى تغييرات في وصف المحكات والسلوك المهنية التي "تتجاوز المتوقع"، ومثلت مراجعات الزملاء الإجراء الأخير من مرحلة العصف الذهني.

## 2. مرحلة التفاصيل

دُرِجَت مستويات الأداء على كل محك وعناصره من خلال ذكر المواصفات التي يجب أن تتوفر في ذلك المحك أو عناصره وفق ثلاثة مستويات أداء (أقل من المتوقع؛ ضمن المتوقع؛ يتجاوز المتوقع) تقابلها ثلاثة تقديرات عديدة (1، 2، 3). إذ يتوقع أن يفهم المؤلفون على نحو أفضل الأداء الذي "يتجاوز المتوقع" أو "أقل من المتوقع" للمهمة إذا تم تحديد مستويات الأداء، خاصةً إذا تم تضمين مواصفات Descriptors لكل مستوى، كما تتيح

مستويات الأداء للمراجع التمييز على نحو أكثر ثباتاً وموضوعية بين الأداء الذي "يتجاوز المتوقع" أو "الأقل من المتوقع"، أو بين الأداء الذي "يتجاوز المتوقع" والأداء "ضمن المتوقع" والأداء "أقل من المتوقع" عند تنفيذ المراجعة لتقييم عمل (المخطوط) المؤلف. وهكذا، تم إعداد الصورة الأولية لسلم التقدير اللفظي.

### 3. مرحلة التجريب

للتحقق من صدق المحكمين "للروبرك" المتمثل في وضوح صياغة المحكات ووضوح مستويات الأداء والوصف المتعلق بكل منهما وتدرج مستويات الأداء الخاصة بكل محك، فقد طُلب من أربعة محكمين من ذوي الاختصاص جرى اختيارهم من أربعة أقسام أكاديمية في الجامعة الأردنية التأكد من تحقيق "الروبرك" لهذه المعايير في ضوء التعريفات الإجرائية للمحكات المتضمنة في الأداة. ولوحظ من خلال تحليل ملاحظات المحكمين أن صيغ السلوك لبعض المحكات ضمن مستويات أداء محددة غير واضحة ويجب تحسينها، كذلك كان من الصعب التمييز بين عناصر بعض المحكات مثل التمييز بين تحديد مشكلة الدراسة وبين صياغتها، كما اقترح عدد من المحكمين حذف "نوع البيانات" أحد العناصر الفرعية للمنهجية وجعلها ضمن عنصر "تحليل البيانات"، واقترح المحكمون دمج بعض المحكات أو العناصر معاً لتشكل محكاً واحداً، إضافة إلى تغيير مسميات بعض هذه المحكات. بعد ذلك، فُحصت هذه المحكات مرة أخرى، ونتج عن ذلك دمج عناصر فرعية معاً لتشكل محكاً رئيساً جديداً ليرتفع بذلك عدد المحكات الأساس ليصبح ستة (6) محكات/مكونات، وهي: المقدمة، المنهجية، النتائج، المناقشة، الاستنتاجات والتوصيات، المراجع وتوثيقها. وينضوي تحتهما ستة (16) عشر عنصراً فرعياً. ولوجود صعوبة في التمييز بين أداء المؤلفين (الباحثين) عبر فئتين أو ثلاث فئات من مستويات الأداء، اقترح المحكمون جعل مستويات الأداء أربع فئات بدلاً عن ثلاث، وكان ذلك حلاً جيداً. إذ اقترح المحكمون إضافة مستوى أداء رابع بين السلوك "أقل من المتوقع" وبين السلوك "ضمن المتوقع" وهو المستوى "يحتاج إلى تحسين" فأصبح بذلك عدد مستويات الأداء أربعة مستويات؛ ومبرر ذلك أن زيادة عدد المستويات يجعل تلك الأحكام أكثر دقة ويقفل من احتمال حدوث خطأ في الحكم على مستوى الأداء، كما أن الأعمال الكتابية - مثل كتابة المخطوط - عادة ما تحتاج إلى مساحة أكبر من التدرج لبيان الفروق أكثر من المهام التعليمية الأخرى.

كما هدفت هذه المرحلة إلى تقييم ثبات المقيمين Interrater reliability في تقدير الدرجات المتحققة على المخطوط عند مراجعته من قبل مقيمين اثنين، ويُقاس ذلك بدرجة الإجماع بين مقيمين اثنين لكل محك في حال راجع كلاهما المخطوط ذاته باستخدام "الروبرك". ولصعوبة الحصول على مخطوط قبل نشره أو صعوبة الطلب من باحثين كتابة مخطوط جديد من أجل تقييمه باستخدام "الروبرك"، فقد تم - عشوائياً - اختيار البحث الموسوم "تقييم واقع امتلاك وممارسة معلمي الصفوف الأولى لاستراتيجيات التقويم البديل في المدارس التابعة لمنطقة حائل" الذي أعده الشمري (Shamari, 2018) والمنشور في مجلة دراسات الصادرة عن الجامعة الأردنية في العام 2019 المجلد 45 العدد 1، وهي إحدى المجالات المعتمدة، وقد تم تدريب اثنين من المقيمين من أعضاء هيئة التدريس في الجامعة الأردنية حول كيفية استخدام "الروبرك" وتقدير درجة المخطوط البحثي باستخدامه، وبصفتي الباحث وباني الأداة (الروبرك)، فقد تم حساب ارتباط سبيرمان المُصحَّح Spearman-Brown correction بين درجات المقيمين وبين درجتي لتقدير ثبات "الروبرك" حيث بلغت قيمة الارتباط 0.86، وهي قيمة عالية، واعتبرت قيمة الارتباط مؤشراً على ثبات درجة تقدير المقيمين (Core.ecu.edu, 2020).

### تقييم أداة الدراسة

تم تقييم الصورة الأولية "الروبرك" من ثلاثة جوانب،

- مدى فائدته في مراجعة الأقران للمخطوط البحثي.

- الكفاءة في الاستخدام، ويُقصد بها الوقت الذي يحتاج إليه المقيم في تقدير الدرجات على "الروبرك".

- الثبات بين المقيمين، ويشير إلى مستوى الاتفاق بين مجموعة من المحكمين على "الروبرك" في وقت معين.

ومن أجل تقييم "الروبرك" في صورته الأولية، جرى اختيار (7) سبعة محكمين من طلبة برنامج الدكتوراه في تخصص القياس والتقويم في الجامعة الأردنية، ومبرر ذلك أن نتائج البحوث تدعو إلى المزيد من إشراك الباحثين في الدراسات العليا بعدد باحثين في بداية حياتهم في عملية مراجعة الأقران وذلك في إطار التوجيه الإشرافي (McNair, Le Phuong, Cseri, & Szekely, 2019). وتم تزويدهم بالبحث ذاته الذي تم استخدامه في الخطوة السابقة إلى جانب تزويدهم بالأداة (الروبرك)، وبعد الانتهاء من تقدير علامة للبحث، طُلب منهم الإجابة عن استبيان قصير حول فائدة سلم التقدير اللفظي وفاعليته، وقد تم تقييم سلم التقدير اللفظي من خلال خمسة أسئلة تم اقتباسها من هيلفورت Helvoort (2010) وجرى تعديلها لتناسب التقييم الحالي، وهي:

1. هل محكات سلم التقدير ذات صلة بتقييم المخطوط البحثي؟  
 نعم  لا
2. هل توجد محكات مفقودة أو غير مضمنة في سلم التقدير؟  
 نعم  لا
3. هل ساعدك سلم التقدير في التركيز على المحكات ذات الصلة بتقييم المخطوط البحثي؟  
 نعم  لا
4. ما الوقت الذي تحتاج إليه لمراجعة المخطوط باستخدام سلم التقدير؟  
 أقل من 5 دقائق  5-10 دقائق  
 10-15 دقيقة  أكثر من 15 دقيقة
5. هل أنت على استعداد لاستخدام سلم التقدير "روبك" في المستقبل من أجل مراجعة البحوث التي تعرض عليك من أجل من مراجعتها؟  
 نعم  لا

وكانت النتائج الرئيسة على النحو التالي:

- رأى جميع المحكمين أن المحكات ذات صلة بتقييم المخطوطات البحثية.
- أشار 2 من 7 محكمين إلى وجود محكات مفقودة أو غير مضمنة في "الروبك".
- أشار جميع المحكمين إلى أن "الروبك" ساعدهم في التركيز على الجوانب ذات الصلة بالمكونات/ المحكات الأساس للبحوث في العلوم التربوية والنفسية.
- تم تقسيم الوقت اللازم لمراجعة المخطوط باستخدام "الروبك" إلى أربعة خيارات. حيث حدّد محكم واحد الخيار الثاني 5-10 دقائق، و4 من المحكمين حددوا الخيار الثالث 10-15 دقيقة. وحدد 2 من المحكمين الخيار الرابع أكثر من 15 دقيقة. في المقابل، لم يحدد أي من المحكمين الخيار الأول وهو أقل من 5 دقائق، ولكن لم يتم فحص سبب ذلك على نحو أكبر؛ لأن المحكمين أكملوا الاستبيان دون ذكر أسمائهم.
- أشار جميع المحكمين إلى أنهم على استعداد لاستخدام "الروبك" في المستقبل من أجل مراجعة المخطوط البحثي الذي يُعرض عليهم من أجل مراجعته.
- كذلك حُسب ثبات المقيمين باستخدام ارتباط سيرمان المُصحح، حيث بلغت قيمة معامل الارتباط 0.96، وهذا دليل آخر على ثبات الدرجات وتفسيراتها المتحققة على "الروبك".
- وهكذا، وفي ضوء آراء المحكمين وملاحظاتهم وتقييمهم للصور الأولية، وتمتعه بدلالات صدق وثبات مقبولة تمكن من استخدامه في مراجعة المخطوط، فقد تألف سلم التقدير اللفظي "الروبك" لمراجعة الأقران للمخطوط البحثي في صورته المقترحة من ستة (6) محكات رئيسة، كما يوضحها الملحق (1). علمًا بأن الفترة الزمنية التي يستغرقها تطبيق "الروبك" تتراوح بين 5 دقائق وأكثر من 20 دقيقة بمتوسط قدره 14.2 دقيقة، ويحصل المخطوط على درجة واحدة في حال التقدير "أقل من المتوقع" وأربع درجات في حال حصوله على التقدير "يتجاوز المتوقع" على كل محك أو عناصره.
- ولحساب الدرجة الكلية على سلم التقدير اللفظي، تُجمَع الدرجات الفرعية على جميع المحكات وعناصرها. وبذلك، يتراوح مدى الدرجة الكلية بين 16 و64 للبحوث الوصفية (المسحية) والارتباطية، في حين تتراوح الدرجة الكلية للبحوث التجريبية بجميع أنواعها بين 18 و72. أما قرار التحكيم، فيكون في أربعة مستويات على النحو المبين في الجدول (4).

الجدول 4. مستويات قرار التحكيم في ضوء سلم التقدير اللفظي المقترح والدرجات المناظر لها

قرار التحكيم	التجريبية	الوصفية (المسحية) والارتباطية
المخطوط صالح للنشر دون تعديلات	60 فأكثر	54 فأكثر
المخطوط صالح للنشر بعد إجراء تعديلات طفيفة	46 – 59	41 – 53
المخطوط صالح للنشر بعد إجراء تعديلات جوهرية	32 – 45	28 – 40
المخطوط غير صالح للنشر في المجلة	أقل من 32	أقل من 28

## نتائج الدراسة ومناقشتها

انطلاقاً من الهدف الرئيس لهذه الدراسة وهو بناء أداة "روبرك" للتقييم تُستخدم في مراجعة الأقران للمخطوط البحثي، فإن إجراءات البناء وما تمخض عنها شكّلت نتائج هذه الدراسة؛ ففي المرحلة الأولى من مراحل البناء وهي مرحلة العصف الذهني، أظهرت نتائج المسح النظري لمصادر عدة (الدراسات التي بحثت أدوات مراجعة البحوث، المنهجيات في العلوم التربوية والنفسية، المكونات الأساس للبحوث في العلوم التربوية والنفسية المنشورة من قبل APA، المحكات الواردة في عينة من نماذج التحكيم لمجلات في العلوم التربوية والنفسية) إجماع هذه المصادر على العناصر الأكثر تكراراً التي يجب أن تتحقق في المخطوط، حيث بلغ عددها ثمانية عشر عنصراً فرعياً (صياغة المشكلة، تحديد المشكلة، الدراسات السابقة، الأدوات والمواد والأجهزة، الإطار النظري، أسلوب المعالجة، الدراسات الاستطلاعية، التصميم، المعاينة، تطبيق الأدوات، نوع البيانات، تحليل البيانات، النتائج، المناقشة، الاستنتاج، حدود البحث، تطبيقات البحث، اقتراح بحوث مستقبلية) تنضوي جميعها تحت خمسة محكات/ مكونات رئيسية، وهي: المشكلة، مواد البحث، المنهجية، النتائج، المناقشة والاستنتاج. وفي ضوء آراء المحكمين وملاحظاتهم وتقييم المحكمين للصور الأولية، وتمتعه بدلالات صدق وثبات مقبولة، فقد تألف سلم التقدير اللفظي "الروبرك" لمراجعة الأقران للمخطوط البحثي في صورته المقترحة من ستة محكات رئيسية، وهي: المقدمة، المنهجية، النتائج، المناقشة، الاستنتاجات والتوصيات، المراجع وتوثيقها. وينضوي تحتها ستة عشر عنصراً فرعياً.

ويذكر هنا، أنّ نتائج المسح النظري للمكونات/المحكات في البحوث في العلوم التربوية والنفسية جاءت متوافقة مع نتائج تقييم المحكمين للصور الأولية لسلم التقدير اللفظي "الروبرك"؛ الذي تمثل في الحكم على مدى فائدة "الروبرك" في مراجعة الأقران للمخطوط البحثي. إذ رأى جميع المحكمين - وعددهم سبعة محكمين - أنّ المحكات ذات صلة بتقييم المخطوط البحثي، ورأى اثنان من سبعة محكمين عدم وجود محكات مفقودة في سلم التقدير، ولكن لم يتم فحص سبب ذلك على نحو أكبر؛ لأنّ المحكمين أكملوا الاستبيان دون ذكر أسمائهم. كما أشار جميع المحكمين إلى أن سلم التقدير اللفظي ساعدهم على التركيز على الجوانب ذات الصلة بالمكونات/المحكات الأساس للبحوث في العلوم التربوية والنفسية، وهذه النتيجة تتفق مع نتائج دراسة ديكنسون وأدمز (2017) Dickinson and Adams التي أظهرت أنّ "الروبرك" أحد الحلول الممكنة لتحديد الجدارة والأهمية للتقييم، وهو وسيلة المقيمين الذين يسعون إلى نقل التقييم من مجرد وصف التقييم إلى تحديد جودته ونجاحه.

وحول إمكانية استخدام الأداة "الروبرك" في المستقبل من أجل مراجعة المخطوطات التي تُعرض عليهم، فقد أظهرت النتائج أنّ جميع المحكمين ذكروا بأنهم على استعداد لاستخدام "الروبرك" في المستقبل من أجل مراجعة المخطوطات التي تُعرض عليهم. وهذا يتفق مع ما تشير إليه الأدبيات بأنّ "الروبرك" التحليلي هو أداة جيدة يُعتمد عليها في تصنيف سلوك الباحثين في أثناء كتابة البحوث (Helvoort, 2010).

وحول الكفاءة في الاستخدام، الجانب الثاني لتقييم الصورة الأولية "الروبرك"، ويُقصد بها الوقت الذي يحتاج إليه المقيم في تقدير الدرجات على "الروبرك". فقد تم سؤال المحكمين حول الزمن المخصص لتقييم المخطوط. حيث أظهرت النتائج أنّ محكماً واحداً حدد الخيار الثاني 5-10 دقائق، و4 من المحكمين حددوا الخيار الثالث 10-15 دقيقة، وحدد 2 من المحكمين الخيار الرابع أكثر من 15 دقيقة. في المقابل، لم يحدد أي من المحكمين الخيار الأول وهو أقل من 5 دقائق ولكن لم يتم فحص سبب ذلك على نحو أكبر لأنّ المحكمين أكملوا الاستبيان دون ذكر أسمائهم.

وحول الخصائص السيكمومترية "الروبرك"، الذي يمثل الجانب الثالث لتقييم الصورة الأولية "الروبرك"، فقد أظهرت نتائج ثبات المقيمين باستخدام ارتباط سبيرمان المُصحح وهو أكثر طرق الثبات استخداماً مع "الروبرك" بحسب نتائج مراجعة الدراسات التي أجرتها بروخارت وتشن Brookhart and Chen (2015) ارتفاع قيمة معامل الارتباط لتبلغ 0.96. وهذا دليل آخر على ثبات الدرجات وتفسيراتها المتحققة على سلم التقدير اللفظي إلى جانب ثبات المقيمين الذي تم التحقق منه في مرحلة التجريب. وهذا يتوافق مع نتائج دراسة جوننج (2006) Gunning بأنّ "الروبرك" التحليلي أكثر ثباتاً من "الروبرك" الكلي. وعلى كل الأحوال فإنّ مزيداً من دلالات الثبات مطلوبة لتأكيد هذه النتائج؛ إذ سساهم النتائج التي توصلت إليها الدراسة الحالية في تطوير أدوات التقييم لضمان ثبات "الروبرك" وكفاءته، كما يبيّن تقدير ثبات المقيمين للدرجات المتحققة على "الروبرك" في تقييم البحث الذي جرى اختياره عشوائياً في هذه الدراسة ارتفاع قيمته، وقد يكون مبرر ذلك التوجيه والتدريب الذي تلقاه المقيمون قبل تقييم الصورة الأولية "الروبرك". وهذا ما تؤكدته نتائج الدراسة التي أجراها جونسون وسيفنجي (2007) Jonsson and Svingby بأنه يمكن تعزيز ثبات التصحيح لتقييم الأداء من خلال استخدام "الروبرك" خاصة إذا كان تحليلياً ومحدد الموضوع ويتضمن أمثلة وتم تدريب المقيم عليه. كما يؤكد تبرير ارتفاع قيمة ثبات المقيمين في الدراسة الحالية نتائج دراسة رضائي ولوفورن (2010) Rezaei and Lovorn التي أظهرت أنّ استخدام "الروبرك" قد لا يحسن من الثبات أو الصدق إذا لم يكن المقيمون مدربين جيداً على كيفية تصميمه وتوظيفه بفاعلية.

كما أظهرت نتائج مرحلة التجريب وهي المرحلة الثالثة من مراحل بناء "الروبرك" أنّ إجراءات التجريب وفرت له دلالة صدق المحكمين التي تمثلت في أحكام أربعة محكمين من ذوي الاختصاص حول تحقيق "الروبرك" لعدة معايير (وضوح صياغة المحكات، وضوح مستويات الأداء والوصف المتعلق بكل منهما، تدريب مستويات الأداء الخاصة بكل محك)؛ إذ ساهم المحكمون في توضيح وتحسين صيغ السلوك لبعض المحكات ضمن مستويات أداء محددة، وفي حذف بعض العناصر الفرعية، كما تم في محك المنهجية، إضافةً إلى دمج بعض العناصر الفرعية معاً لتشكل محكاً جديداً، واقتراحاً

جعل مستويات الأداء أربع فئات بدلاً من ثلاث فئات بعد إضافة مستوى أداء رابع بين السلوك "أقل من المتوقع" وبين السلوك "ضمن المتوقع" وهو المستوى "يحتاج إلى تحسين" فأصبح بذلك عدد مستويات الأداء أربعة مستويات؛ مما جعل الأحكام أكثر دقة. وهذه الإجراءات تختلف من حيث النوع مع أكثر طرق الصدق استخدامًا مع "الروبك"، وهو صدق البناء (المفهوم) بحسب نتائج مراجعة الدراسات التي أجرتها بروخارت وتشن (Brookhart and Chen, 2015)؛ ولعل مبرر ذلك أن صعوبة توفير دالة صدق المفهوم "للروبك" الذي يتطلب وجود عينات من المخطوط التي تغطي جميع مستويات الأداء، إضافة إلى عدد كبير من المقيمين.

#### حدود الدراسة

محور هذه الدراسة هو بناء سلم تقدير لفظي "روبك" يُستخدم في مراجعة المخطوط البحثي، ويمكن أن يكون هذا "الروبك" نقطة انطلاق جيدة لتصميم أدوات المراجعة. ومع ذلك، فإن وجوده لا يكفي لتحسين المخطوط بدون توفر سياسات مراجعة تشجع استخدامه. من أجل ذلك لا يزال هناك الكثير من العمل الذي يتعين القيام به في تقييم المخطوط البحثي وربما إعادة تصميم نماذج المراجعة المستخدمة حاليًا. وكما هو الحال مع غالبية الدراسات، فإن الدراسة الحالية تخضع لقيود عدة، منها الاقتصار على تقييم المحكات (المكونات الأساس) الخاصة بالمخطوطات التجريبية والوصفية (المسحية) والارتباطية في العلوم التربوية والنفسية. ولكون سياسة النشر للعديد من المجالات تُلزم الباحثين بعدد محدد من الكلمات وبما لا يزيد عن 2500 كلمة كما الحال في دورية المراجعة النفسية، وفي حال الزيادة يُطلب منهم تبرير ذلك (Psychological Review, 2019). كما أن إلزام الباحثين بعدد معين من الصفحات يحد من كتابة تفاصيل طريقة أو أسلوب المعالجة في المخطوط؛ وهذا كله يعني عدم استطاعة المؤلف عرض ورقته عرضًا مفصلاً، مما لا يسمح بوصول المخطوط على بعض المحكات إلى مستوى الأداء الرابع "يتجاوز المتوقع". إلا أن المقترح أن يكون أقل تقدير مطلوب لأي مخطوط في جميع الحالات هو التقدير "ضمن المتوقع" ويقابله التدرج العددي (3) لكل محك أو عناصره.

ونظرًا إلى أن بعض المحكات المضمنة في سلم التقدير اللفظي تناسب أنواعًا معينة من البحوث مثل البحوث التجريبية، فإنه يجب تقويم هذه البحوث في ضوء المكونات التي تنطبق عليها فقط. ولتيسير ذلك على المراجع فقد تم وضع إشارة في حاشية سلم التقدير اللفظي تشير إلى عناصر المحكات التي تناسب جميع أنواع البحوث التجريبية والوصفية (المسحية) والارتباطية، وكذلك العناصر التي تناسب البحوث التجريبية على وجه الخصوص. من جانب آخر، اقتصر سلم التقدير على صدق المحكمين وثبات المقيمين، وسيكون من المفيد التحقق من صدقه وثباته بطرق أخرى في دراسات مستقبلية. كما تجدر الإشارة إلى أن عدد المحكمين المشاركين في حساب صدقه كان قليلًا نوعًا ما؛ ويحول هذا دون أي تعميم لنتائج تقييم المحكمين في العلوم التربوية والنفسية على الحقل المعرفية الأخرى، ويمكن تجاوز ذلك في الدراسات المستقبلية بزيادة عدد المحكمين وتنوع تخصصاتهم. كما اقتصر تقييم الأداة "الروبك" في صورته الأولية على تجارب الباحثين المشاركين من طلبة الدراسات العليا في برنامج الدكتوراه على ثلاثة أبعاد (مناسبتها للمراجعة، الكفاءة في الاستخدام، وثبات المقيمين). في المقابل سيكون من المفيد تقييم سلم التقدير من قبل المؤلفين والمراجعين والمحررين في دراسة لاحقة. وأخيرًا، تجدر الإشارة إلى أن البحث الذي تم تقييمه باستخدام سلم التقدير اللفظي اقتصر على البحث الذي جرى اختياره عشوائيًا، الأمر الذي يحول دون أي تعميم لنتائج التقييم على المجالات المعرفية الأخرى.

#### التضمينات العملية للبحث

يشير مصطلح المستفيدين beneficiaries أو المجموعة المستهدفة، أو المستفيدين المستهدفين إلى الذين سيستفيدون من "الروبك"، وهم الأشخاص والكيانات الذين ستتغير لديهم المعطيات بمجرد تطبيقه، ويمكن أن يتأثروا فيه على نحو مباشر أو غير مباشر. وفي ما يلي أهم المستفيدين:

##### 1. المؤلف (الباحث)

يتوقع أن يكون سلم التقدير اللفظي "الروبك" ذي أهمية للباحثين الشباب الذين بدأوا للتو في إجراء مراجعة الأقران، إذ يمكن للباحثين الاستفادة منه إضافة إلى المزيد من كبار المراجعين الذين يريدون تجديد المعلومات حول أفضل الممارسات لضمان مراجعة بناءً وأخلاقية للبحث العلمي. ويمكن أن يساعد الباحثين الجدد حديثي الخبرة في مراعاة العناصر التي يجب توفرها في أثناء الكتابة. وفي هذا الصدد يقول آرتر (Arter, 2000) ونيكوكو (Nitko, 2000) أنه عندما يتم إعلام الطلاب بسلم التقدير اللفظي قبل التعليم والتقييم، فإنهم يعرفون مستوى الأداء المتوقع، ويكونون أكثر دافعية للوصول إلى هذه المحكات. وتؤكد بروخارت (Brookhart, 2013) بأنه عندما يتم بناء سلم التقدير اللفظي واستخدامه على نحو صحيح يصبح أداة قوية تدعم التدريس وتعززه وتدعم تعلم الطلاب كذلك. وهنا يفترض الباحث أن يقوم سلم التقدير اللفظي بهذه الوظيفة عند استخدامه في تقييم المخطوط، فكما يُعزّز سلم التقدير اللفظي تعلم الطلاب، فإنه من المتوقع أيضًا أن يُعزز تعلم الباحثين لأهم المكونات الأساس التي يجب أن يتضمنها المخطوط عند

تقديمه للنشر وفق مواصفات محددة، وهذا يعزز من جعل الباحث "متعلمًا ذاتيًا"، كما يحقق مبدأ التقييم من أجل التعلم والجودة.

## 2. المراجع

يظل دور المراجع في مراجعة المخطوط وإسهامه في المجتمع العلمي والأكاديمي مهمًا (Ali & Watson, 2016). إذ يُمكن للمراجع باستخدام "الروبك" المقترح دراسة واستنتاج ما إذا كانت الورقة تُسهم إسهامًا قيمًا في المعرفة والأدب الحالي من حيث: تطوير النظرية، والبيانات الجديدة، والمنهجية الجديدة.... إلخ؛ إذ إنَّ "الروبك" المقترح يجعل المراجع قادرًا على هذا الاستنتاج وتقدير مستوى جودة المخطوط قبل الموافقة على نشره. كما يمكن أن يساعد "الروبك" المراجعين على تقليل الوقت الذي يمضونه في تصنيف المخطوط. وفي تقليل الحاجة إلى كتابة التعليقات نفسها مرارًا وتكرارًا لعدد من المخطوطات التي تتم مراجعتها. ففي دراسة حديثة أجراها جوناثان وهارفي وهيرشليفير (Jonathan, Harvey, and Hirshleifer (2017) لتبسيط الضوء على ما يجري على نحو صحيح، وما هو الخطأ في عملية المراجعة، حيث وجدوا إجماعًا تقريبيًا على أنَّ المحكمين في المجلات ذات التصنيف المرتفع في الاقتصاد يخطئون أخطاءً متماثلة يمكن تصويبها.

## 3. المحرر

يؤمّل أن يُسهم استخدام المحررين لسلم التقدير اللفظي "الروبك" بعد إجراء دراسات حول فاعليته ومقارنته في نماذج المراجعة المستخدمة حاليًا من تقليل الانتهاكات – التي تمت الإشارة إليها أنفاً- الناتجة عن استخدامها. كما أنَّ رفع "الروبك" إلى الموقع الإلكتروني للمجلة التي تسمح باستخدامه من قبل هيئة التحرير والسماح للباحثين بالاطلاع عليه وإعلامهم مسبقًا بأن تقييم المخطوط يخضع إلى هذا "الروبك"، من شأنه الحد من البحوث العلمية منخفضة الكفاءة، والتقليل من مواطن الضعف والتجاوزات غير البناءة التي تعاني منها بعض البحوث.

## التوصيات

سوف تبقى مراجعة الأقران الوسيلة الأساس التي تعتمدها المجلات المُصنّفة وغير المُصنّفة في عملية المراجعة وبغض النظر عن نموذج المراجعة المستخدم، لكن هذا لا يمنع من تحسين عملية المراجعة وأدواتها باستمرار. إذ تشير نتائج الدراسة الحالية إلى تمتع سلم التقدير اللفظي "الروبك" بدلالات صدق وثبات أولية تجعلنا قادرين على استخدامه في العلوم التربوية والنفسية. وبالرغم من أن "الروبك" المقترح والمستند إلى أفضل الممارسات في مراجعة المخطوط، قد لا يكون قابلاً للتعميم على جميع الحقول الأخرى، إلا إنه من المتوقع أن يُقدم تصوراً عاماً ومهمًا ينبغي مناقشته على نطاق أوسع في المستقبل، وللمضي قدماً يجب أن تُشجع استخدام نماذج المراجعة وتطويرها وربطها بالتكنولوجيا لتعزيز عملية مراجعة الأقران وتسريعها.

يجب أن تأخذ الدراسات المستقبلية في الحسبان استخدام سلم التقدير اللفظي "الروبك" على عينات من البحوث المنشورة في العلوم التربوية والنفسية؛ وذلك لفحص قابلية تطبيقه على نحو أوسع والحصول على صدق وثبات جديدين عند استخدامه كأداة لتقييم المخطوط البحثي، وفحص إمكانية تطبيقه من قبل الباحثين الجدد أو من يحاول منهم أن يقدم مخطوطاً بحثياً لمجلة ما من أجل تقييمها أو فحصها فيما إذا كانت تتضمن انتهاكات في مكوناتها الأساس قبل نشرها؛ الأمر الذي يُمكن المؤلف من أن يضع نفسه مكان المراجع وأن يُقدر الدرجة التي يستحقها المخطوط. إضافةً إلى إجراء دراسات أخرى لاختبار استخدام قالب "الروبك" الحالي على حقول معرفية أخرى، مع الأخذ بالحسبان اختلاف محكات التقييم باختلاف الحقل.

لذا، فإن الباحث يوصي بإجراء التقدير بواسطة درجتين مختلفتين على الأقل ومناقشة النتائج قبل تحديد التقديرات النهائية، رغم أنَّ هذه ليست ممارسة شائعة في مراجعة الأقران فردية التعمية (المغلقة). أما التوصية الثانية الناتجة عن هذا البحث، فهي استخدام المحكات الستة التي تضمنها "الروبك" مع الأخذ بالحسبان اختلاف المحكات باختلاف الحقل. وأخيراً، يمكن القول بأنَّ من يبحث في جودة المحكات وليس فقط في تحقيقها فإن عليه استخدام "الروبك".

## المصادر والمراجع

الشمري، عيد. (2018). تقييم واقع امتلاك وممارسة معلمي الصفوف الأولى لاستراتيجيات التقويم البديل في المدارس التابعة لمنطقة حائل. *دراسات: العلوم التربوية*، 45(4).

## References

- Albarracín, D. (2015). *Psychological Bulletin*, 141(1), 1–5. <https://doi.org/10.1037/bul0000007>
- Albarracín, D., Cuijpers, P., Eastwick, P. W., Johnson, B. T., Roisman, G. I., Sinatra, G. M., & Verhaeghen, P. (2018). *Psychological Bulletin*, 144(3), 223–226. <https://doi.org/10.1037/bul0000147>
- Ali, P., & Watson, R. (2016). Peer review and the publication process. *Nursing Open*, 3(4), 193–202, 2016. <https://doi.org/10.1002/nop2.51>
- AlRwaili, E. (2016). Method of (SCORING RUBRICS) in Performance Assessment and Its Impact in The Achievement of First Secondary Students in Mathematics. *Dirasat: Educational Sciences*, 43(5), 1903–1914.
- Andrade, H. G. (2005). Teaching with Rubrics: The Good, the Bad, and the Ugly. *College Teaching*, 53(1), 27–31. <https://doi.org/10.3200/CTCH.53.1.27-31>
- Arter, J. (2000). Rubrics, scoring guides, and performance criteria: Classroom tools for assessing and improving student learning. In *The annual meeting of the American Educational Research Association*, New Orleans, LA.
- Baldwin, M. (2017). In referees we trust? *Phys Today*, 70(2)44. doi: 10.1063/PT.3.34 63
- Bhattacharjee, A. (2012). Social Science Research: Principles, Methods, and Practices. *Textbooks Collection*, 3. [http://scholarcommons.usf.edu/oa\\_textbooks/3](http://scholarcommons.usf.edu/oa_textbooks/3)
- Bordage, G., & Caelleigh, A.S. (2001). A tool for reviewers: Review criteria for research manuscripts. *Academic Medicine*, 76(9), 904–951. doi:10.1097/00001888-200109000-00013
- Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. Alexandria, VA: ASCD.
- Brookhart, S., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343–368. doi: 10.1080/00131911.2014.929 565
- Chubin, D., & Hackett, E. (1989). *Peerless Science: Peer Review and U. S. Science Policy*.
- Cicchetti, D. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, 14(1), 119–135. doi: 10.1017/S0140525X00065675
- Clement, L., Chauvot, J., Philipp, R., & Ambrose, R. (2003). A method for developing rubrics for research purposes. In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), *Proceedings of the 2003 joint meeting of PME and PMENA* (Vol. 2, pp. 221–227). Honolulu: CRDG, College of Education, University of Hawaii.
- Cockett, A., & Jackson, C. (2018). The use of assessment rubrics to enhance feedback in higher education: An integrative literature review. *Nurse Education Today*, 69, 8–13.
- Cox, G.C., Morrison, J., & Brathwaite, B.H. (2015). The Rubric: An Assessment Tool to Guide Students and Markers. In *The 1st International Conference on Higher Education Advances*. Valencia, Spain. <http://dx.doi.org/10.4995/HEAd 15.2015.414>
- Devine, E., & Frass, W. (2016). Peer review: A global view. Retrieved from: <http://authorservices.taylorandfrancis.com/peer-review-global-view/>
- Dickinson, P., & Adams, J. (2017). Values in evaluation—The use of rubrics. *Evaluation and Program Planning*, (65), 113–116. doi: 10.1016/j.evalprog plan.2017.07.005
- Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, 53, 69–76. <http://doi.org/10.1016/j.stueduc.2017.03.003>
- Gastel, B., & Day, R. E. (2016). *How to write and publish a scientific paper*. Santa Barbara: Greenwood.
- Graue, E & Gallego, M. (2006). The Craft of Reviewing Manuscript for Journals: Building Skill and Learning Through the Process. In *The American Educational Research Conference*, Chicago, IL, April.
- Gunning, T. G. (2006). *Assessing and correcting reading and writing difficulties*. Boston: Pearson Education Inc.
- Helvoort, J. V. (2010). A scoring rubric for performance assessment of information literacy in Dutch Higher Education.

- Journal of Information Literacy*, 4(1), 22-39. <http://dx.doi.org/10.11645/4.1.1256>
- Horbach, S., & Halffman, W. W. (2018). The changing forms and expectations of peer review. *Research Integrity and Peer Review*, 8(3). doi:10.1186/s41073-018-0051-5
- Jonathan, B., Harvey, C., & Hirshleifer, D. (2017). How to Write an Effective Referee Report and Improve the Scientific Review Process. *Journal of Economic Perspectives*, 31 (1): 231-44. doi: 10.1257/jep.31.1.231
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Katarina, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-136. doi.org/10.14689/ejer.2014.57.2
- Kerlinger, F. N. (1986). *Foundations of Behavioral Research*. New York: Holt, Rinehart and Winston.
- Laming, D. (1991). Why is the reliability of peer review so low? *Behavioral and Brain Sciences*, 14(1), 154-156. doi:10.1017/S0140525X00065882
- Leung, D., Law, R., Kucukusta, D., & Guillet, B. (2014). How to review journal manuscripts: a lesson learnt from the world's excellent reviewers. *Tourism Management Perspectives*, 10, 46–56. <https://doi.org/10.1016/j.tmp.2014.01.003>
- Lönngren, J., Adawi, T., & Svanström, M. (2019). Scaffolding strategies in a rubric-based intervention to promote engineering students' ability to address wicked problems. *European Journal of Engineering Education*, 44(1-2), 196-221.
- Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2008). Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, 63(3), 160-168. <http://dx.doi.org/10.1037/0003-066X.63.3.160>
- Martens, K. (2018). Rubrics in program evaluation. *Evaluation Journal of Australasia*, 18 (1), 21-44. <https://doi.org/10.1177/1035719X17753961>
- McKercher, B., Law, R., Weber, K., Song, H., & Hsu, C. (2007). Why Referees Reject Manuscripts. *Journal of Hospitality & Tourism Research*, 31(4), 455–470. <https://doi.org/10.1177/1096348007302355>
- McNair, R., Le Phuong, H. A., Cseri, L., & Szekely, G. (2019). Peer review of manuscripts: A valuable yet neglected educational tool for early-career researchers. *Education Research International*. <http://dx.doi.org/10.1155/2019/1359362>
- Menéndez-Varela, J.L., & Gregori-Giralt, E. (2018). The reliability and sources of error of using rubrics-based assessment for student projects. *Assessment & Evaluation in Higher Education*, 43(3), 488–99. doi: 10.1080/02602938.2017.1360838
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. Practical Assessment. Research & Evaluation, 7(25). Retrieved April 23, 2019 from <http://edresearch.org/pare/getvn.asp?v=7&n=25>.
- Nguyen, V. M., Haddaway, N. R., Gutowsky, L. F., Wilson, A. D., Gallagher, A. J., Donaldson, M. R., & Cooke, S.T. (2015). How long is too long in contemporary peer review? Perspectives from authors publishing in conservation biology journals. *PLoS ONE*, 10(8), e0132557.
- Nitko, A. J. (1996). *Educational Assessment of Students*. Englewood Cliffs, NJ: Prentice-Hall.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9(0), 129-144. doi: <http://dx.doi.org/10.1016/j.edurev.2013.01.002>
- Popham, J. (1997). What's Wrong - and What's Right - with Rubrics. *Educational Leadership*, 55 (2), 72–75.
- Prins, F., de Kleijn, R., & van Tartwijk, J. (2017). Students' use of a rubric for research theses. *Assessment & Evaluation in Higher Education*, 42(1), 128-150. doi: 10.1080/02602938.2015.1085954
- Provenzale, J. M., & Stanley, R. J. (2005). A Systematic Guide to Reviewing a Manuscript. *AJR*, 185, 1–7.
- Raamkumar, A.S., Foo, S., & Pang, N. (2016). Survey on inadequate and omitted citations in manuscripts: a precursory study in identification of tasks for a literature review and manuscript writing assistive system. *Information*



- Research*, 21(4). Retrieved from <http://InformationR.net/ir/21-4/paper733.html> (Archived by WebCite® at <http://www.webcitation.org/6m5HZifOk>).
- Reddy, Y., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448, doi: 10.1080/026 02930902 86 2859
- Rennie, D. (2003). Editorial peer review: Its development and rationale. *Peer Review in Health Sciences*. 1-13.
- Rezaei, A., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15 (1), 18-39. <https://doi.org/10.1016/j.asw.2010.01.003>
- Rockwell, S. (2005). Ethics of Peer Review: A Guide for Manuscript Reviewers. Retrieved 16-7-2012 from: [http://medicine.yale.edu/therapeutic radiology/Images/Ethical\\_Issues\\_in\\_Peer\\_Review\\_tcm307-34211.pdf](http://medicine.yale.edu/therapeutic radiology/Images/Ethical_Issues_in_Peer_Review_tcm307-34211.pdf).
- Rosenfeld, R. M. (2010). How to review journal manuscripts. *Otolaryngology-Head and Neck Surgery*, 142, 472-486.
- Shamari, E. (2018). Evaluating the Status of Possession and Practice of Basic Grade's Teachers of Alternative Assessment Strategies in Schools of Hail Region. *Dirasat: Educational Sciences*, 45(7), 537-551.
- Suskie, L. (2009). *Assessing student learning: A common sense guide*. San Francisco: Jossey-Bass.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. London: Sage.
- Wicherts, J. M. (2016). Peer Review Quality and Transparency of the Peer-Review Process in Open Access and Subscription Journals. *PLOS ONE*, 11(1), e0147913. <https://doi.org/10.1371/journal.pone.0147913>
- Wood, M., Roberts, M., & Howell, B. (2004). The Reliability of Peer Reviews of Papers on Information Systems. *Journal of Information Science*, 30(1), 2–11. <http://doi.org/10.1177/0165551504041673>