الجامعة الإسلامية بالمدينة المنورة
ISLAMIC UNIVERSITY OF MADINAH

مجلة الجامعة الإسلامية

للعلوم التربوية والاجتماعية

مجلة علمية دورية محكمة

تصدر أربع مرات في العام خلال الأشهر:

(مارس، يونيو، سبتمبر، ديسمبر)

العدد 21 - المجلد 40

رمضان 1446 هـ - مارس 2025 م

معلومات الإيداع في مكتبة الملك فهد الوطنية

النسخة الورقية :

رقم الإيداع: 1441/7131

تاريخ الإيداع: 1441/06/18

رقم ردمـــد : 1658-8509

النسخة الإلكترونية :

رقم الإيداع: 1441/7129

تاريخ الإيداع: 1441/06/18

رقم ردمـــد : 1658-8495

الموقع الإلكتروني للمجلة :

https://journals.iu.edu.sa/ESS

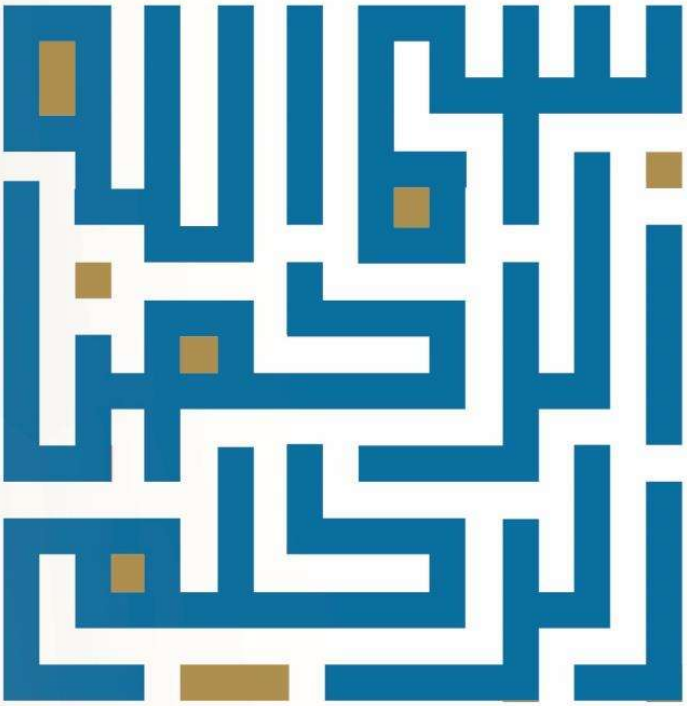
البريد الإلكتروني للمجلة :

ترسل البحوث باسم رئيس تحرير المجلة

iujourna14@iu.edu.sa

البحوث المنشورة في المجلة
تعبرعن آراء الباحثـين ولا تعبر
بالضرورة عن رأي المجلة

جميع حقوق الطبع محفوظة
للجامعة الإسلامية



قواعد وضوابط النشر في المجلة

- أن يتسم البحث بالأصالة والجدية والابتكار والإضافة المعرفية في التخصص.
- لم يسبق للباحث نشر بحثه.
- أن لا يكون مستلاً من أطروحة الدكتوراه أو الماجستير سواء بنظام الرسالة أو المشروع البحثي أو المقررات.
- أن يلتزم الباحث بالأمانة العلمية.
- أن تراعى فيه منهجية البحث العلمي وقواعده.
- أن لا تتجاوز نسبة الاقتباس في البحوث التربوية (25%)، وفي غيرها من التخصصات الاجتماعية لا تتجاوز (40%).
- أن لا يتجاوز مجموع كلمات البحث (12000) كلمة بما في ذلك الملخصين العربي والإنجليزي وقائمة المراجع.
- لا يحق للباحث إعادة نشر بحثه المقبول للنشر في المجلة إلا بعد إذن كتابي من رئيس هيئة تحرير المجلة.
- أسلوب التوثيق المعتمد في المجلة هو نظام جمعية علم النفس الأمريكية (APA) الإصدار السابع، وفي الدراسات التاريخية نظام شيكاغو.
- أن يشتمل البحث على : صفحة عنوان البحث، ومستخلص باللغتين العربية والإنجليزية، ومقدمة، وصلب البحث، وخاتمة تتضمن النتائج والتوصيات، وثبت المصادر والمراجع، والملاحق اللازمة مثل: أدوات البحث، والموافقات للتطبيق على العينات وغيرها؛ إن وجدت.
- أن يلتزم الباحث بترجمة المصادر العربية إلى اللغة الإنجليزية.
- يرسل الباحث بحثه إلى المجلة إلكترونيًا ، بصيغة (WORD) وبصيغة (PDF) ويرفق تعهدًا خطيًا بأن البحث لم يسبق نشره ، وأنه غير مقدم للنشر، ولن يقدم للنشر في جهة أخرى حتى تنتهي إجراءات تحكيمه في المجلة.
- المجلة لا تفرض رسوما للنشر.

الهيئة الاستشارية :

معالـي أ.د : محمد بن عبدالله آل ناجي
رئيس جامعة حفر الباطن سابقاً

معالي أ.د : سـعيد بن عمر آل عمر
رئيس جامعة الحدود الشمالية سابقاً

معالي د : حسـام بن عبدالوهاب زمان
رئيس هيئة تقويم التعليم والتدريب سابقاً

أ. د : سـليمان بن محمد البلوشي
عميد كلية التربية بجامعة السلطان قابوس سابقاً

أ. د : خالـد بن حامد الحازمي
أستاذ التربية الإسلامية بالجامعة الإسلامية سابقاً

أ. د : سـعيد بن فالح المغامسي
أستاذ الإدارة التربوية بالجامعة الإسلامية سابقاً

أ. د : عبداللـه بن ناصر الوليعي
أستاذ الجغرافيا بجامعة الملك سعود

أ.د. محمد بن يوسف عفيفي
أستاذ أصول التربية بالجامعة الإسلامية سابقاً

هيئة التحرير:

رئيس التحرير :

أ.د : عبدالرحمن بن علي الجهني

أستاذ أصول التربية بالجامعة الإسلامية في المدينة المنورة

مدير التحرير :

أ.د : محمد بن جزاء بجاد الحربي

أستاذ أصول التربية بالجامعة الإسلامية في المدينة المنورة

أعضاء التحرير:

معالي أ.د : راتب بن سلامة السعود

وزير التعليم العالي الأردني سابقا
وأستاذ السياسات والقيادة التربوية بالجامعة الأردنية

أ.د : محمد بن إبراهيم الدغيري

وكيل جامعة شقراء للدراسات العليا والبحث العلمي
وأستاذ الجغرافيا الاقتصادية بجامعة القصيم

أ.د : علي بن حسن الأحمدي

أستاذ المناهج وطرق التدريس بالجامعة الاسلامية في المدينة المنورة

أ.د. أحمد بن محمد النشوان

أستاذ المناهج وتطوير العلوم بجامعة الإمام محمد بن سعود الإسلامية

أ.د. صبحي بن سعيد الحارثي

أستاذ علم النفس بجامعة أم القرى

أ.د. حمدي أحمد بن عبدالعزيز أحمد

عميد كلية التعليم الإلكتروني
وأستاذ المناهج وتصميم التعليم بجامعة حمدان الذكية بدبي

أ.د. أشرف بن محمد عبد الحميد

أستاذ ورئيس قسم الصحة النفسية بجامعة الزقازيق بمصر

د : رجاء بن عتيق المعيلي الحربي

أستاذ التاريخ الحديث والمعاصر المشارك بالجامعة الاسلامية في المدينة المنورة

د. منصور بن سعد فرغل

أستاذ الإدارة التربوية المشارك بالجامعة الإسلامية في المدينة المنورة

الإخراج والتنفيذ الفني:

م. محمد بن حسن الشريف

التنسيق العلمي:

أ. محمد بن سعد الشال

سكرتارية التحرير:

أ. أحمد شفاق بن حامد

أ. علي بن صلاح المجبري

أ. أسامة بن خالد القماطي

الجامعة الإسلامية بالمدينة المنورة

ISLAMIC UNIVERSITY OF MADINAH

تقييم أعضاء هيئة التدريس للمهام الأدائية لطلبة قسم علم النفس باستخدام سلم التقدير الكلي والتحليلي

Faculty Members' Assessment of the Performance Tasks of the Psychology Department's Students Using the Holistic and Analytical Rubrics

إعداد

د. هشام بن يحيى بن علي الجبيلي

أستاذ القياس والتقويم المساعد

قسم علم النفس – كلية العلوم الاجتماعية – جامعة الإمام محمد بن سعود الإسلامية

Dr. Hesham Yahya Ali Aljubaily

Assistant Professor of Measurement and Evaluation
Department of Psychology - College of Social Sciences - Imam Mohammad Ibn Saud Islamic University

Email: hyaljubaily@imamu.edu.sa

DOI:10.36046/2162-000-021-019

د. هشام بن يحيى بن علي الجبيلي

تاريخ التقديم: ٢٠٢٤/٨/٢٥ م تاريخ القبول: ٢٠٢٤/٩/٧ م

المستخلص

تهدف الدراسة إلى معرفة تأثير استخدام نماذج سلم التقدير الكلي والتحليلي على تقييم مهام الطلاب الأدائية لطلبة قسم علم النفس ومقارنتها مع طريقة التقييم التقليدية، تم توزيع ٦٠ من أعضاء هيئة التدريس بقسم علم النفس بشكل عشوائي على ثلاث مجموعات تستخدم طرق تقييم مختلفة: (كلية، تحليلية، تقليدية)، تم تحليل درجات تقييم الأداء وتأثير نوع الجنس والمستوى الأكاديمي على تقييم الأداء. كشف اختبار كروسكال-واليس عن اختلافات كبيرة في درجات التقييم عبر المجموعات الثلاث، مع تحقيق مجموعة التقييم التحليلي أعلى رتبة، تليها المجموعة الكلية ثم المجموعة التقليدية، وقد أشارت المقارنات اللاحقة إلى وجود فروق ذات دلالة إحصائية بين المجموعة التقليدية وكل من المجموعتين الكلية والمجموعات التحليلية، كما أشارت النتائج الى عدم وجود فرق دال إحصائيًا بين المجموعتين الكلية والتحليلية. وأشار تحليل حجم التأثير إلى وجود تأثير كبير لطرق التقييم الكلية والتحليلية على تقييم أداء المهام. وأشارت نتائج الدراسة كذلك إلى عدم وجود فروق ذات دلالة إحصائية تعزى إلى نوع الجنس أو المستوى الأكاديمي لأعضاء هيئة التدريس في تقييم الأداء لكل من سلم التقدير الكلي وسلم التقدير التحليلي، مما يشير إلى عدم تحيز هذه الطرق في تقييم الأداء. وبناءً على النتائج توصي الدراسة باستخدام سلم التقدير التحليلي والكلي لتقييم المهام كونها أكثر فعالية في تقييم أداء الطلبة مقارنة بالطريقة التقليدية.

الكلمات الدالة: سلم التقدير الكلي، سلم التقدير التحليلي، التقييم.

Abstract

This study aims at determining the effect of using the holistic and analytical rubrics on assessing students' performance in the Psychology Department and to compare it with the traditional grading method. 60 faculty members in the Department of Psychology were randomly distributed into three groups that used holistic rubric, analytical rubric, and traditional assessment method, to assess a student's performance task. Assessed scores were analyzed as well as the effect of gender and academic level of the faculty members. The Kruskal-Wallis test revealed significant differences in assessment scores across the three groups with the analytical rubric group achieving the highest mean rank, followed by the holistic rubric group, and then the traditional group. Post-hoc comparisons indicated that there were statistically significant differences between the traditional group and both the holistic and analytical groups, but there was no significant statistical difference between the holistic and analytical groups. Effect size analysis indicated that there was a significant effect of the holistic and analytical rubric methods on assessing task performance. The study also resulted in no significant statistical differences in assessing task performance due to the gender or academic level of faculty members in both the holistic and the analytical rubric scales, which indicates that these methods are not biased in assessing students' performance. The study recommends using the analytical and holistic rubrics to assess tasks as it is more effective in evaluating students' performance compared to the traditional method.

Keywords: Holistic rubric scale, Analytical rubric scale, Assessment

المقدمة

يعد تقييم أداء وعمل طلاب الجامعات جزءًا أساسيًا من العملية التعليمية، وعنصرًا مهما جدا في التعليم الجامعي، كما يعدّ تقييم جودة مشاريع الطلاب وواجباتهم في التعليم العالي مسعىً حاسمًا وضروريًا لقياس تطور معارفهم ومهاراتهم. وغالبًا ما يستخدم المعلمون أساليب تقييم مختلفة، وأصبحت نماذج التقييم أداةً لا غنى عنها لتقييم عمل الطلاب بشكل منهجي. ومن تعاريف التقييم التي تركز على عملية تعلم الطلبة هو تعريف (Walvoord, 2010) الذي يعرف التقييم بأنه:" عملية منهجية لجمع المعلومات حول تعلم الطلبة تتم باستخدام مصادر الوقت والمعرفة والخبرة والموارد المتاحة من أجل المساعدة في اتخاذ قراراتٍ مؤثرة تتعلق بتعلم الطلاب" (ص. ٢) .

وهناك العديد من الطرق التي يتم استخدامها لتقييم أداء طلبة الجامعات وإحدى طرق التقييم الشائعة في العملية التعليمية والمستخدمة من قبل أعضاء هيئة التدريس بالجامعات هي استخدام طريقة سلم التقدير(Rubrics) . ويعرف سلم التقدير بأنه:" نوع من المصفوفة التي توفر مستويات متدرجة من الإنجاز أو الفهم لمجموعة من معايير أو أبعاد الجودة لنوع معين من الأداء، على سبيل المثال، ورقة أو عرض تقديمي شفهي أو استخدام مهارات العمل الجماعي" (Allen & Tanner، ٢٠٠٦، ص١٩٧).

وتُستخدم طريقة سلم التقدير بشكل شائع لتقييم واجبات الطلاب ومشروعاتهم. وتعتبر سلالم التقدير أنها مصفوفة تحوي أدلة لتصف معايير تقييم عمل الطلاب وتستخدم لتقييم مجموعة واسعة من المهام مثل الواجبات والمشاريع والمقالات والعروض التقديمية، كما أن سلم التقدير يصنف الى نوعين هما: سلم التقدير الكلي وسلم التقدير التحليلي (Selke, 2013؛ ظاظا، ٢٠٢١).

كما تقدم نماذج سلالم التقدير المصنفة إلى أنواع كلية وتحليلية مزايا ومعلومات متميزة في سياق تقييم مشاريع الطلاب الجامعيين وواجباتهم. فنماذج سلم التقدير الكلي تقدم حكمًا شاملاً وواسعًا على عمل الطالب، فعند استخدام طريقة سلم التقدير الكلي، يقوم المعلمون بتقييم المهام ككل، مع الأخذ في الاعتبار الجودة الشاملة وفعالية المنتج. وتعتمد هذه الطريقة غالبًا على التوصيفات العامة لمستويات الأداء، مما يجعلها خيارًا فعالاً لتقييم المشاريع والمهام بطريقة أكثر شمولاً، وتتميز

نماذج سلم التقدير الكلي لقدرتها على تقديم رؤية شاملة لأداء الطلاب وغالبًا ما تكون أسهل في الاستخدام عند تقييم المشاريع المعقدة بمعايير متعددة الأوجه.(Brookhart, 2013)

ومن ناحيةٍ أخرى، تقوم نماذج سلم التقدير التحليلي بتقسيم معايير التقييم إلى مكوناتٍ أو أبعاد متميزة، مما يتيح إجراء تقييم أكثر تفصيلاً لعمل الطالب. وتعد نماذج سلم التقدير التحليلي مفيدةً بشكلٍ خاص لتقييم نتائج التعلم المحددة وتزويد الطلاب بتعليقات تفصيلية حول نقاط القوة لديهم ومجالات التحسين، وغالبًا ما يتم تفضيل نماذج التقييم هذه عندما يهدف المعلمون إلى تزويد الطلاب بملاحظاتٍ أكثرَ استهدافًا وتقييمًا للمهام التي لها معايير محددة جيدًا (Stevens & Levi, 2013) .

كما يعدّ الاختيار بين سلالم التقدير الكلية والتحليلية لتقييم مشاريع طلاب الجامعات وواجباتهم قرارًا حاسمًا يجب على المعلمين اتخاذه. في حين تتميز سلالم التقدير الكلية ببساطتها وقدرتها على توفير تقييم عملي سريع لعمل الطالب، فإن سلالم التقدير التحليلية تقدم تحليلاً أكثرَ تفصيلاً وشفافية لمعايير التقييم. يأتي كل نهج مع مجموعة من المزايا والقيود الخاصة به، والتي يمكن أن تؤثر بشكل كبير على جودة التقييم والدرجات والتغذية الراجعة المقدمة للطلاب وموثوقية عملية التقييم.(Stevens & Levi, 2013)

ومن الجدير بالذكر أن كلا الطريقتين تتميزان بقدرٍ من الموضوعية وذلك لوجود معايير لتقييم المشاريع أو الواجبات بقدر أكبر من الطريقة التقليدية التي تعتمد على الحكم والتقييم الذاتي للمعلم ولا يقوم على أي معيار (Reddy & Andrade ,2010).

وعلى الرغم من قدرة سلم التقدير الكلي لتقديم التقييم المناسب، إلا أن سلم التقدير التحليلي يعتبر أكثر مصداقية وموضوعية من سلم التقدير الكلي لأنه أكثر تفصيلاً ويعتمد على وصف أدق لمعايير التقييم (Yune,et al., 2018).

وفي الوقت الذي يبحث فيه المعلمون عن أساليب مبتكرة وفعالة للتقييم، يصبح فهم الفروق الدقيقة والآثار المترتبة على استخدام نماذج سلم التقدير الكلية والتحليلية في سياق المشاريع الكلية وواجباتها أمرًا بالغَ الأهمية. وهذا سيوفر رؤىً قيمة للمعلمين وتعزيز تعلم الطلاب وإنجازاتهم، وبالتالي تعزيز أهداف مؤسسات التعليم العالي. ولذلك يهدف هذا البحث إلى التحقق من مدى فعالية استخدام المعلمين لنماذج سلالم التقدير الكلية والتحليلية والمقارنة بينهما لتقييم الأداء على مستوى

التعليم الجامعي، وبالتالي التعرف لوجود الفرق او التشابه بين الطريقتين والذي يوضح مدى الدقة أو عدمها في عملية التقييم لأداء الطلبة والمقارنة مع طريقة التقييم التقليدية.

مشكلة الدراسة:

في مجال التعليم العالي، تقوم عملية التقييم بدور محوري في قياس وتعزيز نتائج تعلّم الطلاب. فهو يوفر رؤىً قيمة حول فهم الطلاب لمحتوى الدورة وفعالية الاستراتيجيات التربوية. وتحتاج عملية ممارسة تقييم أداء ومعرفة ومهارات الطلاب الى إلمام أعضاء هيئة التدريس في الجامعات السعودية لمهارات التقويم والتنوع في استخدام طرق التقييم والقياس المختلفة التي تركز على تعلم الطلبة وتحقق مخرجات التعلم المستهدفة.

ويعد عدم استخدام أعضاء هيئة التدريس لطرق متنوعة للتقييم والاعتماد على الاختبارات فقط من المشاكل التي لا تساعد في الحصول على تصور كامل لمستوى أداء طلبة الجامعات ومستوى المعرفة والمهارات المكتسبة خلال الدراسة في المرحلة الجامعية. وترى (عبود، ٢٠١٩، ص. ٧٩) أن حصر عملية التقييم في المرحلة الجامعية، اعتمادها على الاختبارات الفصلية والنهائية فقط تعد أحد المشاكل الرئيسية في العملية التعليمية لأنها لا تساعد طلبة الجامعات على معرفة مستواهم الفعلي إلا بعد العملية التعليمية وليس خلالها، بالإضافة أن هذه الطريقة لا تعزز مهارات التفكير والتعلم الذاتي وحل المشكلات التي تعتبر متطلبات لخريج الجامعة في العصر الحديث .

ومن مشكلات عملية التقييم في الجامعات أيضًا أنه عند استخدام طرق التقييم لما يقدمه الطلبة من أعمال مثل المشاريع والواجبات، لا يتم تقييمها بشكل دقيق وموضوعي، بل يتم تقييمها بالطريقة التقليدية المعتادة المعتمدة على التقييم الذاتي بدون استخدام معايير ومحكات واضحة للتقييم مثل وضع مستويات الأداء المتوقعة ووصف الأداء لكل مستوى بدقة والدرجة المستحقة لكل مستوى. ويعد استخدام طريقة سلم التقدير أو مصفوفة التقييم المتدرج أحد الطرق التي يمكنها الحد من عيوب طريقة التقييم التقليدية، لأنها توفر إطارًا منظمًا للمعلمين لتقييم المهام وبالذات التي تحتوي على عناصر وأجزاء متعددة مثل المشاريع والواجبات والأداء العملي. كما أن هذه الطريقة تعد مفيدة للمعلم والمتعلم، حيث تجعل المعلم يقيم بشكل موضوعي، وتمكن من معرفة نواحي القوة والضعف لدى الطالب وبالتالي يمكن للطالب التعرف على هذه النواحي، وتساعد أيضا على معرفة التقدم

الحاصل بمستوى الطالب ومساره التعليمي وتحصيله الدراسي (عبود، ٢٠١٩، ص. ٨١). وأشارت العديد من الدراسات الى فعالية طريقة سلم التقدير (مصفوفة التقدير) الإيجابية في تعزيز الأداء الأكاديمي، وتحسن التقييم الذاتي، والمشاركة بالمقرر ومستوى الرضى عن المقرر، وتقدم الإنجاز، وفهم الطلاب لما يجب القيام به بوضوح وكيف سيتم تقييم الأداء، وتحسين عملية التدريس، والمساهمة في التقييم بالشكل السليم، ومصدر مهم للمعلومات لتحسين البرنامج الأكاديمية (Panadero, et al.,2023؛ García-Ros, et al.,2021؛ Tshering, & Phu-ampai,2022؛ Wolf, & Stevens, 2007).

ومن أكثر طرق سلم التقدير (مصفوفة التقدير) شيوعاً وانتشاراً، طريقة سلم التقدير الكلي وسلم التقدير التحليلي. حيث توفر نماذج سلم التقدير الكلي حكمًا شاملاً على عمل الطالب، وتقدم تقييمًا عاما لمستوى الأداء للمهمة وفقا لمستويات تقدير متعددة. من جهةٍ أخرى، تقوم نماذج سلم التقدير التحليلي بتقسيم التقييم إلى معايير محددة، مما يوفر تقييمًا أكثرَ تفصيلاً للجوانب المختلفة لأداء الطالب. ولذلك قامت العديد من الدراسات بمقارنة فعالية وفائدة سلم التقدير بنوعيه؛ الشاملة والتحليلية لتقييم الطلاب وتقديم مزايا وعيوب كل طريقة. وقد وجدت إحدى الدراسات أن نماذج سلم التقدير التحليلي كانت أكثر موثوقية من نماذج سلم التقدير الكلية في التحقق من المحتوى الرئيسي، وتقديم تغذية راجعة محددة وفقًا لعدة معايير أو محكات (Goulden,1994).ووجدت دراسة أخرى أن نماذج سلم التقدير الكلية كانت أكثر فائدةً في الحالات التي لا يحتاج أداؤها وقتًا طويلاً للمهام القصيرة، أو المهام التي لا تحتاج إلى تغذيةٍ راجعةٍ محددةٍ مثل الاختبارات المقننة، في حين كانت نماذج سلم التقدير التحليلي أكثر فعاليةً في تقليل أخطاء المقيّمين لقياس المهام، لأنها تعتمد على محتوىً مفصلاً للأداء بدلاً من الاعتماد على التقييم الكلي العام (Yune, et al., 2018). ووصفت العديد من الدراسات مزايا وعيوب كلا النوعين من نماذج سلم التقدير، وبشكل عام تتميز نماذج سلم التقدير الكلية بأنها أسهل وأسرع في الاستخدام وتعطي تصوراً عاماً للإنجاز، في حين قدمت نماذج سلم التقدير التحليلي محتوىً أكثرَ تفصيلاً وتفيد بتقديم تغذية راجعة لمستوى للمعلم والمتعلم وبالتالي تقديم تصور حول نقاط القوة والضعف في أداء المهام المطلوبة (Yune, et al., 2018؛ Barathrum& Patel, 2017؛Mertler,2001). ومما سبق، يتضح اختلاف فعالية نماذج سلم التقدير الكلية، مقابل سلم التقدير التحليلية لتقييم الطلاب،

وتستخدم وفقًا للسياق المحدد والغرض من التقييم. لذلك يجب أن يعتمد الاختيار بين هذين النوعين على احتياجات التقييم المحددة للسياق البحثي أو التعليمي (Chukwu ere, 2021).

ومع ذلك لم يتم إجراء، سوى القليل من الأبحاث وبالذات في الجامعات العربية والسعودية على معلمين الجامعات والذين عليهم المسؤولية في تقييم أداء الطلبة للمشاريع والمهام الجامعية. ولذلك يهدف هذا البحث إلى تصميم نماذج سلم التقدير الكلي وسلم التقدير التحليلي واستخدامهما في عملية تقييم أعضاء هيئة التدريس لمهام الطلبة، والمقارنة بينهما وبين طريقة التصحيح التقليدية في تقييم المهام الأكاديمية التي يؤديها الطلبة، من خلال الكشف عن الفرق بين الطرق في تقييم مهمة أداء محددة على عينة من أعضاء هيئة التدريس بقسم علم النفس بجامعة الإمام محمد بن سعود الإسلامية.

وبناءً على ما سبق، يُمكن تحديد مشكلة الدراسة في التساؤل الرئيسي التالي: هل يوجد فرق دال إحصائيًا بين المجموعات (المجموعة التجريبية من أعضاء هيئة التدريس اللذين يستخدموا طريقة سلم التقدير الكلي، والمجموعة التجريبية من أعضاء هيئة التدريس اللذين يستخدموا طريقة سلم التقدير التحليلي والمجموعة الضابطة التي تستخدم الطريقة التقليدية) في تقييم المهام الأدائية للطلبة؟

فروض الدراسة:

- يوجد فرق دال إحصائيًا بين المجموعة التجريبية الأولى والمجموعة التجريبية الثانية والمجموعة الضابطة في تقييم المهام الأدائية للطلبة.
- يوجد فرق دال إحصائيًا بين أعضاء هيئة التدريس اللذين يستخدموا طريقة سلم التقدير الكلي في تقييم درجة الأداء تعزى لنوع الجنس والدرجة العلمية.
- يوجد فرق دال إحصائيًا بين أعضاء هيئة التدريس اللذين يستخدموا طريقة سلم التقدير التحليلي في تقييم درجة الأداء تعزى لنوع الجنس والدرجة العلمية.

أهمية الدراسة:

يعد التقييم عنصرًا حاسمًا في العملية التعليمية ووسيلةً أساسيةً لقياس فهم الطلاب ومهاراتهم وأدائهم الأكاديمي العام. ولذلك تكمن أهمية الدراسة بما يلي:

أ– الأهمية النظرية:

تساهم هذه الدراسة في التعرف على أبرز طرق التقييم ودورها في تعزيز تعلم الطلاب، وإبرازها كطرق فعالة تتفوق على طريقة التقييم التقليدية، التي تعتمد في التقييم على درجةٍ واحدة، وعلى الانطباع العام عن العمل. ولذلك تبرز الدراسة الحالية أهمية التقييم القائم على تقسيم التقييم إلى معايير فرعية محددة، مما يؤدي إلى تصحيح أعمال الطلبة ووضع درجات بطريقة فعالة وعادلة، بالإضافة إلى تأثيرها الإيجابي على تعلم الطلاب وتحفيزهم. وتمثل طرق التقييم باستخدام نماذج سلم التقدير الكلي ونماذج سلم التقدير التحليلي منهجيتين متميزتين لتقييم مشاريع الطلاب وواجباتهم لضمان تقييمات صحيحة وموثوقة؛ ولذلك يعد إجراء دراسة حول استخدام هذه الطريقتين التي يستخدمها أعضاء هيئة التدريس في تقييم مشاريع ومهام وواجبات الطلاب له أهمية نظرية كبيرة تؤدي إلى الإرشاد لأفضل الممارسات في مجال تقييم العملية التعليمية ونواتج التعلم. ولذلك تسعى الدراسة إلى المساهمة في تقديم المعرفة حول أفضل طرق الممارسات لتقييم أداء الطلاب في التعليم العالي، والتي توجه المعلمين ومصممي المناهج ومؤسسات التعليم العالي الذين يسعون إلى تحسين عملية التقييم بشكل خاص، والعملية التعليمية بشكل عام.

ب–الأهمية التطبيقية:

تقدم نتائج هذه الدراسة أهمية عملية وتطبيقية لأعضاء هيئة التدريس بالجامعات والمسؤولين عن تصميم أدوات التقييم وتنفيذها. ومن خلال مقارنة نماذج سلم التقدير الكلي ونماذج سلم التقدير التحليلي في سياق المشاريع والواجبات، يسعى هذا البحث إلى توفير الإرشادات حول التقييم باستخدام نماذج سلم التقدير، التي يتم استخدامها في مواقف محددة أو عند تقييم المشاريع أو المهام المعقدة التي تتطلب تقييمًا متعدد الأوجه، وتعد هذه الطرق للتقييم أكثر ملائمة لأهداف تعليمية محددة، أو تخصصات أو أنواع من المشاريع؛ لذلك ستساعد نتائج هذه الدراسة المعلمين على اتخاذ قرارات مستنيرة عند تصميم أدوات التقييم المصممة خصيصًا لأهدافهم التعليمية، وقياس نواتج التعلم بدقة ومصداقية وأكثر فعالية، وتمتد الآثار العملية إلى تطوير المناهج الدراسية، والتخطيط التعليمي، وتطوير مهارات أعضاء هيئة التدريس في التقييم، مما يضمن توافق ممارسات التقييم مع الأهداف التربوية والمساهمة في تحقيق نتائج تعلم مفيدة للطلاب.

أهداف الدراسة:

- تسعى الدراسة الحالية الى إجراء مقارنة منهجية بين نماذج سلم التقدير الكلي، ونماذج سلم التقدير التحليلي ومعرفة الفرق بينهما عندما يتم استخدامها من قبل أعضاء هيئة التدريس في تقييم أعمال الطلاب من مشاريع أو مهام مختلفة.
- تصميم نموذج لسلم تقدير كلي ونموذج لسلم تقدير تحليلي والتحقق من خصائصهما السيكومترية لتستخدم في تقييم الأداء.
- معرفة الاختلاف في تقييم أداء الطلبة بسبب نوع الجنس والدرجة العلمية لأعضاء هيئة التدريس باستخدام كل من سلم التقدير الكلي وسلم التقدير التحليلي.

حدود الدراسة:

١. **الحدود الموضوعية**: يقتصر موضوع الدراسة على المقارنة بين استخدام أعضاء هيئة التدريس لنموذج سلم تقدير كلي ولنموذج سلم تقدير تحليلي، ومعرفة الفرق بين الطريقتين بالإضافة لطريقة التقييم التقليدية في تقييم أداء عمل طالب في أحد المواضيع أو المبادئ العامة في مجال علم النفس، وبالتالي يمكن تصحيحها وتقييمها من أي عضو هيئة تدريس بالقسم.

٢. **الحدود المكانية**: تم تطبيق هذه الدراسة في قسم علم النفس بجامعة الإمام محمد بن سعود الإسلامية بمدينة الرياض.

٣. **الحدود الزمانية**: طبقت الدراسة الحالية خلال الفصل الدراسي الثالث من العام الدراسي ١٤٤٥هـ.

تعريف مصطلحات الدراسة:

سلم التقدير (Rubric): "هو عبارة عن مجموعة مترابطة ومتسقة من المعايير لعمل الطلاب، تتضمن وصفًا لمستويات جودة الأداء وفقًا للمعايير" (Brookhart, 2013, p. 4).

سلم التقدير الكلي:" هو وصف العمل من خلال تطبيق جميع المعايير في نفس الوقت وتمكين الحكم الشامل حول جودة العمل" (Brookhart, 2013, p. 6).

التعريف الإجرائي: تقييم لتمثيل الجودة الشاملة لعمل الطالب بدرجة أو قيمة واحدة ممثلة بمستوى واحد محدد من قبل المقيمين (أعضاء هيئة التدريس)، عن طريق استخدام سلم التقدير الكلي المصمم لتقديم تقييم شامل باعتبار عمل الطالب أداء متكامل.

سلم التقدير التحليلي:" تعرض قواعد التقييم التحليلية معايير محددة مسبقًا ومستويات أداء محددة عادةً في مصفوفة. وفي التصحيح التحليلي، توفر الأحكام على المعايير الفردية أساسًا لاشتقاق العلامات باستخدام قواعد واضحة" (Tomas et al., 2019, p. 2).

التعريف الإجرائي: تقييم أعضاء هيئة التدريس لأداء الطالب عن طريق استخدام سلم التقدير التحليلي حيث يقسم مشروع الطالب أو مهمته إلى معايير لها مستويات متدرجة موصوفة بمكونات محددة مما يسمح بتقييم الجوانب الفردية في أداء المهمة مثل المحتوى والتنظيم والوضوح. ويتم جمع درجة المستوى بكل معيار مع المعايير الأخرى لتحدد الدرجة النهائية لأداء الطالب.

الطريقة التقليدية للتقييم: تطلق عليه عبود (٢٠١٩، ص. ٨١) النظام التقليدي للتصحيح، وتعرفه بأنه:" النظام المتبع للتقويم في الجامعة والذي يتبعه أعضاء الهيئة التدريسية بشكل اعتيادي لتقويم أداء الطلبة على المهمات التي تطلب منهم دون الاستناد إلى محكات تمثل سلوك الطلبة وتعكس النشاط المطلوب أداؤه بطريقة واضحة."

الإطار النظري والدراسات السابقة:

أولاً: سلم التقدير الكلي والتحليلي:

لمعرفة سلالم التقدير الكلية والتحليلية بشكل أفضل، فإنه من المهم معرفة المعنى العام لسلم التقدير (Rubric). ويصف (Brookhart, 2013) سلم التقدير بأنه نموذج تقييم ودليل تصحيح منهجي أو أداة تقييمية تستخدم في التقييم التعليمي وتتألف من معايير ومستويات أداء محددة مسبقًا، مما يوفر إطارًا موحدًا لتقييم وتوضيح التوقعات الخاصة المطلوبة بالمهام أو المشاريع.

ومن خلال المعرفة للمفهوم العام لسلم التقدير فيمكن فهم سلم التقدير بنوعيه الكلي والتحليلي. فسلم التقدير الكلي عبارة عن أداة تقييم توفر تقييمًا عامًا لعمل الطالب، مع تعيين درجة واحدة بناءً على الانطباع العام للمعلم عن الأداء بأكمله. ويحتوي على معايير متدرجة مختلفة،

ولكن يتم فقط اختيار معيار واحد ليقدم حكم موحد وعام عن الأداء. (Stevens &Levi, 2013). أما سلم التقدير التحليلي فهو أداة تقييم منظمة تستخدم في التقييم التعليمي والتي تحلل عملية التقييم إلى مصفوفة من المعايير أو الأبعاد المحددة. على عكس سلم التقدير الكلي، توفر نماذج التقييم التحليلية مستويات متدرجة منفصلة لكل معيار محدد مسبقًا، مما يسهل إجراء تحليل مفصل ودقيق لمكونات محددة ضمن أداء الطالب، وتسمح هذه الطريقة بإجراء تقييم أكثر دقة وتركيزًا، مما يمكّن المعلمين من تقديم تعليقات مستهدفة حول الجوانب والأبعاد المختلفة للمهمة أو المشروع (Moskal, 2000).

ثانيا: الأطر النظرية للتقييم بطريقتي سلم التقدير الكلي والتحليلي:

من المؤكد أن استخدام نماذج سلم التقدير الكلي والتحليلي يمكن أن تستمد من أطر نظرية مختلفة، وتوفر الأدبيات الأكاديمية وجهات نظر نظرية تدعم تطبيقها في ممارسات التقييم. كما وتوفر هذه الأطر النظرية أساسًا لفهم الأساس المنطقي وراء استخدام نماذج التقييم الشاملة والتحليلية في ممارسات التقييم، والتي تشمل جوانب القياس ونظرية التعلم والتقييم الحقيقي واعتبارات العبء المعرفي. وفيما يلي نظرة عامة على بعض هذه الأطر النظرية:

١- التوافق مع نظرية القياس التربوي:

يجد استخدام نماذج سلم التقدير الكلية والتحليلية الدعم، في نظريات القياس التربوي، وخاصة في أعمال علماء النفس التربوي مثل إيبل وفريسبي (١٩٨٦). وتؤكد نظرية القياس التربوي على أهمية أدوات التقييم المتصفة بالصدق والثبات التي تقيس بدقة ما يعتزمون تقييمه. ويتوافق عمل كل من إيبل وفريسبي مع المبادئ الأساسية للقياس التعليمي، مع التركيز على أهمية إنشاء تقييمات موثوقة وصالحة وعادلة. وكان لمساهماتهم تأثيراً دائماً على هذا المجال، حيث أثرت على ممارسات بناء الاختبار، وفهم الموثوقية والصلاحية، وتعزيز استراتيجيات التقييم العادلة والمنصفة، وبالتالي المبادئ الأساسية للقياس التربوي يتم الاعتماد عليها في تصميم وإجراء نماذج سلم التقدير الكلية والتحليلية في مجال التقييم التعليمي (Ebel & Frisbie, 1986).

٢- التطبيق ضمن أطر التقييم التكويني:

إن استخدام كل من سلالم التقدير الكلية والتحليلية يوائم ويتناسب مع نظريات التقييم التكويني التي قدمها بلاك ويليام (١٩٩٨). وتؤكد طرق التقييم التكويني على العملية المستمرة للتغذية الراجعة والتحسين مما يسهل تعلم الطلاب. وتتماشى كل من طرق التقدير الكلية والتحليلية عند تطبيقها في إطار التقييم التكويني مع نظريات بلاك وويليام من خلال تقديم تغذية راجعة مستمرة، وتشجيع تفكير الطلاب، ودعم عملية التحسين المتكررة. فسلالم التقدير الكلية توفر رؤية شاملة للأداء، في حين توفر سلالم التقدير التحليلية تفصيلاً تفصيليًا للملاحظات المستهدفة وتنمية المهارات، مما يعزز بشكل جماعي عملية التقييم التكويني (Black &Wiliam, 1998).

٣- الدمج في نماذج التقييم الحقيقية (الواقعية):

يمكن وضع نماذج التقييم باستخدام سلالم التقدير الكلية والتحليلية ضمن نماذج التقييم الحقيقية (الواقعية)، كما دعا إليها جرانت ويجينز وجاي ماكتيغي (١٩٩٨). ويؤكد التقييم الحقيقي على التطبيق الواقعي والفعلي للمهارات والمعرفة المقاسة. ويشتهر كل من جرانت ويجينز وجاي ماكتيغي بعملهما في مجال التقييم الموثوق وتطوير تصميم المناهج الدراسية القائمة على الفهم. ويؤكد عملهم، الموضح بشكل خاص في "الفهم حسب التصميم" (١٩٩٨)، على أهمية التقييم الحقيقي في قياس نتائج التعلم ذات المغزى. وتتوافق كل من نماذج التقدير الكلية والتحليلية مع إطار التقييم الموثوق أو الحقيقي لجرانت ويجينز وجاي ماكتيغي من خلال التأكيد على تكامل المعرفة والمهارات في سياقات العالم الحقيقي والواقعي. فنماذج التقييم الكلية تقدم تصورا شاملاً وعاماً عن الأداء الواقعي، في حين توفر نماذج التقييم التحليلية تقييمًا تفصيليًا قائمًا على المعايير يعكس العناصر المحددة للمهام. وتساهم هذه الأنواع من نماذج التقييم في التقييم الهادف لقدرات الطلاب في تطبيق تعلمهم بشكل أصيل وموثوق (Wiggins& McTighe, 1998).

ثالثا: المقارنة بين سلم التقدير الكلي والتحليلي:

تختلف نماذج التقدير الكلية ونماذج التقدير التحليلية في أسلوبهما في تقييم الطلاب، نماذج التقدير الكلية تقدم انطباعًا عامًا عن العمل، أما نماذج التقدير التحليلية تقسم التقييم إلى معايير

ومعايير فرعية محددة. حيث إن كل نوع من نماذج سلم التقدير له مزاياه وعيوبه، ويعتمد اختيار طريقة التقييم على السياق المحدد وأهداف التقييم للمهام أو الواجبات أو المشاريع.

سلم التقدير الكلي:

توفر نماذج التقدير الكلية تقييمًا شاملاً للأداء العام، وتقييم المهام ككل. تشير الأبحاث التي أجراها موسكال (٢٠٠٠) إلى أن نماذج التقييم الشاملة يمكن أن تعزز كفاءة وضع الدرجات، مما يوفر منظورًا أوسع لعمل الطلاب. ومع ذلك، يرى النقاد أنها قد تفتقر إلى التحديد، ومن المحتمل أن تتجاهل الجوانب الدقيقة لأداء الطلاب (Reddy & Andrade, 2010).

المميزات: توفر نماذج سلم التقدير الكلية درجة واحدة إجمالية للمهمة، مما يجعل عملية التقدير أكثر كفاءة من حيث الوقت (Reddy & Andrade, 2010). وتسمح نماذج سلم التقدير الكلية بإجراء تقييم عالمي لأداء الطلاب، مما يوفر فهمًا واسعًا لنقاط القوة والضعف (Moskal, 2000). وقد تؤدي نماذج سلم التقدير الكلية إلى تقليل العبء المعرفي لكل من المقيمين والطلاب من خلال التركيز على الانطباعات العامة بدلاً من المعايير التفصيلية (Panadero, et al.,2016).

العيوب: قد تفتقر نماذج سلم التقدير الكلية إلى تقديم تغذية راجعة محددة حول المعايير الفردية، مما يحد من فهم ومعرفة الطلاب لمجالات محددة تحتاج إلى التحسين (Brookhart, 2013). ويمكن أن يكون تقييم سلم التقدير الكلي مبني على الأحكام الذاتية للمقيمين، مما يؤدي إلى تباين محتمل في الدرجات بين المقيمين المختلفين (Jonsson &Svingby, 2007).

سلم التقدير التحليلي:

في المقابل، تقوم نماذج سلم التقدير التحليلية بتقسيم التقييمات إلى معايير محددة، مما يسمح بإجراء تقييم أكثر تفصيلاً للمكونات الفرعية للمهام الطلابية. ويؤكد بلاك ووليام (١٩٩٨) على الدقة التي توفرها نماذج سلم التقدير التحليلية، مما يساعد المعلمين والطلاب على حد سواء في تحديد مجالات التحسين. ومع ذلك، فقد أثيرت مخاوف بشأن طبيعة التقييمات التحليلية التي تستغرق جهداً ووقتًا طويلاً (Gikandi, et al.,2011).

المميزات: تسمح نماذج سلم التقدير التحليلية بتقديم تغذية راجعة وملاحظات تفصيلية حول معايير محددة، مما يوفر فهمًا أكثر شمولاً لأداء الطلاب (Andrade & Du, 2005). وتمكن نماذج

سلم التقدير التحليلية المعلمين من تحديد المجالات التي يتفوق فيها الطلاب أو يكافحون فيها، مما يسهل جهود التحسين المستهدفة، وكذلك تعزز الموضوعية في تقييم الدرجات من خلال توفير معايير واضحة لكل مكون (Putri &Melani, 2022).

العيوب: يمكن أن يستغرق تصميم وتطبيق نماذج سلم التقدير التحليلية وقتًا طويلاً بالنسبة للمعلم ووقتا طويلاً للطالب لفهم وإنجاز عناصر المهمة ويعود ذلك نظرًا للحاجة إلى وجود معايير ومعايير فرعية متعددة التفصيل. (Panadero& Jonsson, 2013). كما أن المعايير التفصيلية في نماذج سلم التقدير التحليلية قد تطغى على الطلاب مما يؤدي إلى صعوبات في تحديد أولويات التغذية الراجعة من أجل التحسين (Taras,2005).

ومن المهم ملاحظة أن المزايا والعيوب المذكورة قد تختلف اعتمادًا على السياق وطبيعة المهمة ووضوح تصميم وطريقة التقييم. وغالبًا ما يختار المعلمون أنواع وطرق التقييم أو دمجها بناءً على الأهداف والمتطلبات المحددة للتقييمات، ولطبيعة المهام المطلوبة. وتشير الدراسات التي أجراها بانا‌ديرو وجونسون (٢٠١٣) إلى أن نماذج سلم التقدير الكلية والتحليلية يمكن أن تكون فعالة، اعتمادًا على طبيعة مهمة التقييم. كما وقد تكون نماذج سلم التقدير الكلية مفيدة للمهام التي تتطلب حكمًا شاملاً، بينما توفر نماذج سلم التقدير التحليلية تقييمًا أكثر تفصيلاً ومناسبًا للمهام التي تتطلب عمل مهام متفرعة وتفصيلية. بالإضافة إلى ذلك، تشير الأبحاث التي أجراها Brookhart and Chen (2015) إلى أن الجمع بين كلا النوعين الكلي والتحليلي قد يوفر نهجًا متوازنًا يلبي نقاط القوة لكل منهما. ومن خلال دمج الطريقة الكلية والتحليلية للتقييم، يمكن الاستفادة من زيادة الموثوقية والمصداقية والصلاحية والشمولية للتقييم، مما يمكّن المعلمين من اكتساب رؤيةً أعمق لأداء الطلاب وتعزيز التحسن المستمر (Yamanishi, et al.,2019؛Putri &Melani, 2022).

الدراسات السابقة:

أعد الغامدي (٢٠٢٢) دراسة هدفت إلى قياس التحصيل الدراسي لطلبة التربية الفنية بجامعة الباحة في المقررات العملية باستخدام مقياس تقدير متدرج (روبرك) أعده الباحث بغرض تطبيقه بالدراسة، كما سعت الدراسة الى تحديد مستوى أداء الطلبة وقياس التحصيل الدراسي وفقا لمعايير ومؤشرات مصممة ومحددة في مقياس التقدير المتدرج . وتكونت عينة الدراسة من (٥٥) طالب و

(٣٠) طالبة تم اختيارهم عشوائيا من طلاب قسم التربية الفنية بكلية التربية جامعة الباحة، وأعد الباحث مقياس تقدير متدرج مكون من خمسة معايير و٣٣ مؤشر لقياس التحصيل الدراسي، وأشار البحث الى تحقق ثبات وصدق الأداة المصممة. اعتمدت الدراسة على المنهج الوصفي الارتباطي، وتوصلت الدراسة إلى عدة نتائج منها ارتفاع مستوى التحصيل الدراسي للطلبة باستخدام مقياس التقدير المتدرج، ووجود علاقة طردية دالة إحصائيًا بين تقييم أعضاء هيئة التدريس وتقييم الطلبة الذاتي في جميع المعايير والمؤشرات ولأداة ككل. كما أظهرت النتائج وجود فروق ذات دلالة إحصائية في تقييم مقياس التقدير المتدرج للتحصيل الدراسي بين المراحل الدراسية، حيث وجد أن تقييم التحصيل الدراسي لطلبة المستوى الثالث والرابع والخامس أعلى من طلبة المستوى السادس، وتقييم التحصيل الدراسي لطلبة المستوى السابع أعلى من طلبة المستوى السادس. وكذلك نتج عدم وجود فرق دال إحصائيًا بين الطلاب والطالبات باستخدام مقياس التقدير المتدرج في التقييم الذاتي للتحصيل الدراسي.

أجرت عبود (٢٠١٩) دراسة هدفت الى معرفة فاعلية وأثر استخدام مقاييس التقدير المتدرجة (الروبرك) في تحسين التحصيل الأكاديمي، وفي تنمية الكفاية الذاتية الأكاديمية، ومهارات التقويم الذاتي لدى طالبات كلية التربية بجامعة الملك فيصل، وتكونت عينة الدراسة من ١٢٤ طالبة من طالبات السنة الثالثة بكلية التربية في جامعة الملك فيصل. وقامت الباحثة بتطوير مقياس تقدير متدرج كلي ومقياس تقدير متدرج تحليلي لمقرر بناء وتطوير المناهج المقدم للطالبات في الكلية، وبناء مقياس لقياس مهارات التقويم الذاتي، كما تم تطبيق مقياس الكفاءة الذاتية الأكاديمية، وأظهرت نتائج الدراسة تمتع المقاييس بمؤشرات صدق وثبات مناسبة .كما واعتمدت الدراسة على المنهج شبه التجريبي واشتملت على مجموعتين تجريبيتين، أحدهما تستخدم مقاييس التقدير التحليلية، والأخرى تستخدم مقاييس التقدير الكلية، وتتكون كل مجموعة منها من ٤٢ طالبة، كما يوجد مجموعة ضابطة تكونت من ٤٠ طالبة، واستخدم معها طريقة التقويم التقليدية، وأظهرت نتائج الدراسة وجود اختلاف دال إحصائيًا بين المجموعة التجريبية باستخدام الطريقة التحليلية و المجموعة التجريبية باستخدام الطريقة الكلية، وكان مستوى التحصيل الأكاديمي للطالبات في المجموعة التجريبية باستخدام الطريقة التحليلية أعلى من المجموعة التجريبية الأخرى. وكذلك وجود فروق دالة إحصائيًا في مستوى التحصيل الأكاديمي للطالبات ولصالح المجموعة التجريبية الأولى والثانية مقارنة مع

المجموعة الضابطة، وتعني هذه النتائج أن الطريقة التحليلية تتفوق على الطريقة الكلية في التقويم، وان كلا الطريقتين أفضل من الطريقة التقليدية في تقويم التحصيل الدراسي للطالبات. وأشارت النتائج أيضًا الى وجود فروق دالة إحصائيًا في مقياس الكفاءة الذاتية الأكاديمية باستخدام الروبرك بنوعيه الطريقة التحليلية والطريقة الكلية في التقويم، حيث كانت نتائج الطريقة التحليلية أعلى من الطريقة الكلية، كما أن الطريقتين كانتا أفضل مقارنةً بالطريقة التقليدية، مع عدم وجود فرق بين التقويم بالطريقة التحليلية والطريقة الكلية لمهارات التقويم الذاتي للطالبات، بينما يوجد فروق دالة إحصائيًا لصالح الطريقتين على الطريقة التقليدية.

كما وأجرى السرحاني (٢٠١٦) دراسة هدفت إلى معرفة فاعلية استخدام سلم التقدير التحليلي للتحصيل الدراسي لمادة الرياضيات على مستوى التذكر والفهم لدى طلاب الصف الأول المتوسط في منطقة الجوف. وقد تكونت عينة الدراسة من طلاب الصف الأول المتوسط في مدرسة متوسطة حكومية في مدينة سكاكا بمنطقة الجوف، موزعة على مجموعة تجريبية عددها ٢٤ طالب، ومجموعة ضابطة عددها ٢٢ طالب. وقد استخدم الباحث دليل نشاط للطالب ودليل للمعلم من تصميم الباحث، وفقا لسلم التقدير التحليلي لأداء الطلاب، وأعد اختبار تحصيلي مقنن يشمل أسئلة حول مستوى التذكر والفهم. وقد اعتمدت الدراسة على المنهج التجريبي بتطبيق اختبار قبلي وبعدي على المجموعة التجريبية والمجموعة الضابطة. وتشير نتائج الدراسة الى وجود فرق دال إحصائيًا في نتائج التحصيل بين المجموعة التجريبية والمجموعة الضابطة لصالح المجموعة التجريبية، مما يعني تحسن أداء الطلبة باستخدام سلم التقدير التحليلي في مستوى التذكر والفهم.

وفي دراسة أبو عبيد (٢٠١١) كان الهدف منها معرفة تأثير استخدام سلم التقدير المتدرج التحليلي والكلي في تقييم تحصيل الطلاب واتجاهاتهم لمادة الرياضيات. وقد طبقت الدراسة على عينة عشوائية من طلاب الصف الأول الثانوي في محافظة القويعية. وطبق فيها المنهج التجريبي، حيث اشتملت العينة على مجموعة تجريبية أولى مكونة من ٤٢ طالب، وتم التصحيح فيها بالطريقة التحليلية، ومجموعة تجريبية ثانية مكونة من ٤٣ طالب، وتم التصحيح فيها بالطريقة الكلية، ومجموعة ضابطة مكونة من ٤٣ طالب. وقد تم إعداد سلم التقدير المتدرج التحليلي وآخر كلي، وإعداد اختبار تحصيل لمادة الرياضيات، وتم استخدام مقياس الاتجاهات نحو الرياضيات. وقد نتج عن الدراسة وجود فروق دالة إحصائيًا في الاختبار التحصيلي وفي مقياس الاتجاهات نحو الرياضيات

وفقا لطرق التقييم المستخدمة. وكانت الفروق دالة إحصائيًا بين طريقة التصحيح التحليلية والتقليدية لصالح الطريقة التحليلية، وبين طريقة التصحيح الكلية والتقليدية لصالح الطريقة الكلية، كما ولا توجد فروق دالة بين طريقتي التقييم التحليلية والتقليدية في أداء واتجاهات الطلبة.

كما وقدم ظاظا (٢٠٢١) دراسة هدفت إلى بناء سلم تقدير لفظي "روبرك" لمراجعة المخطوطات البحثية في مجال العلوم التربوية والنفسية والتحقق من صدق وثبات سلم التقدير. واشتملت عينة الدراسة على سبعة محكمين لسلم التقدير اللفظي من طلبة برنامج الدكتوراه في تخصص القياس والتقويم في الجامعة الأردنية. وقام الباحث ببناء سلم التقدير اللفظي وفقا لنموذج هيلفورت المناسب للمهام غير التعليمية، وتم تحديد محكات التقييم، ومستويات الأداء المطلوبة، ووصف مستويات الأداء للمحكات. وقد استخدم المنهج الوصفي التحليلي في هذه الدراسة. وأشارت نتائج الدراسة الى تكون سلم التقدير من ستة محكات تشمل ١٦ عنصرا لمراجعة الأبحاث المنشورة في المجلات العلمية وهي: المقدمة، المنهجية، النتائج، المناقشة، الاستنتاجات والتوصيات، والمراجع وتوثيقها. كما وأظهرت النتائج أن نسب اتفاق المحكمين كانت مرتفعة من حيث شمولية سلم التقدير، ومناسبة المحكات وعناصرها المكونة لها، وفي اختصار زمن تقييم مراجعة الأبحاث، وبقيمة ثبات مرتفعة للمقيمين بلغت ٠,٩٦، وتمتع سلم التقدير بدلالة صدق المحكمين وتحقق عدة معايير في المقياس والتي تمثلت بوضوح صياغة المحكات، وضوح مستويات الأداء، والوصف لكل مستوى، ومناسبة تدريج مستويات الأداء للمحكات.

وهدفت دراسة Putri و Melani (2022) إلى المقارنة بين استخدام المعلمين لطريقتي نماذج سلم التقدير التحليلية والكلية لتقييم اختبارات الكتابة التحريرية لطلاب الصف الثامن. وطبقت الدراسة باستخدام المنهج التجريبي على إثنين من المعلمين من نفس المنطقة التعليمية في اندونيسيا، لتقييم اختبارات كتابية لأربعة وعشرون طالبا من طلبة الصف الثامن، وقسموا الى مجموعتين، حيث يقيم المعلمين المهام بالطريقة التحليلية ومرة أخرى باستخدام الطريقة الكلية. وقد تم استخدام مقياس تقييم كلي مكون من سته درجات، واستخدام مقياس باكمان وبالمر التحليلي لتقييم القدرة الكتابية والمكون من خمسة معايير وخمسة مستويات للأداء. وتم حساب ثبات المقاييس باستخدام اتفاق المقيمين والتحقق من صدق المقاييس من خلال تقديم المقيمين تفسيرهم لأسباب إعطاء الدرجات للمهام المصححة. وأظهرت النتائج وجود فرق دال إحصائيًا بين تقييم المعلمين بالطريقة التحليلية

والطريقة الكلية من حيث الصعوبة، حيث تشير النتيجة أن الطريقة التحليلية أصعب من حيث التصحيح ولها متطلبات أكثر من الطريقة الكلية في تقييم المهام. وأظهرت النتائج أيضا وجود ارتباط طردي مرتفع ودال إحصائيًا بقيمة ٠،٨٣ بين الطريقة التحليلية والطريقة الكلية في تصحيح المهام الكتابية للطلبة. وتشير الدراسة الى وجود فروق دالة إحصائية واختلافات في التقييم بين المعلمين عند استخدام الطريقة الكلية، بينما لم يكن هنالك اختلافات بينهم عند استخدام الطريقة التحليلية. وبشكل عام، أظهرت طريقة التقييم التحليلية موثوقية وثبات أعلى من طريقة التقييم الكلية وتقدم تغذية راجعة ومعلومات مفيدة ومحددة للطلاب والمعلمين حول الموضوع الذي يتم تقييمه.

قام Koizumi و In 'Nami و Fukazawa (2020) بدراسة هدفها المقارنة بين استخدام سلم التقدير التحليلي والكلي لأداء التحدث في الاختبارات الشفهية، والتحقق من خصائص القياس لكل طريقة والعلاقة بينهما، والتعرف على أوجه التشابه والاختلاف بين الطريقتين في تقييم المهام، وقد تم تطبيق الدراسة على ١١٠ طالب من ثلاث جامعات في اليابان، وكان أغلبهم في تخصصات العلوم والهندسة والطب، أما بقية التخصصات فمثلت نسبة ٣,٦٤٪ من عينة الدراسة. ومثل الطلاب الذكور ٧١,٨٢٪ من العينة. وتم استخدام ستة مهام لعب أدوار، وأربع مهام مناقشات ثنائية مصممة كجزء من التدريس في فصل اللغة الإنجليزية. وتم تصميم سلم تقدير تحليلي يتكون من المعايير التالية: النطق والقواعد والمفردات والطلاقة والتواصل التفاعلي، ويتم تقيمه من ثلاثة مستويات. وصمم أيضا سلم تقدير كلي يتضمن إنجاز المهام، والتواصل التفاعلي، والطلاقة، ويقيم على مقياس من واحد إلى ثلاثة. اشترك أربعة مقيمين لتقييم المهام باستخدام الطريقة الكلية، وثلاثة مقيمين من الأربعة لتقييم المهام باستخدام الطريقة التحليلية. وأظهرت نتائج التحليل باستخدام نموذج راش بشكل عام أن كلا الطريقتين متشابهين ويعملان بشكل فعال في عملية التقييم للمهام. ومع ذلك، فإن طريقة التقييم التحليلية كانت أفضل قليلاً من الطريقة الكلية في أربعة جوانب. أولاً، أفضل قليلاً في نتيجة ملائمة النموذج العام. ثانيًا، التمتع بدرجة أعلى قليلاً في الثبات من حيث المتقدم للاختبار وأداء المهمة. ثالثًا، أفضل من حيث التمييز بين المختبرين بشكل أكثر دقة. رابعاً، كانت نسبة خارج حدود الملائمة للمختبرين (overfit) أصغر. ومن نتائج الدراسة أيضا وجود ارتباط مرتفع ودال إحصائيًا بين التقييم بالطريقتين بلغ ٠،٨٤. وأخيرا، أشارت النتائج بأن طريقة سلم التقدير التحليلي ساهمت بشكل كبير في درجات تقييم طريقة سلم التقدير التحليلي، حيث

أظهرت نتائج تحليل الانحدار المتعدد أن ثلاثة معايير تحليلية (التواصل التفاعلي، والطلاقة، والقواعد والمفردات) تفسر حوالي ٧٥٪ من الدرجات الشاملة، وأن التواصل التفاعلي فقط ساهم في حوالي ٦٩٪ من الدرجات، و٥٪ إضافية كانت من مساهمة معيار الطلاقة.

وفي دراسة اجراها Jönsson و Balan (2018) هدفت إلى المقارنة بين تقييم المدرسين لأداء الطلبة باستخدام طريقة سلم التقدير الكلية، وطريقة سلم التقدير التحليلية، ومعرفة أي من الطريقتين أكثر اتفاقا وثباتا بين المقيمين، ومبررات التقييم للأداء للتحقق من صدق الطريقتين. وقد طبقت الدراسة في دولة السويد، وتكونت عينة الدراسة من ٢٤ معلما من مدارس مختلفة من نفس المنطقة. واستخدم المنهج التجريبي، حيث تم توزيعهم عشوائيا إلى مجموعتين. المجموعة الأولى مكونة من ١٣ معلما استخدموا نموذج تصحيح بالطريقة التحليلية في التقييم، والمجموعة الثانية مكونة من ١١ معلما استخدموا الطريقة الكلية لتقيم أداء الطلبة باستخدام نموذج تصحيح تم تصميمه لغرض الدراسة. وقام المعلمون في المجموعة الأولى بتقييم مهام أداء كتابية لأربعة طلبة بعمر ١٢ عاما وبمجموع ١٦ مهمة لنفس الطلبة على مدار فصل دراسي واحد، بالإضافة لتفسير أو مبررات لتقييمهم للأداء. وقام المعلمون في المجموعة الثانية بتقييم مهام الأداء لنفس الطلاب الأربعة في أجراء واحد بنفس الوقت، وأيضًا مع تقديم المبررات لدرجات التقييم. وقد أشارت النتائج إلى أن المعلمين المستخدمين للطريقة التحليلية أظهروا نسبة اتفاق أعلى بفارق كبير بواقع ٦٦،٢٪ من بين المعلمين المستخدمين للطريقة الكلية والذين حصلوا على ٤٥،٩٪، كما أن المجموعة المستخدمة للطريقة التحليلية حصلت على قيم أعلى بحساب كابا كوهين بلغت ٠،٦٠٢ وقيمة ارتباط رتبي أعلى بحساب رو سبيرمان بلغت ٠،٩٧٣، بينما حصلت المجموعة الثانية المستخدمة للطريقة الكلية على قيمة كابا كوهين بلغت ٠،٤٠٥، وقيمة رو سبيرمان بلغت ٠،٩٤٣. ولم تظهر النتائج وجود اختلافات كبيرة بين المعلمين في المجموعتين في كيفية تبرير الدرجات، بل وجد تشابه كبير بين المعلمين في تبرير التقييم لأداء الطلبة.

التعقيب على الدراسات السابقة:

يتضح من خلال عرض الدراسات السابقة بأن جميع الدراسات تأكد على أهمية استخدام طريقتي سلم التقدير التحليلي أو سلم التقدير الكلي في عملية تقييم مهام الطلاب المعرفية والعلمية

والعملية من واجبات أو مشاريع أو اختبارات، وأنها أساليب تقييم فعالة كونها تبتعد عن التقييم الذاتي وتعتبر من أكثر طرق التقييم موضوعية. وهدفت معظم الدراسات الى تصميم أدوات تقييم باستخدام طريقتي سلم التقدير التحليلي وسلم التقدير الكلي. إلا أن بعض الدراسات ركزت على تصميم أداة تقييم من نوع واحد فقط كدراسة الغامدي (٢٠٢٢) والسرحاني (٢٠١٦)، ودراسات أخرى قامت بتصميم أداة تقييم بالطريقتين التحليلية والكلية كدراسة عبود (٢٠١٩) وJönsson وBalan (2018). كما يلاحظ اختلاف مجالات التخصص التي طبقت بها الدراسات بسبب اختلاف تخصصات المهتمين بهذا الموضوع، وبالتالي تباينت الأهداف الخاصة للدراسات السابقة وخصائص العينات وأبعاد وجوانب المحتوى من مهارات ومتطلبات لأدوات التقييم التي تم تصميمها. فيلاحظ مثلا اختلاف عينات الدراسة، فمنها ما طبق على الطلبة من مراحل التعليم العام كدراسة السرحاني (٢٠١٦) وPutri وMelani (2022)، ومن الدراسات ما طبق على مرحلة التعليم العالي كدراسة الغامدي (٢٠٢٢) وعبود (٢٠١٩). وقد تنوعت كذلك المنهجيات المستخدمة بالدراسة، فتم استخدام المنهج الوصفي التحليلي مثل دراسة ظاظا (٢٠٢١)، واستخدام المنهج الارتباطي مثل دراسة الغامدي (٢٠٢٢)، واستخدام المنهج التجريبي كدراسة أبو عبيد (٢٠١١)، والمنهج شبه التجريبي مثل دراسة عبود (٢٠١٩). وتعد الدراسة الحالية إضافة للدراسات السابقة في مجال تقييم أداء الطلبة والتحصيل الدراسي، حيث تهتم هذه الدراسة ببناء أدوات تقييم بطريقتي سلم التقدير التحليلي وسلم التقدير الكلي والمقارنة بينهما في تقييم أعضاء هيئة التدريس من قسم علم النفس لأداء الطلبة للمهام الدراسية.

منهج الدراسة:

تم استخدام المنهج التجريبي للإجابة على الفرض الأول للدراسة الحالية وذلك لمناسبة هذا المنهج في إستخدام التحليل الإحصائي المناسب لتصميم الفرض. وتم توزيع المعلمين بشكل عشوائي إلى ثلاث مجموعات: ١- المجموعة التجريبية الأولى: استخدمت طريقة سلم التقدير الكلي للتقييم، ٢- المجموعة التجريبية الثانية: استخدمت طريقة سلم التقدير التحليلي للتقييم، ٣- المجموعة الضابطة: استخدمت الطريقة التقليدية للتصحيح، أي بدون استخدام أي نوع من سلالم التقدير. وتتيح هذه المنهجية المقارنة بين طرق التقييم والتحقق من تأثير المتغير المستقل (نوع طريقة التقييم:

الكلية والتحليلية والتقليدية) على المتغير التابع (درجة تقييم أداء الطالب) والتي يتم قياسها فقط بعد تطبيق طرق التقييم لنفس المهمة. ويستخدم هذا التصميم بشكل شائع في الأبحاث التجريبية لتقييم آثار التدخلات أو المعالجات التجريبية على النتائج في المتغير التابع (Trochim, & Donnelly, 2008). كما تم إستخدام المنهج الوصفي المقارن للإجابة على الفرض الثاني و الثالث لهذه الدراسة كونها تعتبر الطريقة المناسبة لتحديد الاختلافات وفقا لنوع الجنس والدرجة العلمية لأعضاء هيئة التدريس في تقييم درجة الأداء باستخدام سلالم التقدير.

مجتمع الدراسة:

تكون مجتمع الدراسة من أعضاء هيئة التدريس بقسم علم النفس في جامعة الإمام محمد بن سعود الإسلامية. ويبلغ عدد أعضاء هيئة التدريس من الذكور والإناث (١١٤) حسب إحصائية الموقع الرسمي للقسم في الجامعة (جامعة الإمام محمد بن سعود الإسلامية، د. ت.).

عينة الدراسة:

بعد تحديد مجتمع الدراسة، تم إعداد قائمة بجميع أعضاء هيئة التدريس في القسم وتعيين رقم لكل فرد من أفراد المجتمع. وتم اختيار ٦٠ من أعضاء هيئة التدريس بطريقة العينة العشوائية البسيطة. وتم توزيعهم عشوائيا إلى مجموعتين تجريبيتين، ومجموعة ضابطة بحيث تتكون كل مجموعة من ٢٠ عضو هيئة تدريس: ١- المجموعة الأولى: التي تستخدم سلم التقدير الكلي للتقييم، ٢- المجموعة الثانية: التي تستخدم سلم التقدير التحليلي للتقييم، ٣- المجموعة الضابطة: التي تستخدم الطريقة التقليدية للتقييم وبدون سلالم التقدير. ويوضح الجدول رقم (١) وصف عينة الدراسة. بالنظر إلى عينة الدراسة، يتضح أن هناك تفاوتات طفيفة في نسب الذكور والإناث وفي في توزيع المشاركين حسب الدرجة العلمية بين المجموعات، إلا أن هذه التباينات ليست كبيرة بحيث تؤثر على تكافؤ المجموعات، مما يعني أن المجموعات متكافئة إلى حد كبير.

جدول (١) توصيف عينة الدراسة

		المجموعة الضابطة	المجموعة التجريبية ١	المجموعة التجريبية ٢	المجموع
نوع الجنس	ذكر	٩ (٤٥٪)	١٣ (٦٥٪)	١١ (٥٥٪)	٣٣ (٥٥٪)
	أنثى	١١ (٥٥٪)	٧ (٣٥٪)	٩ (٤٥٪)	٢٧ (٤٥٪)

		المجموعة الضابطة	المجموعة التجريبية ١	المجموعة التجريبية ٢	المجموع
الدرجة العلمية	معيد	٢ (١٠٪)	١ (٥٫٠٪)	٢ (١٠٪)	٥ (٨٫٣٪)
	محاضر	٤ (٢٠٪)	٤ (٢٠٪)	٣ (١٥٪)	١١ (١٨٫٣٪)
	أستاذ مساعد	٦ (٣٠٪)	٧ (٣٥٪)	٧ (٣٥٪)	٢٠ (٣٣٫٣٪)
	أستاذ مشارك	٦ (٣٠٪)	٥ (٢٥٪)	٥ (٢٥٪)	١٦ (٢٦٫٧٪)
	أستاذ	٢ (١٠٪)	٣ (١٥٪)	٣ (١٥٪)	٨ (١٣٫٣٪)
المجموع		٢٠	٢٠	٢٠	٦٠

أدوات الدراسة:

تم إعداد وتصميم مقياس سلم تقدير كلي ومقياس سلم تقدير تحليلي بغرض تحقيق أهداف الدراسة الحالية. وبسبب اختلاف طبيعة كل مقياس كما اتضح سابقا، تم تصميم نماذج التقييم الكلية والتحليلية بعدة خطوات لضمان تقييم أداء المهمة التي سيقيمها أعضاء هيئة التدريس بشكل فعال. وفيما يلي خطوات وإجراءات إعداد وتصميم كل نوع من أنواع التقييم:

أولا: سلم التقدير الكلي:

١- تحديد أهداف التعلم: تم تحديد أهداف التعلم أو النتائج التي يهدف التقييم إلى قياسها بوضوح. والأهداف الخاصة بالمهمة هي: أن يعرّف الطالب مفهوم الثبات في القياس النفسي، وأن يعدد الطالب أنواع الثبات والطرق الإحصائية لحسابه.

٢- تحديد المعايير: تم تحديد المعايير أو الأبعاد الرئيسية التي تمثل الجوانب الأساسية لأداء الطالب المتعلقة بأهداف التعلم. وتغطي هذه المعايير جميع الجوانب ذات الصلة بالمهمة.

٣- تحديد مستويات الأداء: تم وضع أوصاف واضحة وما يشكل الأداء العام والكلي لكل مستوى من مستويات الأداء. وبالتالي كل خلية من المصفوفة تحتوي على وصف للأداء يتوافق مع مستوى الأداء المحدد. وحددت بالمستويات التالية: ممتاز (٤ درجات)، وجيد جدا (٣ درجات)، وجيد (٢ درجتين)، وضعيف (١ درجة) .

٤- المراجعة والتحسين: تم مراجعة نموذج التقييم للتأكد من الوضوح والشمول والمواءمة مع أهداف التعلم. وتم عرضها على مجموعة من الخبراء في المجال لتحسين نموذج التقييم حسب آرائهم، والتحقق من الخصائص السيكومترية للمقياس.

ثانيا: سلم التقدير التحليلي:

١. تحديد أهداف التعلم: تم تحديد أهداف التعلم أو النتائج التي يهدف التقييم إلى قياسها بوضوح. والأهداف الخاصة بالمهمة هي: أن يعرّف الطالب مفهوم الثبات في القياس النفسي، وأن يعدد الطالب أنواع الثبات والطرق الإحصائية لحسابه.

٢. تقسيم سلم التقدير إلى معايير: تم تحليل المهمة إلى الأجزاء المكونة لها والمعايير المحددة التي تضمنها الأداء في المهمة. وتتناول المعايير جوانب مختلفة من المهمة، وهي كالتالي: المحتوى المعرفي للعرض، وتحليل المعلومات المقدمة، وتنظيم العرض، ووضوح الأفكار وتوصيل المعلومات، واللغة والتنسيق العام.

٣. تحديد مستويات الأداء: حددت مستويات متعددة من الأداء لكل معيار، والتي تعكس درجات متفاوتة من الكفاءة أو الجودة. وحددت بالمستويات التالية: ممتاز (٤ درجات)، وجيد جدا (٣ درجات)، وجيد (٢ درجتين)، وضعيف (١ درجة).

٤. كتابة وصف الأداء: تم تحديد وكتابة أوصافًا واضحة ومحددة لكل مستوى من مستويات الأداء ضمن كل معيار. وتوضح هذه الصفات والخصائص أو الميزات المرتبطة بكل مستوى من مستويات الكفاءة والأداء .

٥. تنظيم المصفوفة: نظمت المعايير ومستويات الأداء في تنسيق منظم وعلى شكل جدول لمصفوفة. تم إدراج كل معيار رأسياً، مع إدراج مستويات الأداء أفقياً.

٦. المراجعة النهائية: أجريت مراجعة نهائية لسلم التقدير التحليلي للتأكد من أنها تصف بدقة الجوانب المطلوبة لأداء الطلاب وتوفر إرشادات واضحة للتقييم. كما تم أيضا عرض المقياس على عدد من الخبراء في المجال للاستفادة من آرائهم، والتحقق من الخصائص السيكومترية للمقياس.

الخصائص السيكومترية لمقياس سلم التقدير الكلي والتحليلي:

أولا: صدق المحتوى لمقياس سلم التقدير الكلي والتحليلي:

تم عمل تحكيم لمقياس سلم التقدير الكلي ومقياس سلم التقدير التحليلي ومعاييره والتوصيف لكل مستوى من مستويات الأداء من قبل سبعة محكمين من أعضاء هيئة التدريس في مجال علم النفس والقياس والتقويم. وقد بلغ متوسط نسبة اتفاق المحكمين على السلامة اللغوية للمعايير وتوصيف مستويات الأداء ٩٨,٤١ %، ونسبة اتفاق المحكمين على وضوح المعايير وتوصيفها لمستويات الأداء وفهمها كانت ٩٩,٢١ %، وبلغت نسبة اتفاق المحكمين على انتماء المعايير وتوصيفها لمستويات الأداء لعرض موضوع الثبات في القياس النفسي بنسبة ١٠٠ %. وتم الأخذ بملاحظات المحكمين كالتعديل او إعادة الصياغة أو أي إضافة لتوصيف مستويات الأداء للمعايير لكلا المقياسين.

ثانيا: ثبات مقياس سلم التقدير الكلي والتحليلي:

للتحقق من ثبات المقياسين، تم استخدام طرق مختلفة لكل مقياس ويعود ذلك لاختلاف طبيعة التكوين والبناء لكل مقياس.

١. سلم التقدير الكلي:

تم مواجهة مشكلة أساسية لقياس ثبات المحكمين أو ثبات الاتساق الداخلي لسلم التقدير الكلي تعود للتصميم في هذه الدراسة، والذي يعتمد على تقييم المحكمين لأداء واحد فقط ولطالب واحد فقط. وذلك لكون حساب ثبات المحكمين أو ثبات الاتساق الداخلي يتطلب المقارنة بين التقييمات المتعددة لعدد من المهام وهي لا تتوفر في هذه الحالة. ولحل هذه المشكلة، تم استخدام حساب متوسط النسبة المئوية لدرجة الاتفاق للمحكمين (The Average Agreement Percentage) والتي تعد بديلاً عمليًا ومناسباً للأغراض العملية عندما لا يمكن تطبيق الطرق المعروفة مثل (Intraclass Correlation Coefficient ICC) أو Cohen's Kappaأو كرونباخ ألفا وذلك بسبب القيود في العمليات الحسابية للبيانات (Cohen, R., & Swerdlick, M.,2017). وكانت نسبة متوسط النسبة المئوية لدرجة الاتفاق للمحكمين تساوي ٧٥ %، وتشير هذه النتيجة بوجود اتساق كبير

وأن الاتساق في تقييم المحكمين مرتفع، وبالتالي تشير النتيجة بأن مقياس سلم التقدير الكلي يتصف بالثبات.

٢. سلم التقدير التحليلي:

تم التحقق من ثبات سلم التقدير التحليلي عن طريق حساب ثبات المحكمين (Intraclass Correlation Coefficient ICC) وحساب معامل ثبات ألفا كرونباخ للمقياس. وكانت قيمة ثبات المحكمين (ICC) تساوي ٠,٨٩٦ . وتشير هذه النتيجة إلى ثبات مرتفع للمقياس، مما يعني أن سلم التقدير التحليلي متسق عند استخدامه من قبل العديد من المحكمين من أعضاء هيئة التدريس. وكانت قيمة ألفا كرونباخ تساوي أيضا٠,٨٩٦، وهذا يشير أيضا بأن سلم التقدير التحليلي يتصّف بثبات مرتفع. وبشكل عام تدل النتائج السابقة على وجود اتفاق قوي بين المحكمين، ووجود اتساقًا داخليًا عاليًا، مما يعني أن تقييمات المحكمين تقيس نفس الأداء المراد قياسه بشكل موثوق وبثبات مرتفع وهذا أمر بالغ الأهمية للتقييم العادل وغير المتحيز.

وبعد الانتهاء من خطوات التصميم والإعداد لأدوات الدراسة والتحقق من خصائصهما السيكومترية، تم إعداد الشكل النهائي لكل مقياس. والغرض من المقياسين هو تقييم وتحكيم أعضاء هيئة التدريس بالقسم لعمل طالب تم اختياره بشكل عشوائي ويحتوي على عرض كتابي مقدم عن موضوع (الثبات) في القياس النفسي. ويوضح الجدولين (٢ و٣) التاليين الشكل النهائي لسلمي التقدير الكلي والتحليلي واللذان تم استخدامهما على أفراد عينة الدراسة.

جدول (٢) سلم التقدير الكلي

وصف الأداء	الدرجة
يوضح العرض المكتوب فهمًا عميقاً ومتميزاً لموضوع الثبات في القياس النفسي، مع تنظيم واضح، وتحليل دقيق وعميق، وعرض متكامل وفعال وثري بالمعلومات العلمية، وجودة وكفاءة شاملة وفائقة ومتميزة للعرض.	٤ ممتاز
يُظهر العرض المكتوب فهمًا شاملاً لموضوع الثبات في القياس النفسي، مع محتوى منظم بصورة جيدة، وتحليل شامل، وعرض متكامل بالمعلومات العلمية، وجودة مرتفعة لكفاءة العرض.	٣ جيد جدا
يوضح العرض المكتوب الفهم الأساسي لموضوع الثبات في القياس النفسي، مع محتوى منظم إلى حد ما، وتحليل محدود، ويقدم المعلومات العلمية الأساسية، وجودة جيدة.	٢ جيد
يُظهر العرض المكتوب فهمًا محدودًا لموضوع الثبات في القياس النفسي، مع محتوى غير منظم، وتحليل سطحي، وعرض غير كافٍ للمعلومات العلمية بمستوى ضعيف من الجودة.	١ ضعيف

جدول (٣) سلم التقدير التحليلي

المعيار ١	٤ ممتاز	٣ جيد جد	٢ جيد	١ ضعيف
المحتوى المعرفي للعرض	يُظهر فهمًا عميقاً ومتميزاً لموضوع الثبات في القياس النفسي، بما في ذلك جميع طرق تقدير الثبات والمفاهيم والنظريات الأساسية بدقة وفعالية.	يُظهر فهمًا قويًا لموضوع الثبات في القياس النفسي، ويغطي الطرق الأساسية لتقدير الثبات والمفاهيم والنظريات الأساسية.	يوضح الفهم الأساسي لموضوع الثبات في القياس النفسي إلى حد ما، لكن يفتقر إلى العمق ويتجاهل بعض الطرق الأساسية في تقدير الثبات وبعض المفاهيم الأساسية.	يُظهر فهمًا محدودًا لموضوع الثبات في القياس النفسي، مع وجود قصور كبير أو عدم دقة في المعرفة، ويتجاهل المفاهيم والطرق الأساسية لتقدير الثبات.
المعيار ٢	٤ ممتاز	٣ جيد جد	٢ جيد	١ ضعيف
تحليل المعلومات المقدمة	يقدم تحليلاً مميز وواضح لموضوع الثبات في القياس النفسي، ويقيم بشكل نقدي الطرق المختلفة لتقدير الثبات وأهميته وتحدياته وآثاره بعمق ووضوح مع الأمثلة والأدلة المناسبة.	يقدم تحليلاً شاملاً لموضوع الثبات في القياس النفسي، ويناقش الطرق المختلفة لتقدير الثبات وأهميته وقيوده مع الأمثلة والأدلة المناسبة.	يقدم بعض التحليلات لموضوع الثبات في القياس النفسي، ولكنه يفتقر إلى العمق أو يتجاهل جوانب معينة.	تحليل لموضوع الثبات في القياس النفسي محدود أو سطحي، مع افتقاده للعمق في طرح الموضوع أو صلة محدودة بالموضوع.
المعيار ٣	٤ ممتاز	٣ جيد جد	٢ جيد	١ ضعيف
تنظيم العرض	العرض منظم بشكل جيد للغاية، مع مقدمة واضحة وفقرات المحتوى متماسكة، وانتقالات فعالة بين الأقسام وخاتمة موجزة تشمل جميع النقاط المهمة.	العرض منظم بشكل جيد، مع مقدمة واضحة، وتطور منطقي للأفكار، وخاتمة موجزة تشمل أغلب النقاط المهمة.	العرض التقديمي منظم إلى حد ما، ولكنه يفتقر إلى الوضوح في البنية أو الانتقالات بين الأقسام، وخاتمة موجزة تشمل بعض النقاط المهمة.	يفتقر العرض التقديمي إلى التنظيم والهيكل الواضحين، مما يجعل من الصعب متابعته، مع عدم وجود خاتمة أو موجز يشمل النقاط المهمة.
المعيار ٤	٤ ممتاز	٣ جيد جد	٢ جيد	١ ضعيف
وضوح الأفكار وتوصيل المعلومات	يتميز العرض المكتوب بوضوح التعبير عن الأفكار، واستخدام لغة واضحة تنقل الأفكار المعقدة بشكل فعال وموجزة.	العرض المكتوب واضحًا ومتماسكًا في الغالب، مع لغة واضحة ونقل الأفكار المعقدة بشكل جيد، مع افتقار بعض المجالات والأجزاء إلى الوضوح والدقة.	العرض المكتوب غير واضح أو مطول إلى حد ما، مع لغة قد تكون غير واضحة أو معقدة في بعض الأحيان. ويقل نقل الأفكار المعقدة بوضوح.	العرض المكتوب غير واضح ومطول جداً، ويصعب متابعته، مع وجود أجزاء كثيرة مكتوبة بأفكار غامضة غير مفهومة. ويفتقر لنقل الأفكار المعقدة بوضوح.
المعيار ٥	٤ ممتاز	٣ جيد جد	٢ جيد	١ ضعيف

المعيار ١	٤ ممتاز	٣ جيد جد	٢ جيد	١ ضعيف
اللغة والتنسيق العام	يُظهر العرض المكتوب لغة واضحة جداً، وتنسيقًا مثاليًا، مع عدم وجود أخطاء في القواعد النحوية أو الإملاء.	العرض المكتوب خالٍ في الغالب من الأخطاء اللغوية والإملائية، وتنسيق مناسب، مع وجود أخطاء بسيطة في بعض الأحيان لا تنتقص من الوضوح العام.	يحتوي العرض المكتوب على العديد من الأخطاء اللغوية وفي التنسيق، والتي قد تتكرر مما يجعل القراءة صعبة في بعض الأقسام.	العرض المكتوب مليء بالأخطاء اللغوية وغير مرتب، مما يجعل من الصعب فهمه، ويفتقد بشكل كبير لمهارة التنسيق في أغلب صفحات العرض.

الأساليب الإحصائية المستخدمة بالدراسة:

تم تطبيق أساليب تحليل سيكومترية مختلفة مثل نسبة صدق المحتوى ومؤشر صدق المحتوى للتحقق من صدق المقياسين. وتم حساب ثبات كل من سلم التقدير الكلي والتحليلي بطرق تتناسب مع طبيعة كل مقياس، حيث استخدم تحليل متوسط النسبة المئوية لدرجة الاتفاق للمحكمين لسلم التقدير الكلي، واستخدم حساب ثبات المحكمين وحساب معامل ثبات ألفا كرونباخ للمقياس لسلم التقدير التحليلي. وتم أيضا تحليل البيانات إحصائيًا باستخدام اختبار كروسكال واليس واختبار مان ويتني يو للكشف عن الفروق بين طرق تقييم أداء الطلبة باستخدام طريقة سلم التقدير الكلية وطريقة سلم التقدير التحليلية والطريقة التقليدية، ولمعرفة وجود فروق في التقييم تعزى لمتغيري نوع الجنس والدرجة العلمية لأعضاء هيئة التدريس.

نتائج الدراسة وتفسيرها:

للتحقق من فروض الدراسة واختبارها، تم استخدام الأساليب الإحصائية المناسبة. وفيما يلي عرض نتائج الدراسة وتفسيرها ومناقشتها وفقا لفروض الدراسة كالتالي:

الفرض الأول: يوجد فرق دال إحصـــائيًا بين المجموعة التجريبية الأولى والمجموعة التجريبية الثانية والمجموعة الضابطة في تقييم المهام الأدائية للطلبة.

ولعدم تحقق شروط استخدام التحليل الإحصائي البارامتري، تم إجراء اختبار كروسكال واليس لتحديد ما إذا كانت هناك فروق ذات دلالة إحصائية في درجات التقييم بين المجموعات الثلاث: الكلية والتحليلية والتقليدية (الضابطة). وتشير النتائج إلى وجود فرق إحصائي دال (H =14.716،

df = 2، p <.001)، مما يشير إلى أن درجات تقييم مجموعة واحدة على الأقل تختلف بشكل دال عن المجموعات الأخرى. ويوضح الجدول التالي نتيجة التحليل الإحصائي لاختبار كروسكال واليس.

جدول (٤) نتائج اختبار كروسكال واليس

طريقة التقييم	N	Mean Rank	Df	Statistic	P
الكلية	٢٠	٣٤,٢٥	٢	١٤,٧١٦	٠,٠٠١
التحليلية	٢٠	٣٨,١			
التحليلية	٢٠	١٩,١٥			

ولتحديد أين تكمن هذه الاختلافات، تم إجراء مقارنات لاحقة باستخدام HSD الخاص باستخدام اختبار توكي (Tukey) والموضحة في الجدول التالي.

جدول (٥) نتائج المقارنات اللاحقة باستخدام اختبار توكي

المجموعات	Test Statistic	Std. Error	Std. Test Statistic	Sig.
الضابطة – الكلية	١٥,١	٥,٢٢٢	٢,٨٩٢	٠,٠٠٤
الضابطة – التحليلية	١٨,٩٥	٥,٢٢٢	٣,٦٢٩	٠,٠٠٠
الكلية – التحليلية	٣,٨٥ -	٥,٢٢٢	٠,٧٣٧ -	٠,٤٦١

وتشير نتائج الجدول السابق إلى عدم وجود فرق ذو دلالة إحصائية في درجات التقييم بين المجموعتين الكلية والتحليلية (p =٠,٤٦١)، ووجود فرق ذو دلالة إحصائية في درجات التقييم بين المجموعتين الكلية والضابطة (p =٠,٠٠٤) لصالح المجموعة الكلية، كما أظهرت النتائج وجود فرق ذو دلالة إحصائية في درجات التقييم بين المجموعتين التحليلية والضابطة (p <٠,٠٠٠) لصالح المجموعة التحليلية. وبشكل عام تشير النتائج السابقة بأن نوع طريقة التقييم المستخدمة (الكلية، والتحليلية، والتقليدية) تؤثر بشكل كبير على درجات التصحيح عند تقييم المهام الأدائية للطلبة. وتتشابه نتائج هذه الدراسة مع دراسة عبود (٢٠١٩) التي توصلت إلى أن كلا طرق سلالم التقدير الكلية والتحليلية أفضل من الطريقة التقليدية في تقييم الأداء الدراسي. وأيضًا تشابهت النتائج مع دراسة أبو عبيد (٢٠١١) والتي تشير إلى وجود فروق دالة إحصائيًا بين طريقة التصحيح التحليلية والتقليدية لصالح الطريقة التحليلية، وبين طريقة التصحيح الكلية والتقليدية لصالح الطريقة الكلية مع عدم وجود فرق دال بين طريقتي التقييم التحليلية والتقليدية في أداء واتجاهات الطلبة. كذلك

تشابهت نتائج الدراسة الحالية مع دراسةKoizumi و In'nami و Fukazawa (2020) والتي أظهرت نتائج التحليل باستخدام نموذج راش بأن سلالم التقدير الكلية و التحليلية متشابهين و يعملان بشكل فعال في تقييم المهام الشفوية. وتختلف نتائج هذه الدراسة مع نتائج دراسة عبود (٢٠١٩) والسرحاني (٢٠١٦) و Putri و Melani (2022) بوجود اختلاف دال بين المجموعة التجريبية التحليلية والمجموعة التجريبية الكلية في تقييم الأداء لصالح المجموعات المستخدمة لطريقة سلم التقدير التحليلي .

تم أيضا استخدام الضبط الإحصائي لعزل تأثير المتغيرات الدخيلة، والتحقق من تأثير المعالجة التجريبية باستخدام سلم التقدير الكلي والتحليلي وذلك بتطبيق تحليل النموذج الخطي المعمم (GLM)لقياس أثر المتغيرات الدخيلة (نوع الجنس ومستوى الدرجة العلمية) لأعضاء هيئة التدريس، وللتحقق من أن التأثير على درجات التقييم يعزى الى طرق التقييم التجريبية المستخدمة في هذه الدراسة. وقد تم تلخيص نتائج تحليل (GLM)في الجدول أدناه .

جدول (٦) تحليل النموذج الخطي المعمم(GLM)

Source	Wald Chi-Square	Df	Sig.
(Intercept)	٨٥٨,٩٤٤	١	٠,٠٠٠
طريقة التقييم	٢٤,١٣٣	٢	٠,٠٠٠
نوع الجنس	٣,١٩١	١	٠,٠٧٤
الدرجة العلمية	٨,٠٩٢	٤	٠,٠٨٨

تشير النتائج في الجدول السابق أن قيمة والد كاي سكوير (Wald Chi-Square) تساوي (٨٥٨,٩٤٤) مع قيمة (p=٠,٠٠٠)، مما يشير إلى أن النموذج الإحصائي دال إحصائيًا وبشكل عام يعتبر مناسب جداً، وأشار تحليل طرق التقييم بأن قيمة والد كاي سكوير تساوي (٢٤,١٣٣) مع قيمة (p=٠,٠٠٠)، مما يشير إلى أن طريقة التقييم (الكلية، التحليلية، التقليدية) تؤثر بشكل كبير على درجات التقييم مع وجود فروق ذات دلالة إحصائية في الدرجات تبعا لطريقة التقييم المستخدمة، أما تحليل نوع الجنس تشير نتيجته بأن قيمة والد كاي سكوير تساوي (٣,١٩١) مع قيمة (p=٠,٠٧٤)، وهذا يعني أن نوع الجنس ليس ذا دلالة إحصائية وليس له تأثير على تقييم الدرجات، وأخيراً يظهر تحليل المستوى الأكاديمي لأفراد عينة الدراسة بأن قيمة والد كاي سكوير

تساوي (٨,٠٩٢) مع قيمة (p =٠,٠٨٨)، وهذا يعني أيضًا عدم وجود دلالة إحصائية وبالتالي يعتبر هذا المتغير كذلك ليس له تأثير على تقييم الدرجات.

وتم أيضا التحقق من حجم التأثير لطرق التقييم لأهمية فحص هذا الجانب، لأن الدلالة الإحصائية تخبرنا بوجود تأثير ما، إلا أنها لا تنقل حجم التأثير، وقد يكون هنالك دلالة إحصائية، ولكنه بدون تأثير أو صغير جدًا، وقد لا يكون له معنى عمليًا. في حين أن حجم التأثير يوفر مقياسًا لقدر وقوة التأثير، مما يساعدنا على فهم الآثار العملية للنتائج ومن المحتمل أن تكون ذات معنى في تطبيقات العالم الحقيقي. وقد تم حساب قيمة حجم التأثير إيتا سكوير(η2) والذي بلغ (٠,٢٢٣) مما يعني وجود تأثير كبير لطريقتي التقييم التحليلية والكلية، وفقًا لدليل تفسير حجم التأثير الشائعة (Cohen, 1988). وهذا يعني أن الاختلافات في تقييم الدرجات بين المجموعات (الكلية والتحليلية والتقليدية) كبيرة ولها تأثير قوي في تقييم أداء الطلبة. وتعزز نتيجة حجم التأثير الكبير هذه النتائج السابقة لاختبار كروسكال واليس والاختبارات اللاحقة، ونتائج النموذج الخطي المعمم (GLM)، مما يشير إلى أن طريقة التقييم المستخدمة لها تأثير كبير وجوهري على درجات التقييم للمهام الطلابية. وهذه النتائج توضح أن على الممارسين التربويين أن يأخذوا في الاعتبار الاختلافات الكبيرة بين طرق التقييم لتحسين دقة التصحيح وفعاليتها في التقييمات التعليمية.

وبشكل عام تشير هذه النتائج أن نوع وطريقة التقييم المستخدمة (الكلية، تحليلي، تقليدي) تؤثر بشكل كبير على درجات التقييم. كذلك توضح النتائج أن كلاً من نماذج سلم التقدير الكلية والتحليلية فعالة في تقييم الدرجات للمهام التي يقدمها الطلبة مقارنة بطرق التقييم التقليدية، مع وجود أفضلية أكبر قليلاً لنماذج التقييم التحليلية. ومع ذلك، نظرًا لعدم وجود فرق كبير بين نماذج سلم التقدير الكلية والتحليلية، يمكن اعتبار أي من الطريقتين فعالة في تقييم الأداء وبالتالي في دقة نتائج التقييم. وبالتالي يمكن استنتاج أن طريقة التقييم المستخدمة لها تأثير كبير على تقييم أداء الطلبة، مما يدعم الفرضية القائلة بأن سلالم التقدير الكلية والتحليلية مقارنة مع الطريقة التقليدية يمكن أن تؤدي إلى تقييم الدرجات بشكل مختلف. يمكن لهذه النتائج أن تفيد الممارسات التعليمية واختيار طرق التقييم لضمان التقييم العادل والفعال لأداء الطلاب، وتدعم هذه النتيجة العديد من الدراسات التي توفر رؤى حديثة حول كيفية تأثير أنواع نماذج سلالم التقدير على نتائج التقييم، وتدعم التفسير حول فعالية النماذج الكلية والتحليلية مقارنة بالطرق التقليدية في تقييم أداء الطلاب.

كما توضح هذه الدراسات أن النماذج التحليلية توفر درجة أعلى من التناسق في الدرجات مقارنة بالنماذج الكلية، وهذا يرجع إلى الطبيعة التفصيلية للنماذج التحليلية التي تحدد معايير أداء محددة ومستويات تقييم واضحة (Gibbs & Simpson, 2023 ؛Reddy & Andrade, 2021).

الفرض الثاني: يوجد فرق دال إحصائيًا بين أعضاء هيئة التدريس اللذين يستخدموا طريقة سلم التقدير الكلي في تقييم درجة الأداء تعزى لنوع الجنس والدرجة العلمية.

للتحقق من هذا الفرض، تم تطبيق اختبار مان ويتني يو لمقارنة درجات تقييم الذكور والإناث من أعضاء هيئة التدريس في المجموعة التجريبية التي استخدمت سلم التقدير الكلي وذلك لمعرفة إذا ما يوجد فرق دال إحصائيًا يعزى لمتغير نوع الجنس. ويظهر الجدول التالي نتائج هذا الاختبار .

جدول (٧) نتائج اختبار مان وتني لسلم التقدير الكلي وفقا لنوع الجنس

نوع الجنس	N	Mean Rank	Mann-Whitney U	p
ذكر	١٣	١١,٦٥	٣٠,٥	٠,١٩٦
أنثى	٧	٨,٣٦		
المجموع	٢٠			

يتضح من الجدول السابق أن قيمة Mann-Whitney U تساوي (٣٠,٥) وقيمة الدلالة الإحصائية (p =٠,١٩٦)، وهذا يعني أنه لا يوجد فرق ذو دلالة إحصائية بين درجات الذكور والإناث من أعضاء هيئة التدريس اللذين يستخدموا طريقة سلم التقدير الكلي في تقييم درجة الأداء. وبناءً على هذه النتائج، يمكن استنتاج أن الجنس لا يؤثر بشكل كبير على تقييم الدرجات المعطاة للمهمة التي تم تقييمها. وهذا يشير إلى أن طريقة التقييم عادلة وغير متحيزة فيما يتعلق بنوع الجنس. ولا توجد دراسات سابقة عرضت في هذا البحث، تتفق وتختلف مع نتائج الدراسة الحالية وذلك يعود بسبب اختلاف هذه الدراسات من حيث الأهداف وأسئلة البحث ومنهجية البحث وبالتالي النتائج، ولندرة الأبحاث التي تهتم بدراسة سلالم التقدير الكلية والتحليلية، وبالذات في العالم العربي. ولذلك سيتم عرض أوجه التشابه والاختلاف مع دراسات أخرى. ومن الدراسات التي اتفقت نتائجها مع نتائج الدراسة الحالية، دراسة Reddy and Andrade (2010) ودراسة Bresciani (2006) حيث وجدت هذه الدراسات بأن أدوات التقييم التي تصمم بشكل جيد ومنها التقييم باستخدام سلم

التقدير لا تتأثر بخصائص نوع الجنس. ولا تتفق نتائج هذه الدراسة مع نتائج دراسة Timmerman & Karp (2011) التي وجدت أن نوع جنس المعلم يمكن أن يؤثر على تقييمات الطلاب، مما يشير إلى إمكانية وجود التحيز في أنظمة التقييم حتى عند استخدام أدوات تقييم مقننة.

كما تم استخدام اختبار كروسكال واليس لمقارنة درجات تقييم أعضاء هيئة التدريس في المجموعة التجريبية التي استخدمت سلم التقدير الكلي وذلك لمعرفة إذا ما يوجد فرق دال إحصائيًا يعزى لمتغير الدرجة العلمية. والجدول التالي يوضح نتائج هذا الاختبار.

جدول (٨) نتائج اختبار كروسكال واليس لسلم التقدير الكلي وفقا للدرجة العلمية

الدرجة العلمية	N	Mean Rank	Df	Kruskal-Wallis H	P
معيد	١	١٠,٥	٤	٦,٤٩٦	٠,١٦٥
محاضر	٤	١٤,٢٥			
أستاذ مساعد	٧	٧,٢٩			
أستاذ مشارك	٥	١٣,٥			
أستاذ	٣	٨			
المجموع	٢٠				

تشير النتائج في الجدول أن قيمة Kruskal-Wallis H تساوي (٦,٤٩٦) وقيمة الدلالة الإحصائية تساوي (p =٠,١٦٥)، أي أنه لا يوجد فرق دال إحصائيًا بين أعضاء هيئة التدريس حسب الدرجة العلمية في تقييم أداء الطلبة باستخدام سلم التقدير الكلي. وبناءً على هذه النتائج، يمكن استنتاج أن الدرجة العلمية لا تؤثر بشكل كبير على تقييم أداء الطلبة. وهذا يشير إلى أن طريقة التقييم عادلة وغير متحيزة فيما يتعلق بالدرجة الأكاديمية لأعضاء هيئة التدريس المشاركين في التقييم. واتفقت نتائج الدراسة الحالية مع نتائج دراسة Reddy and Andrade (2010) والتي توصلت إلى أن التقييم لا يتأثر بالخلفية الأكاديمية للمعلمين بالتعليم العالي. واختلفت نتائج هذه الدراسة مع نتائج دراسة Maclellan(2004) التي وجدت أن الخلفية الأكاديمية من العوامل التي تأثر على التقييم المنصف ومدى استفادة التلاميذ من عملية التقييم.

يشير عدم وجود اختلافات كبيرة في درجات تقييم أعضاء هيئة التدريس بسبب نوع الجنس والدرجة الأكاديمية إلى أن معايير التقييم المستخدمة في الدراسة واضحة ومصممة بدقة. وقد تم

تصميم نموذج سلم التقدير الكلي لتوفير معايير عامة وموحدة لتقييم مستوى أداء الطلبة، مما يقلل من التحيز الشخصي. ومن خلال تحديد المعايير ووصفها بوضوح، تستند التقييمات باستخدام نموذج سلم التقدير الكلي إلى مؤشرات أداء محددة بدلاً من التحيزات الشخصية أو الأحكام الذاتية. ونستنتج من خلال النتائج السابقة أن نماذج سلم التقدير الكلي المصممة جيدًا يمكن أن تقلل من تحيز التقييم، وتقلل من التباين في تقييم الدرجات والتصحيح التي قد تتأثر بخصائص المحكمين مثل نوع الجنس. و تشير العديد من الدراسات بأن التصميم الواضح لسلم التقدير الكلي يساعد في تقليل التحيزات التي قد تنشأ نتيجة لخصائص المحكمين، ويقدم تقييما أكثر شفافية وعدلاً لأداء الطلبة (Bloxham & Boyd, 2022 ؛Dawson, 2017).

الفرض الثالث: يوجد فرق دال إحصائيًا بين أعضاء هيئة التدريس اللذين يستخدموا طريقة سلم التقدير التحليلي في تقييم درجة الأداء تعزى لنوع الجنس والدرجة العلمية

تم إجراء اختبار مان ويتني يو لمقارنة درجات تقييم الذكور والإناث من أعضاء هيئة التدريس في المجموعة التجريبية التي استخدمت سلم التقدير التحليلي، وذلك لمعرفة إذا ما يوجد فرق دال إحصائيًا يعزى لمتغير الدرجة العلمية. وعرضت نتائج هذا الاختبار في الجدول التالي.

جدول (٩) نتائج اختبار مان وتني لسلم التقدير التحليلي وفقا لنوع الجنس

نوع الجنس	N	Mean Rank	Mann-Whitney U	P
ذكر	١١	١١,٥٩	٣٧,٥	٠,٣٢٧
أنثى	٩	٩,١٧		
المجموع	٢٠			

يتضح من الجدول السابق أن قيمة Mann-Whitney U تساوي (٣٧,٥) وقيمة الدلالة الإحصائية تساوي (p =٠,٣٢٧)، ولذلك لا يؤثر نوع الجنس لأعضاء هيئة التدريس على تقييم الأداء باستخدام سلم التقدير التحليلي، وبناءً على هذه النتائج، يمكن استنتاج أن نوع الجنس لا يؤثر على تصحيح المهام الطلابية عند استخدام سلم التقدير التحليلي. وهذا أيضًا يشير إلى أن طريقة التقييم غير متحيزة وفقا لنوع الجنس لأعضاء هيئة التدريس. وتتفق نتائج الدراسة الحالية مع نتائج دراسة الغامدي (٢٠٢٢) حيث وجد أنه لا يوجد فرق دال إحصائيًا بين الطلاب والطالبات

للتقييم الذاتي للتحصيل الدراسي باستخدام مقياس سلم التقدير التحليلي. وتختلف نتائج هذه الدراسة مع نتائج دراسةTimmerman & Karp (2011) التي وجدت أن نوع جنس المعلم يمكن أن يؤثر على تقييمات الطلاب وقد يكون أحد العوامل لوجود التحيز في تقييم أداء الطلبة.

وتم أيضًا استخدام اختبار كروسكال واليس لمقارنة درجات تقييم أعضاء هيئة التدريس في المجموعة التجريبية التي استخدمت سلم التقدير التحليلي وذلك لمعرفة إذا ما يوجد فرق دال إحصائيًا يعزى لمتغير الدرجة العلمية. وتم الوصول إلى النتائج المعروضة في الجدول التالي.

جدول (١٠) نتائج اختبار كروسكال واليس لسلم التقدير التحليلي وفقا للدرجة العلمية

الدرجة العلمية	N	Mean Rank	df	Kruskal-Wallis H	P
معيد	٢	١٦,٥	٤	٤,١٢٢	٠,٣٩٠
محاضر	٣	١١,٨٣			
أستاذ مساعد	٧	٧,٩٣			
أستاذ مشارك	٥	١٠,٥			
أستاذ	٣	١١,١٧			
المجموع	٢٠				

يتضح من الجدول السابق أن قيمة Kruskal-Wallis H تساوي (٤,١٢٢) وقيمة الدلالة الإحصائية تساوي (p =٠,٣٩٠)، أي أنه لا يوجد فرق دال إحصائيًا بين أعضاء هيئة التدريس حسب الدرجة العلمية في تقييم أداء الطلبة باستخدام سلم التقدير التحليلي. وبناءً على هذه النتائج يمكن استنتاج أن الدرجة الأكاديمية لأعضاء هيئة التدريس ليس لها تأثير ذو دلالة إحصائية على درجات التقييم للمهام الأدائية للطلبة. وهذا يعني أيضا أن أسلوب التقييم باستخدام سلم التقدير التحليلي عادل وغير متحيز باختلاف الدرجات العلمية لأعضاء هيئة التدريس. وتتفق نتائج دراسة(2010) Reddy and Andrade مع الدراسة الحالية في أن عملية التقييم لا تتأثر بالخلفية الأكاديمية للمعلمين. وتختلف نتائج دراسة Sadler (2009) مع نتيجة الدراسة الحالية في أن هنالك درجة معينة من تأثير الذاتي على التقييم بسبب الدرجة العلمية والخلفيات والخبرات التي يمتلكها المقيمين .

ويشير غياب الاختلافات وعدم وجود فروق ذات دلالة إحصائية في درجات التقييم بسبب نوع الجنس والدرجة الأكاديمية لأعضاء هيئة التدريس عند استخدام سلم التقدير التحليلي إلى أن

نماذج التقييم عادلة ومصممة بشكل جيد، من حيث التوصيف الواضح للمعايير ومستويات الأداء. ولذلك يمكن اعتبار هذه النتيجة إيجابية من الناحية السيكومترية مما يعزز الصدق والموضوعية والاتساق والدقة لهذا النوع من المقاييس. حيث يتم تصميم نماذج سلم التقدير التحليلية لتوفير إطار موحد للتقييم، وبالتالي تقليل التحيز الشخصي. ومن خلال تحديد المعايير ومستويات الأداء ووصفها بوضوح، تضمن هذه النماذج أن تستند التقييمات إلى مؤشرات أداء محددة بدلاً من التحيزات الشخصية أو الأحكام الذاتية. أوضحت الكثير من الدراسات أنه عندما يتم تصميم نماذج سلم التقدير التحليلي بمعايير ومستويات أداء واضحة سيعزز ذلك من الصدق والموضوعية لأداة التقييم، وسيقلل من تأثير التحيزات و الخصائص الشخصية مما يضمن تقييمًا أكثر عدالة في عملية تقييم أداء الطلبة (Feldman & March, 2022 ؛Panadero & Alonso-Tapia, 2021).

توصيات الدراسة

تم استخلاص التوصيات من خلال نتائج الدراسة الحالية، وهي كالتالي:

١. اعتماد استخدام سلم التقدير التحليلي والكلي لتقييم المهام الأدائية كالمشاريع والواجبات والعروض حيث أظهرت هذه الطرق أنها أكثر فعالية في تقييم أداء الطلاب مقارنة بالطريقة التقليدية التي لا تعتمد على معايير واضحة للتقييم؛ لذلك يجب على المؤسسات التعليمية النظر في اعتماد نماذج سلم التقدير التحليلية والكلية كأسلوب تقييم مهم وضروري لتعزيز دقة وشمولية تقييمات الطلاب، وعدم الاعتماد الكلي على أساليب التقييم التقليدية لصالح نماذج تقييم أكثر حداثة وفعالية لتحسين نتائج التقييم وموثوقيته.

٢. يعد استخدام نماذج سلم التقدير التحليلية والكلية أكثر عدلاً وأقل تحيزاً، حيث أظهرت النتائج عدم وجود فروق ذات دلالة إحصائية في درجات التقييم تعزى إلى نوع الجنس والمستوى الأكاديمي لأعضاء هيئة التدريس عند استخدام نماذج التقييم الكلية والتحليلية.

٣. تنفيذ برامج التدريب على الاستخدام الفعال لنماذج سلم التقدير الكلي والتحليلي لأعضاء هيئة التدريس والتأكد على أن لديهم دراية جيدة بتصميم وتطبيق طرق التقييم الكلية والتحليلية. وهذا يشمل ورش العمل والندوات وفرص التطوير المهني المستمر.

مقترحات الدراسة:

تقترح الدراسة الحالية بالاتجاهات البحثية المستقبلية التالية لمزيد من استكشاف موضوع طرق التقييم باستخدام سلالم التقدير (الكلية والتحليلية) وتأثيرها على مخرجات التعليم:

١. دراسة الآثار طويلة المدى لاستخدام نماذج سلالم التقدير الكلية والتحليلية على أداء الطلاب ونتائج التعلم على مدى سنوات أكاديمية متعددة. حيث ان فهم التأثير المستدام لهذه المعايير يمكن أن يوفر رؤى أعمق حول فعاليتها وفوائدها المحتملة للتطوير المستمر للطلاب.

٢. دراسة مقارنة فعالية نماذج مصفوفات التقييم الكلية والتحليلية عبر مختلف المواضيع والتخصصات. قد تتطلب المواضيع المختلفة أساليب تقييم مختلفة، ولكن ستساعد هذه الدراسات في تحديد المزايا الخاصة بالمواضيع التي تندرج تحت المجالات والتخصصات العلمية المختلفة.

٣. دراسة تقييم تأثير برامج التدريب المستهدفة لأعضاء هيئة التدريس على دقة ومصداقية التقييمات القائمة على نماذج التقييم باستخدام سلالم التقدير الكلية والتحليلية. ويعد كفاءة أعضاء هيئة التدريس في استخدام نماذج التقييم أمرًا ضروريًا للحفاظ على جودة التقييم ومعرفة تأثير التدريب على تقليل التباين في تقييم الدرجات.

٤. تقديم دراسات تستهدف فحص تصورات واتجاهات الطلاب حول عملية وفائدة استخدام نماذج التقييم الكلية والتحليلية في تقييماتهم. وكون الطلبة العنصر الأهم في العملية التعليمية، فإن فهم كيفية إدراك الطلاب لهذه الطرق من التقييم يمكن أن يوفر تغذية راجعة قيمة لتحسين ممارسات التقييم.

المراجع العربية

أولًا: المراجع العربية:

أبو عبيد، أحمد علي خلف. (٢٠١١). طريقة قواعد التصحيح (SCORING RUBRICS) في تقييم الأداء وأثرها في تحصيل واتجاهات طلاب الصف الأول الثانوي نحو مادة الرياضيات. مجلة دراسات نفسية وتربوية، ع ٧، ٢٥ – ٥٧.

جامعة الإمام محمد بن سعود الإسلامية (د. ت.). قسم علم النفس: التعريف بالقسم. مسترجع بتاريخ أبريل ١٣، ٢٠٢٤، من موقع https://units.imamu.edu.sa/administrations/Psychology/profile/Pages/default.aspx

السرحاني، محمد بن فاهد سالم. (٢٠١٦). فاعلية استخدام قواعد تقدير الأداء التحليلية لحل المشكلات الرياضية في تنمية التحصيل الدراسي لدى عينة من طلاب الصف الأول المتوسط بمنطقة الجوف. *مجلة التربية*، ع١٧١, ج١، ٦٦٤ – ٦٨٤.

ظاظا، حيدر. (٢٠٢١). بناء سُلَّمْ تقدير لفظي" روبرك" لمراجعة الأقران للمخْطُوطُ البحثي في العلوم التربوية والنفسية. دراسات: العلوم التربوية, ٤٨(٢) , ١٢٧–١٤٣.

عبود، يسري زكي. (٢٠١٩). أثر استخدام مقاييس التقدير المتدرجة على التحصيل الدراسي والكفاءة الذاتية الأكاديمية ومهارات التقويم الذاتي لدى طالبات كلية التربية في جامعة الملك فيصل. رسالة الخليج العربي، س٤٠, ع١٥١، ٧٧ – ٩٥.

الغامدي، احمد إبراهيم ساعد. (٢٠٢٢). تقييم التحصيل الدراسي للمقررات العملية باستخدام مقياس التقدير المتدرج لطلاب التربية الفنية بجامعة الباحة. بحوث في التربية الفنية والفنون, ٢٢(٢) , ١٨٩–٢٠٨.

ثانيًا: ترجمة المراجع العربية

Abu Obaid, Ahmed Ali Khalaf. (2011). The SCORING RUBRICS method in performance evaluation and its impact on the achievement and attitudes of first-year secondary school students towards mathematics. Journal of Psychological and Educational Studies, No. 7, 25 - 57.

Imam Muhammad bin Saud Islamic University (n.d.). Department of Psychology: Introduction to the Department. Retrieved on April 13, 2024, from https://units.imamu.edu.sa/administrations/Psychology/profile/-Pages/default.aspx.

Al-Sarhani, Muhammad bin Fahed Salem. (2016). The effectiveness of using analytical performance assessment rules to solve mathematical problems in developing academic achievement among a sample of first-year middle school students in Al-Jawf region. Journal of Education, No. 171, Vol. 1, 664 - 684.

Dhada, Haider. (2021). Building a Rubric Verbal Rating Scale for Peer Review of Research Manuscripts in Educational and Psychological Sciences. *Studies: Educational Sciences,* 48(2), 127–143.

Aboud, Yasser Zaki. (2019). The Effect of Using Graduated Rating Scales on Academic Achievement, Academic Self-Efficacy, and Self-Assessment Skills of Female Students of the College of Education at King Faisal University. *Arabian Gulf Message,* Vol. 40, No. 151, 77–95.

Al-Ghamdi, Ahmed Ibrahim Saed. (2022). Evaluation of Academic Achievement of Practical Courses Using the Graduated Rating Scale for Art Education Students at Al-Baha University. *Research in Art Education and Arts,* 22(2), 189–208.

ثالثًا: المراجع الأجنبية:

Allen, D., & Tanner, K. (2006). Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE life sciences education,* 5(3), 197–203.

Andrade, H., & Du, Y. (2005). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education,* 30(1), 87–102.

Bharuthram, S., & Patel, M. (2017). Co-constructing a rubric checklist with first year university students: A self-assessment tool. *Apples - Journal of Applied Language Studies,* 11(4), 35–55.

Black, P. J., &Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice,* 5(1), 7-74.

Bloxham, S., & Boyd, P. (2022). Developing Holistic Rubrics for University Assessments: Ensuring Fairness and Transparency. *Assessment in Education: Principles, Policy & Practice,* 29(2), 170-185.

Bresciani, M. J. (2006). Outcomes-Based Academic and Co-Curricular Program Review: A Compilation of Institutional Good Practices. Stylus Publishing, LLC.

Brookhart, S. M. (2013). How to create and use rubrics for formative assessment and grading. Alexandria, VA: Association for Supervision & Curriculum Development (ASCD).

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review,* 67(3), 343-368.

Chukwuere, J. E. (2021). The comparisons between the use of analytic and holistic rubrics in information systems discipline. *Academia Letters,* Article 3579.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2. Auflage). Hillsdale, NJ: Erlbaum.

Cohen, R. J., & Swerdlik, M. E. (2017). Psychological Testing and Assessment: An Introduction to Tests and Measurement (9th ed.). McGraw-Hill Education.

Dawson, P. (2017). Assessment Rubrics: Towards Clearer and More Replicable Design, Research and Practice. *Assessment & Evaluation in Higher Education,* 42(3), 347-360.

Ebel, R. L., &Frisbie, D. A. (1986). Essentials of educational measurement (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

García-Ros, R., Ruescas-Nicolau, M. A., Cezón-Serrano, N., Carrasco, J. J., Pérez-Alenda, S., Sastre-Arbona, C., San Martín-Valenzuela, C., Flor-Rufino, C., & Sánchez-Sánchez, M. L. (2021). Students' Perceptions of Instructional Rubrics in Neurological Physical Therapy and Their Effects on Students' Engagement and Course Satisfaction. *International journal of environmental research and public health,* 18(9), 4957.

Gibbs, G., & Simpson, C. (2023). Assessing Student Work: Holistic vs. Analytical Rubrics and Their Impact on Grading Consistency. *Teaching in Higher Education,* 28(3), 340-355.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education,* 57(4), 2333-2351.

Goulden, N. R. (1994). Relationship of analytic and holistic methods to raters' scores for speeches. *Journal of Research & Development in Education,* 27(2), 73–82.

Jönsson, A., & Balan, A. (2018). Analytic or holistic: a study of agreement between different grading models. *Practical Assessment, Research & Evaluation,* 23(12).

Jonsson, A., &Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review,* 2(2), 130–144.

Koizumi, R, In'nami Y., & Fukazawa, M. (2020). Comparison between holistic and analytic rubrics of a paired oral test. *JLTA Journal,* 23, 57–77.

Lawshe, C. H. (1975). The quantitative approach to content validity. *Personnel Psychology,* 28,563-575.

Maclellan, E. (2004). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education,* 29(3), 309-322.

Mertler, Craig A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research, and Evaluation*, Vol. 7, Article 25.

Moskal, B. M. (2000). Scoring Rubrics: What, When, and How?. *Practical Assessment, Research & Evaluation,* 7(1): 3.

Panadero, E., & Alonso-Tapia, J. (2021). Rubrics for Assessment: A Review and New Directions. *Assessment & Evaluation in Higher Education,* 46(1), 17-31.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review,* 9, 129-144.

Panadero, E., Jonsson, A., &Strijbos, J. W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. *International Journal of Technology and Design Education,* 26(2), 221–236.

Panadero, E., Jonsson, A., Pinedo, L., & Fernández-Castilla, B. (2023). Effects of Rubrics on Academic Performance, Self-Regulated Learning, and self-Efficacy: a Meta-analytic Review. *Educational Psychology Review, 35(4),* 113.

Putri, E. S., &Melani, M. (2022). Holistic Vs. Analytic Evaluation In Writing Test Of Eighth Grade Students. *Journal Of English Language Studies,* 7(1), 60–77.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education,* 35(4), 435-448.

Reddy, Y. M., & Andrade, H. L. (2021). Impact of Rubric Type on Student Learning Outcomes: Holistic vs. Analytical Rubrics. *Assessment & Evaluation in Higher Education,* 46(4), 567-581.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education,* 34(2), 159-179.

Selke, M. J. G. (2013). Rubric assessment goes to college: Objective, comprehensive evaluation of student work. Lanham, MD: Rowman & Littlefield.

Stevens, D. D., & Levi, A. (2013). Introduction to rubrics: an assessment tool to save grading time, convey effective feedback, and promote student learning. 2nd ed. Sterling, Va., Stylus.

Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies,* 53(4), 466–478.

Timmerman, T. A., & Karp, D. R. (2011). The influence of gender on students' evaluation of teaching: A metaanalysis. *Research in Higher Education,* 52(7), 545-567.

Tomas, C., Whitt, E., Lavelle-Hill, R., & Severn, K. (2019). Modelling holistic marks with analytic rubrics. *Frontiers in Education: Assessment, Testing and Applied Measurement*, 4(89).

Trochim, W. M., & Donnelly, J. P. (2008). The research methods knowledge base (3rd ed.).

Tshering, &Phu-ampai, S. (2022). Effects of Using Rubrics on the Learning Achievement of Students in Educational Assessment and Evaluation. *Educational Innovation and Practice,* 3(1), 75–88.

Walvoord, B.E. (2010). Assessment Clear and Simple: A practical guide for institutions, departments, and general education, 2E. San Francisco, CA: Jossey Bass.

Wiggins, G., & McTighe, J. (1998). Understanding by Design. Alexandria, VA: Association for Supervision and Curriculum Development.

Wolf, K., & Stevens, E.A. (2007). The Role of Rubrics in Advancing and Assessing Student Learning. *The Journal of Effective Teaching,* 7, 3-14.

Yamanishi, H., Ono, M., &Hijikata, Y. (2019). Developing a scoring rubric for L2 summary writing: a hybrid approach combining analytic and holistic assessment. *Language Testing in Asia*, 9, 1-22.

Yune, S. J., Lee, S. Y., Im, S. J., Kam, B. S., &Baek, S. Y. (2018). Holistic rubric vs. analytic rubric for measuring clinical performance levels in medical students. *BMC medical education,* 18(1), 124.