الشبكات البيزية وطرق بنائها وتطبيقاتها في التقييم التربوي

د. إقبال زين العابدين درندري[1]

(قدم للنشر 1445/08/22-وقبل 1445/10/06 هـ)

المستخلص: يتناول هذا البحث التعريف بالشبكات البيزية كأحد النماذج الحديثة في القياس والتقييم النفسي والتربوي المبنية على نظرية الاحتمالات ونماذج الرسوم البيانية. كما يستعرض مفاهيم الشبكات البيزية وأسسها النظرية، ومميزاتها، إضافة إلى مكوناتها، وعلاقتها بالتصميم المرتكز على الأدلة، والتقييمات التي تتضمن نماذج كفايات متعددة الأبعاد، ومؤشرات المطابقة وطرق التقدير وأهم البرامج الإحصائية لها. كما يتناول البحث خطوات بنائها وتطبيقاتها وكيفية استخدامها في مجال التقييم التربوي، وبعض الدراسات التي طبقتها في مجال التقييم، وأهم أوجه النقد والتوجهات المستقبلية لها. وخلص البحث لأهمية التوعية بنماذج الشبكات البيزية واستخداماتها عند بناء أدوات التقييم بغرض التشخيص والتنبؤ ومتابعة وتحسين التعلم، وإجراء الدراسات حول الخصائص السيكومترية لها في سياقات القياس والتقييم النفسي والتربوي.

الكلمات المفتاحية: الشبكة البيزية، التصميم المرتكز على الأدلة، نماذج التقييم التربوي.

Bayesian Networks and Their Applications in Educational Assessment

Eqbal Z. Darandari[1]

(Submitted 03-03-2024 and Accepted on 15-04-2024)

Abstract: This research introduces the Bayesian Networks (BNs) as one of the modern models in psychological and educational measurement and assessment that are based on probability theory and graph models. It also reviews the concepts of Bayesian Networks (BNs) and their theoretical foundations, their advantages, in addition to their components and their relation to evidence centered design -ECD and assessments that include multidimensional proficiency models, fit indices, estimation methods, and their important statistical programs. The research also addresses the steps of their construction, and their applications, how to use them in the field of educational assessment, some studies that applied them in the area of assessment, the important criticisms to them, and their future directions. The research concluded that it is important to raise the awareness of Bayesian network (BN) models and their uses in constructing assessment tools for the purposes of diagnosis, prediction, and learning monitoring and improvement, and to conduct studies on their psychometric properties in the contexts of psychological and educational measurement and assessment.

Keywords: Bayesian network, evidence-centered design, educational assessment models.

(1) Department of Psychology – King Saud University | *(1) قسم علم النفس – جامعة الملك سعود*

E-mail: eqbal@ksu.edu.sa

مقدمة

ازداد الاهتمام مؤخراً بأنظمة التعليم وتحسين كفاءتها في جميع أنحاء العالم. وتطلب ذلك رفع كفاءة التعليم ووضع أنظمة للمساءلة، تثبت تحقيق الأهداف والنواتج التعليمية المتوقعة. وتزايد اعتماد أنظمة المساءلة على أساليب التقييم التربوي التقليدية، مما تسبب في التركيز على التقييم التربوي المقنن وتحسينه بشكل أكبر، لتوفير أدوات تساعد في اتخاذ القرارات التعليمية بشكل أكثر كفاءة. ولكن معظم أساليب التقييم التربوي التقليدية، وخاصة تلك الموجودة بالفعل كجزء من أنظمة المساءلة، توفر معلومات عالية المستوى، ولكنها حول المجالات التعليمية الواسعة (مثل الرياضيات والقراءة والعلوم، وما إلى ذلك). كما تشترك هذه التقييمات التربوية في أنها تستهدف تفسير الأشياء التي يقولها الطلاب، أو يقومون بها، أو يؤدونها، للتوصل إلى استنتاجات حول ما يعرفونه أو ما يمكنهم القيام به على نطاق أوسع. وخلال القرن الماضي، طور عدد من أساليب التقييم لتحقيق ذلك بطريقة منهجية. فظهرت نماذج القياس لنظرية الاختبار الكلاسيكية (Classical Test Theory-CTT)، ونظرية الاستجابة للمفردة (Item Response Theory-IRT) وتحليل الفئات الكامنة (Latent Class Analysis-LCA)، ونمذجة التشخيص المعرفي (Cognitive Diagnosis Modeling-CDM) (علام، 2005؛ Almond et al., 2015).

ورغم أن أساليب التقييم المبنية على هذه النماذج قد أثبتت مناسبتها للاختبارات واسعة النطاق والاختبارات في الفصول الدراسية، إلا أنها لا تساعد كثيرًا في توجيه التعلم وتقييم تقدم الطلاب. وبتقدم العلوم المعرفية والتعليمية، زادت أنواع المعرفة والمهارات التي يلزم تطويرها لدى الطلاب، وظهرت الحاجة إلى أنواع أدلة مناسبة لتثبت تحققها. وفي الآونة الأخيرة، أصبحت المناهج الدراسية تُبنى على الكفايات التي يجب على كل طالب اكتسابها؛ وأدى ذلك للتحول من قياس التقدم باستخدام النموذج القائم على المعرفة (Knowledge-based model) إلى النموذج القائم على الكفاية (Competence-based model). ويتطلب تحسين الكفاية التعليمية تزويد المعلم والطالب بالمحتوى والمهارات المتقنة وغير المتقنة وبشكل تفصيلي، كما يتطلب تقديم تحليل لتحصيل الطلاب لمجموعة متنوعة من المجالات الفرعية وبعدد أصغر من المفردات المناسبة. وهنا تظهر المشكلة الأساسية في استخدام أساليب التقييم التقليدية، حيث يصعب استخلاص استنتاجات دقيقة حول ما يعرفه الطالب أو يستطيع أن يعمله، من خلال ملاحظات محدودة لما يقوله أو يقوم به بالفعل، في مواقف محدودة نسبيًا (Mislevy et al., 2005).

كما تتعامل طرق التقييم التربوي التقليدية عادة مع المجال الذي يتم تقييمه باعتباره بنية واحدة، وتستخدم درجة كلية واحدة لتحصيل الطالب، ونموذجًا كامناً أحادي البعد لتمثيله، وبالتالي تكون مفيدًة في القرارات التعليمية البسيطة، وليس المتقدمة. وهناك عدة فئات من النماذج التي تقدم تقديرًا لقدرات المجالات الفرعية، بما في ذلك نماذج نظرية الاستجابة للمفردة متعددة الأبعاد (Multidimensional Item Response Theory -MIRT) (Reckase, 2009) ونماذج التصنيف التشخيصي (Diagnostic Classification Models -DCM) (Rupp et al., 2010). ولكن معظم هذه النماذج تترك البنية للقدرات الكامنة غير محددة، وتتجاهل غالباً بنية المجالات الفرعية، وتعاملها على أنها مستقلة أو مرتبطة بشكل عشوائي (Culbertson, 2014).

ومن هنا ظهرت الحاجة لتطوير أساليب تقييم تربوي تقدم معلومات تفصيلية دقيقة وقابلة للتنفيذ في البيئات التعليمية للمعارف والمهارات الفرعية والعلاقات بينها. كما ظهرت الحاجة إلى نماذج قياس

تحقق دقة أكبر واتساقاً أفضل من خلال نمذجة العلاقات بين المجالات الفرعية وبشكل واضح. وتمكِّن النماذج المبنية على الكفايات من تصميم العلاقات الهيكلية بين المجالات الفرعية بشكل واضح، كما تقدم النمذجة بالرسوم البيانية، مثل الشبكات البيزية (Bayesian Networks-BNs) ونماذج المعادلات البنائية (Structural Equation Models-SEM)، طرقاً بديلة لتمثيل ونمذجة هذه العلاقات الهيكلية بوضوح (Mislevy et al., 2000; Mislevy et al., 2005).

وتوفر الشبكات البيزية (BNs) إطارًا لنمذجة العلاقات المعقدة؛ وبالتالي فهي مناسبة تمامًا لنمذجة الأوصاف التفصيلية لمعرفة الطلاب. وعلى عكس معظم أساليب التقييم التقليدية (مثل الاختبارات) التي تقيس مقدار ما يعرفه الطالب، فإن النموذج الشبكي البيزي (BN) يحتوي على "نموذج المجال" (Domain model) المراد قياسه، والذي يقوم على تحديد احتمال لكل مجموعة فرعية من نموذج المجال، يشير إلى مدى احتمالية أن يعرف الطالب تلك المجموعة الفرعية فقط من المجال؛ مما يساعد بطرق مختلفة ومناسبة على اتخاذ القرار. وتستفيد الشبكات البيزية (BNs) من الربط بين نظرية الاحتمالات (Probability theory) ونماذج الرسم البياني (Graphical models) لتمثيل العلاقات الاحتمالية بين عدد كبير من المتغيرات؛ لذا تم استخدامها كأدوات نمذجة مرنة، عبر مجموعة واسعة من مجالات البحث والتطبيقات (Mislevy, 1994; Cruz et al., 2020).

ومن أكثر تطبيقات الشبكات البيزية (BNs) انتشاراً في التعليم الأنظمة التعليمية التكيفية (Adaptive Educational Systems-AES)، وأنظمة التدريس الذكي (Intelligent Tutoring Systems - ITS)، حيث تعد إحدى تقنيات الذكاء الاصطناعي التي يمكن استخدامها للتعامل مع مشكلة عدم اليقين في نمذجة تعلم الطلاب. وقد أصبحت تُستخدم لنمذجة تعلم الطلاب، ولتصنيفهم بناءً على الأداء المتوقع، وللتنبؤ بسلوكهم، وطرق مساعدتهم وأساليبهم المحتملة لحل المشكلات، ولاتخاذ قرارات التدريس (Bekele & Menzel, 2005; Almond et al., 2007; Chanthiran et al., 2022).

ورغم أن الشبكات البيزية (BNs) تتمتع بالقدرة على التعامل مع تعقيد نماذج القياس المتاحة في المجال التعليمي، إلا أن استخدامها في مجال القياس والتقييم النفسي والتربوي يعتبر قليلاً مقارنة بالمجالات الأخرى ذات العلاقة. وفي العقود الأخيرة، زاد الاهتمام بها، لمناسبتها لنمذجة البيانات ذات البنية المتقدمة، حيث تساعد على تصور العلاقات الأكثر تعقيدًا بين المتغيرات، إضافة لقدرتها على نمذجة عدم اليقين المتأصل في النمذجة المعرفية، حيث تعمل الشبكات البيزية (BNs) كنموذج قياس (Measurement model)؛ وبالرغم من ذلك، فإن تطبيقات هذه النماذج ما زالت محدودة عالميًا في مجال القياس والتقييم النفسي والتربوي (Desmarais & Gagnon, 2006; Almond et al., 2015; Culbertson, 2016; Uglanova, 2021).

وتعد الشبكات البيزية (BNs) أسلوباً سيكومترياً جديداً، ولكن جوانب تطبيقه غير معروفة جيدًا في بعض المجالات. وتتفاقم هذه المشكلة في العالم العربي، حيث لا توجد دراسات -في حدود علم الباحثة- تناولت الشبكات البيزية (BNs) كنوع جديد نسبيًا من نماذج القياس في مجتمع القياس والتقييم النفسي والتربوي؛ مما أدى لضعف الاستفادة من هذه النماذج وتطبيقاتها لتحسين دقة نماذج القياس المستخدمة لنمذجة تعلم الطلاب وتقييم التعلم. وهذا بدوره يقودنا للتساؤل عن الإطار النظري للشبكات البيزية (BNs)، وطبيعتها وميزاتها، ومكوناتها، وخطوات بنائها، وطرق التقدير والمعايرة فيها، وكيفية استخدامها في التقييم التربوي،

وذلك بهدف توفير معلومات نظرية وتطبيقية عنها تساعد في التعرف عليها والاستفادة منها.

وبناءً على ذلك، فإن البحث الحالي يتناول المحاور التالية: 1) التعريف بالشبكات البيزية (BNs) والإطار النظري لها؛ 2) مميزات الشبكات البيزية (BNs)؛ 3) خطوات بناء الشبكات البيزية (BNs) ومكوناتها في مجال التقييم التربوي؛ 4) مؤشرات مطابقة النموذج وطرق التقدير والمعايرة للشبكات البيزية (BN)؛ 5) أهم البرامج الإحصائية لتحليلات الشبكات البيزية (BNs)؛ 6) تطبيقات الشبكات البيزية (BNs) في التقييم التربوي؛ و7) التعقيب والخلاصة والتوصيات.

وتكمن أهمية البحث الحالي في أنه يوجه اهتمام الباحثين في مجال القياس والتقييم النفسي والتربوي نحو الشبكات البيزية (BNs)، كما سيوفر معلومات تساعدهم في توظيفها وصنع القرارات بأعلى كفاءة ممكنة. كما سيقدم البحث معلومات وإرشادات تساعد مطوري أساليب التقييم النفسي والتربوي في الحصول على بدائل للتعامل مع البناءات المعقدة للمعرفة والتحصيل وتقييمها ومتابعة تقدمها والتنبؤ بها.

أولاً: التعريف بالشبكات البيزية (BNs) والإطار النظري لها:

الشبكة البيزية (BN) هي بنية رسوم بيانية احتمالية توضح العلاقات السببية بين المتغيرات المترابطة، وتوفر طريقة مناسبة لتحديد التوزيعات الاحتمالية المشتركة (Joint probability distributions) المعقدة لمجموعة من المتغيرات، سواء الملاحظة أو الكامنة. وهي مزيج من نظرية الرسم البياني (Graph theory) والاستدلال البيزي (Bayesian inference)، حيث يمثل الرسم البياني الجوانب الرئيسة لهيكل التوزيع (Uglanova, 2021). وقد قدم نابوليتان (Neapolitan, 2004) وبيرل (Pearl, 1988) معلومات حول الأسس النظرية والمنهجية لها. وفي تطبيقات التعليم، يمكن أن تستخدم الدوائر للإشارة إلى نموذج الكفاية، والمثلثات للإشارة إلى متغيرات نموذج الأدلة.

وتفترض الشبكات البيزية (BNs) تطبيق نظرية بيز (Bayes' theorem) للاستنتاج الاحتمالي حول متغير كامن (Pearl, 1988)، حيث تشير (x) إلى المتغير الملاحظ، والذي يمثل عمل الطالب، وتشير(θ) إلى المتغير الكامن، والذي يمثل خاصية أو سمة للطالب، ومن ثم يتم التعبير عن العلاقة الشرطية بين هذه المتغيرات من خلال نظرية بيز بالمعادلة التالية:

الاستدلال حول المتغيرات الكامنة (θ) للفرد، بالنظر إلى الملاحظات (x):

$$P(\theta \mid x) = P(x \mid \theta) * P(\theta) P(x) \quad (1)$$

$$p(\theta|\boldsymbol{x}) \propto p(\boldsymbol{x}|\theta)p(\theta)$$

حيث P (θ | x) هو التوزيع الاحتمالي البعدي (Posterior probability distribution)، وهو توزيع المتغير الكامن المشروط بالمتغير الملاحظ؛ و P (θ) هو الاحتمال القبلي (Prior probability) الذي يمثل المعرفة السابقة حول توزيع المتغير الكامن، دون النظر إلى البيانات التجريبية؛ و P (x | θ) هو ما يسمى بالاحتمالية التي توضح احتمالية (Plausibility) البيانات، بالنظر إلى مَعلمات النموذج؛ و P (x) هو التوزيع الاحتمالي للمتغير الملاحظ غير المشروط (Unconditional) بأي متغير آخر (Uglanova, 2021). ويمكن التعرف على التوزيع القبلي من الأبحاث السابقة، أو البيانات التجريبية، أو الاختبارات المسبقة أو آراء الخبراء (Almond et al., 2015).

ثانياً: مميزات الشبكات البيزية (BNs):

هناك عدة جوانب يمكن أن تقدمها الشبكات البيزية (BNs)، حيث يمكن لها أن تتعامل بسهولة مع المعلومات غير الدقيقة أو المشكلات المتعلقة بالبيانات،

حيث تقوم بنمذجة العلاقات المعقدة بين المتغيرات العشوائية بطريقة مباشرة، وتسمح للفرد بالتعرف على العلاقات السببية، بالإضافة إلى أنه يمكن أيضًا تَصوُّر هذه العلاقات. ويمكن أن تسهّل الجمع بين المعرفة بالمجال والبيانات، كما توفر نهجًا فعالاً لتجنب الإفراط في مطابقة البيانات (Chanthiran et al., 2022).

ويتميز التقييم باستخدام الشبكات البيزية (BNs) بتقديم تقرير حول الاحتمالات المختلفة للنتائج، ودمج المعرفة المتخصصة حول المجال المعرفي بالمعرفة الاحتمالية، حيث يمكن التعامل مع معرفة الخبراء بالمجال كتوزيع قبلي للبيانات الإحصائية بطريقة عملية. كما يمكن لهذه الشبكات "التعلم" من البيانات، والتعامل مع النماذج والمهام المعقدة (مثل: المشكلات متعددة الخطوات، والاستجابات المعقدة المفتوحة، الخ). وهي تزود صانعي القرارات التعليمية بفكرة واضحة عن احتمالية كل جزء من الكفاية (Mislevy et al., 2005; Almond et al., 2015).

ثالثاً: خطوات بناء الشبكات البيزية (BNs) ومكوناتها في مجال التقييم التربوي:

هناك عدة خطوات لبناء الشبكة البيزية (BN) كما يلي:

1- **تحليل مجال المعرفة (Knowledge domain):** وهي عملية جمع وتنظيم المتطلبات والمعلومات حول مجال التقييم، وتتضمن "تحديد المتغيرات الكامنة" وبنيتها السيكومترية، وهذا ينطوي على تحليل المعرفة المستهدفة أو مجال المهارة إلى أجزاء يصبح كل منها عقدة.

2- **نمذجة المجال (Domain modeling):** تطوير الأدلة للتقييم ورسم النماذج الأساسية، حيث يتم تحديد "نموذج القياس" عن طريق تمثيل المجالات الفرعية كمتغيرات كامنة موزعة بشكل طبيعي مع علاقات تعويضية (Compensatory)، كما هو الحال في نموذج المعادلة البنائية أو نموذج المسار؛ ووضع هيكل النموذج (Structure of the model)، والعلاقات الهيكلية بين المتغيرات الكامنة. ويمكن تضمين علاقات أخرى بين المتغيرات على أي مستوى في التسلسل الهرمي. وتكون من نوع "متطلب قبلي Prerequisite of" أو"جزء من part of" أو "مرتبطة بـ Correlated with "، أو "محفزة Induces "، أو مانعة Inhibits " (Almond et al., 2007).

3- **تحديد إطار التقييم المفاهيمي (Conceptual assessment framework):** يتم تحويل هيكل الرسوم البيانية إلى توزيع احتمالي لمتغيرات الكفاية والنتائج التي يمكن ملاحظتها (الشبكة البيزية). ويوضع مخطط التقييم حيث يتم تقسيم الإطار إلى أجزاء تسمى النماذج. ويوفر كل نموذج مواصفات تجيب عن الأسئلة المهمة: ماذا نقيس؟ وكيف نقيسه؟ وكم نقيس؟ وكيف يتم التخصيص؟ والتجميع؟ والعرض؟ (Mislevy et al., 2005). وهي موضحة في جزء مكونات الشبكة البيزية المستخدمة في التقييم.

4- **وضع مخطط الرسم البياني (Graph):** ويحدد العلاقات الشرطية بين المتغيرات، ويتم تحديد قيمة الاحتمال القبلي (a priori) للعلاقات الهيكلية بين المتغيرات، كما يتم الاختيار بين النماذج المرشحة بعد وضع الشبكة البيزية (BN)، ثم تحديد المعلمات لها (أي الاحتمالات الشرطية Conditional probabilities والاحتمالات الهامشية Marginal probabilities) والتوزيعات القبلية (Prior distributions) لمعلمات النموذج (Model parameters)، بهدف استخدام هذه الشبكة للاستدلال. ويمكن استخدام تقديرات الخبراء كاحتمالات أولية، حيث يستعان بالخبراء لبناء مجموعة أولية من المسارات، وتحديد العلاقات بين مفاهيم المجالات الفرعية. ويتم

تقدير مَعلمات النموذج باستخدام طرق التقدير، مع الأخذ في الاعتبار المعلومات السابقة والبيانات التجريبية. وبعد الحصول على المعلمات، يمكن حساب الاحتمال البعدي للسمات المعرفية باستخدام المدخلات من قيم المتغيرات الملاحظة (مثل الاستجابات على المفردات) (Culbertson, 2014; Wang et al., 2023).

5- **التحقق من جودة ومطابقة النموذج (Model fit):** حيث يتم تقدير فترات الثقة باستخدام المحاكاة لاستجابات المفردات من التوزيعات التنبؤية البعدية لنماذج مختارة؛ وتحديد مدى الدقة في استعادة المعلمات (Parameter recovery).

6- **إجراء عمليات "التقدير" (Estimation):** حيث يتم استخراج مَعلمات الأفراد، ومَعلمات المفردات ومَعلمات المسار وإجراء الاستدلال (وضع الدرجات) باستخدام الشبكة البيزية (BN) (Culbertson, 2014).

وتتكون الشبكة البيزية (BNs) من جزءين:

1-الجزء الأول: هو عبارة عن نموذج مخطط رسم بياني (Graphical models) يعكس البنية السببية للمجال ويوضح "التمثيل النوعي للمعرفة" حول كيفية ارتباط مكونات النظام. ويحتوي على رسم بياني دوري موجه (Directed Acyclic Graph- DAG)، ومجموعة مقابلة من التوزيعات الاحتمالية المشروطة (Conditional probability distributions). وتمثل كل عقدة (Node) متغيراً، وتمثل كل حافة (Edge) أو سهم (Arrow) علاقة سببية بين متغيرين.

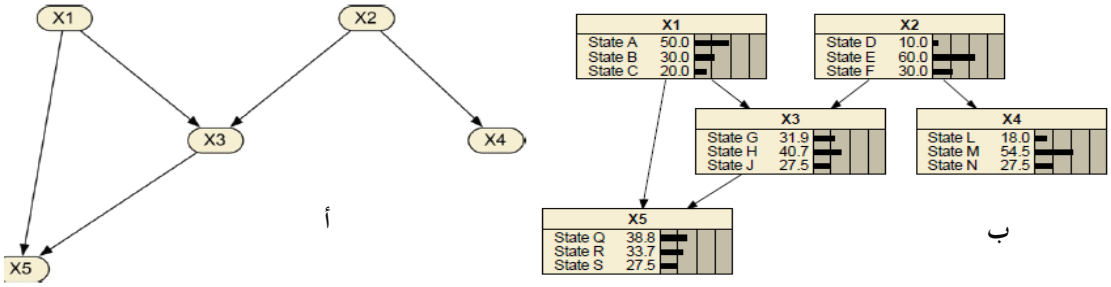
ويمكن الإشارة إلى DAG بالرمز G = (X, E)، حيث يتكون X من متغيرات تسمى العقد و E عبارة عن زوج من العقد المتصلة تمثل العقد الموجودة على DAG متغيرات عشوائية، بينما تشير الحواف أو الأسهم إلى علاقات التبعية الاحتمالية (Probabilistic dependence relationship) بين المتغيرات الممثلة بواسطة العقدتين المرتبطتين، وتدل على وجود تأثير بينها. ويكون خط السهم في اتجاه واحد، ولا يعود إلى العقدة الأصلية. وكل زوج من العقد المتصلة له حافة. والعقدة الأصل (Parent) هي التي تنبثق منها الحافة الموجهة، والعقدة الفرعية (Child) هي التي تتجه لها الحافة. ويحدد التوزيع الاحتمالي المشروط لكل عقدة، اعتمادًا على أصولها (Almond et al., 2015; Chanthira, 2022).

ويمكن ربط زوج من العقد بواسطة الحواف الموجهة أو غير الموجهة. ويكون المتغيران مستقلين بشكل مشروط، إذا لم تكن هناك حافة تربط بينهما، مع الأخذ في الاعتبار المتغيرات الأخرى في الرسم البياني وبالتالي، يمكن وصف العلاقات المعقدة بين عدد من المتغيرات من خلال العلاقات الشرطية بينها (Pearl, 1988; Mislevy et al., 2005; Almond et al., 2007).

2- **الجزء الثاني:** كمي، وهو جداول الاحتمالية الشرطية (Conditional Probability Tables-CPTs) وهي تمثل التبعيات الشرطية التي تحدد قيمة العلاقات بين المتغيرات الممثلة في الشبكة. وكل متغير لديه جدول احتمالية شرطية. ويمثل جدول الاحتمالية الشرطية توزيع الاحتمال القبلي (Prior probability distribution) للمتغير (Mislevy et al., 2005).

وتمثل كل خلية من خلايا الجدول، للمتغير الملاحظ (x)، الاحتمال الشرطي للتواجد في كل حالة من المتغير (x)، بالنظر إلى حالة المتغير الكامن (θ). والقيم الموجودة داخل جداول الاحتمالية هي أيضًا مَعلمات يتم تقديرها من خلال تطبيق الشبكة البيزية (BN) مع الأخذ في الاعتبار المعلومات السابقة والبيانات التجريبية (Uglanova, 2021).

ويوضح الشكل (1) مثالاً لنموذج رسم بياني والشبكة البيزية (BN) المقابلة له، حيث يتكون نموذج الرسم البياني (أ) من خمسة عقد، وخمسة حواف موجهة. وتتوافق الشبكة البيزية (BN) ذات المعلمات الكاملة (ب) مع نموذج الرسم البياني (Mislevy et al., 2005; Mahjoub & Kalti, 2011).



شكل (1): نموذج رسم بياني والشبكة البيزية (BN) المقابلة له (Schultz et al., 2011, p. 4).

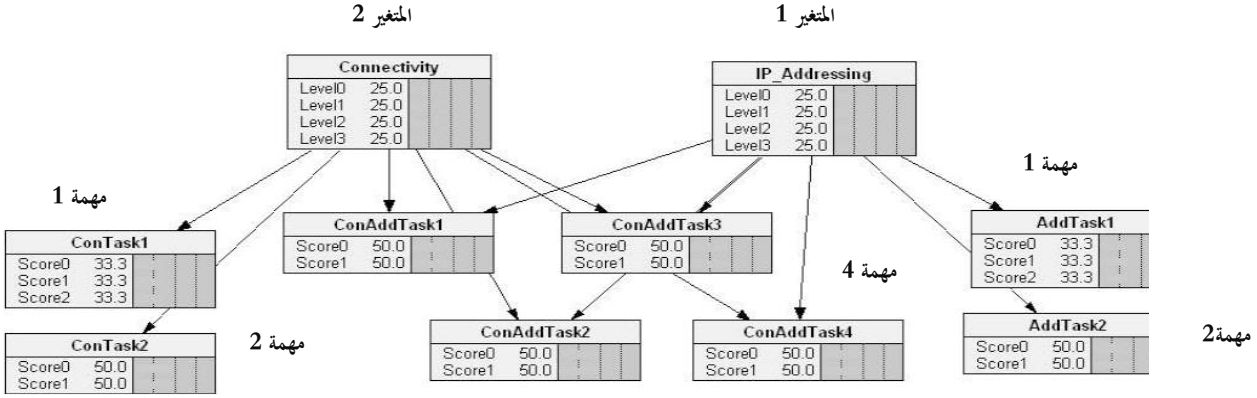
وفي سياق التقييم التربوي، تمثل العقد الكامنة الخصائص المعرفية، على سبيل المثال: مهارات الرياضيات أو التفكير النقدي؛ حيث تمثل العقد الملاحظة سلوك أو أعمال الطلاب، مثل إجابة على مفردة متعددة الخيارات، أو إجراء في مشروع على الحاسب (Uglanova, 2021). وتمثل الحواف التبعيات الشرطية (Conditional dependencies) بين الخصائص الكامنة والإجراءات الملاحظة أو بين درجات الطلاب في التقييم. وعادة ما تكون المتغيرات القابلة للملاحظة ثنائية منفصلة (Dichotomous) (حيث يمثل (0) الإجابة الخاطئة، و (1) الإجابة الصحيحة) أو متعددة البدائل (Polytomous) (حيث تتم إضافة درجات الإجابات الصحيحة جزئيًا).

ويمكن النظر إلى نماذج الاستجابة للمفردة (IRTs) ونماذج التصنيف التشخيصي (DCMs) ونماذج المعادلات البنائية (SEM) على أنها حالات خاصة من الشبكات البيزية (Culbertson, 2014). وعلى النقيض من تلك النماذج المنتشرة على نطاق واسع، فإن الشبكات البيزية (BNs) تعتبر المتغيرات الكامنة أيضًا منفصلة، ويتم وصفها عادةً بالفئات الكامنة (Latent classes)، وتظهر مستويات كفاية مختلفة (Proficiency levels)، على سبيل المثال: منخفضة أو متوسطة أو عالية (Levy, 2009). ويمثل المخطط البياني في نماذج المعادلات الهيكلية العلاقات الدلالية بين المتغيرات، في حين تستخدم الرسوم البيانية مع الشبكات البيزية (BNs) للتعبير عن العلاقات المستقلة المشروطة ضمن تمثيل لتوزيعهم المشترك (Almond et al., 2015; Uglanova, 2021).

وتمثل بنية الشبكة البيزية (BN) التوقعات النظرية لخبراء المجال أو مطوري الاختبار أو علماء القياس النفسي حول الارتباط بين الخصائص المعرفية والسلوك الملاحظ. ويمكن أيضًا تطبيق الأسلوب المبني على البيانات في حالة عدم وجود نظرية قوية أو مصدر للمعلومات الإضافية (Uglanova, 2021).

ولكل متغير (أصل)، مَهمّة ملاحظة تعتمد على كفاية الطلاب. وهناك مجموعة مرتبطة من التوزيعات الاحتمالية الشرطية المقابلة لكل مجموعة ممكنة من قيم المتغيرات (الأصل)، حيث احتمالات الحالات في كل توزيع شرطي مجموعها واحد. وبعد ذلك يمكن استخراج جداول الاحتمالية الشرطية (CPTs)

للمتغيرات الملاحظة. ويوضح الشكل (2) نموذجاً لشبكة بيزية بمتغيرين يجسد كل منهما تقدم تعلم الطلبة من خلال ثمانية متغيرات ملاحظة، والمهام المطلوبة.



شكل (2): نموذج لشبكة بيزية بمتغيرين يجسد كل منهما تقدم تعلم الطلبة من خلال ثمانية متغيرات ملاحظة والمهام المطلوبة (West et al., 2010, p. 13).

كما يوضح الشكل التالي (3) عينة من مخرجات جدول الاحتمالية الشرطية للمتغير الملاحظ (1). وقيمه المحتملة هي (0 و1 و2) باستخدام التقدير الجزئي. ويمثل كل صف توزيع الاحتمال الشرطي لكل مستوى أداء على المهام. ويتضح أن هذه مهمة تميز بشكل أفضل في المستوى (2)، وهو المستوى الذي يوجد فيه احتمال 70٪ على الأقل للحصول على درجة (1) فأكثر. وتؤدي هذه العملية إلى تصميم مخطط رسم بياني للتقييم، يطلق عليه إطار التقييم المفاهيمي.

Connectivity	Score0	Score1	Score2
Level0	85.000	10.000	5.000
Level1	60.000	30.000	10.000
Level2	30.000	50.000	20.000
Level3	10.000	60.000	30.000

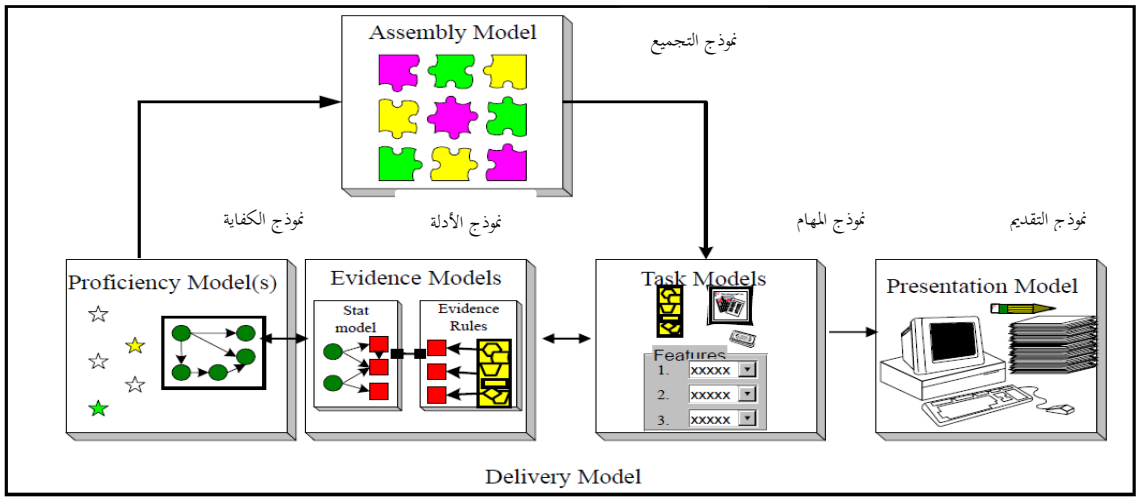
شكل (3): عينة من مخرجات جدول الاحتمالية الشرطية للمتغير الملاحظ (1) لكل مستوى وقيمه المحتملة (West et al., 2010, p. 14).

مكونات الشبكات البيزية (BNs) المستخدمة لنمذجة أنظمة التقييم:

تعمل نمذجة الشبكة البيزية (BNs) على مجموعة واسعة من التقييمات، ولا سيما التقييمات المبتكرة والأكثر تعقيدًا التي تتضمن نماذج كفاية متعددة الأبعاد (Multidimensional proficiency models) وبيئات الأداء التي تتم محاكاتها، وأنواع المهام المتقدمة مع العديد من المفردات. وعادة يستخدم أسلوب التصميم المرتكز على الأدلة (Evidence Centered Design - ECD) ويشتمل على الإطار الاحتمالي لنموذج الشبكة البيزية، ونموذج احتمال وإطار مفاهيمي (Jenkins, 2001; Almond et al., 2007). وتمر العملية بسلسلة من الإجراءات التي تتمحور حول الأسئلة التالية: "ما الذي يمكننا أن نلاحظه بشأن أداء الفرد ويوفر دليلاً على أنه يمتلك/ أو لا يمتلك المعرفة والمهارات والقدرات التي يرغب في إثباتها؟"، و"كيف يمكن هيكلة المواقف للتمكن من تقديم تلك الملاحظات؟"(Culbertson, 2014).

وتنقسم الشبكات البيزية المستخدمة لنمذجة أنظمة التقييم إلى قسمين متداخلين: 1) نموذج الكفاية (Proficiency model) المكون من العُقد التي تمثل كفايات الطلاب المراد قياسها، إلى جانب الروابط الموجهة (Directed links) التي تمثل علاقات معروفة أو مفترضة بين الكفايات؛ و 2) نموذج الأدلة (Evidence model) مع العُقد التي تتوافق مع المهام التي يمكن ملاحظتها، مع روابط من كفايات يتم قياسها مباشرة من خلال ما يمكن ملاحظته. ويوضع مخطط الرسم البياني من الكفايات إلى المهام التي يمكن ملاحظتها، حيث يمكن لمهمة معينة قياس كفايات متعددة (Almond et al., 2007; Almond et al., 2015; Yan et al., 2021).

ويتطلب تحديد إطار التقييم المفاهيمي (Conceptual Assessment Framework- CAF)، باستخدام نموذج الشبكة البيزية (BN) تحويل مخطط الرسوم البيانية إلى توزيع احتمالي لمتغيرات الكفاية والنتائج التي يمكن ملاحظتها (شبكة بيز) ويشكل مخطط التقييم؛ حيث يتم تقسيم الإطار إلى أجزاء تسمى النماذج. ويوفر كل نموذج مواصفات تجيب عن الأسئلة المهمة: "ماذا نقيس؟" و "كيف نقيسه؟" و "كم نقيس؟" و "كيف يتم التخصيص؟"، ويجاب عنها ضمن مكونات نظام تقييم متكامل يطبق في الشبكة البيزية، كما في الشكل التالي (4) (Mislevy et al., 2005; Almond et al, 2015; Yan et al., 2021).



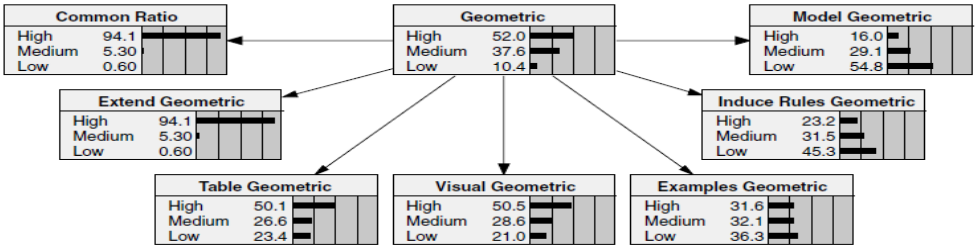
شكل (4): مكونات نماذج تطبيق الشبكة البيزية لأنظمة التقييم (Almond et al., 2015, p. 27).

وفيما يلي توضيح لكل منها:

أ- ماذا نقيس = نموذج الكفاية (Proficiency model):

هناك متغير إتقان واحد يُطلق عليه مَعلمة القدرة أو ثيتا (θ)، وتأخذ القيم (-2، -1، 0، 1، 2) باحتمالات مسبقة وهي (0.1، 0.2، 0.4، 0.2، 0.1). ومتغيرات النتائج القابلة للملاحظة تكون مستقلة عن القدرة (θ). ويتم تصنيف الطلاب حسب قدراتهم، حيث يتم تحديد متغيرات الكفاية، وبنية نموذج الكفاية، مع استخلاص مَعلمات نموذج الكفاية. ويتم بناء نموذج الكفاية على نموذج الانحدار للعقدة الفرع من الأصل. ويحتوي النموذج على متغيرات غير ملاحظة، يُشار إليها بـ (θ)، للفرد i، والتي تصف جوانب المعرفة والمهارة التي هي أهداف الاستدلال في التقييم. ويقوم نموذج الطالب في النماذج الشبكية بتشكيل فهمنا حول i(θ) من حيث التوزيع الاحتمالي. ويُشار إلى المتغيرات لجميع الطلاب (N) في العينة محل الاهتمام بـ (θ).

ويتطلب تحديد نموذج الشبكة البيزية (BN) تحديد عقد الكفاية (Proficiency nodes)، وعدد المستويات أو مجموعة القيم لكل كفاية، وتحديد بنية الارتباط مع العقدة (الأصل). ويحدد لكل عقدة جداول الاحتمالات الشرطية المقابلة، المشروطة بقيم أصول العقدة، ويتم تحديد الملاحظات التي ستستخدم كمهام، وهيكل الارتباط من الكفاية إلى الملاحظة، بشكل يشبه معلومات مصفوفة كيو (*Q*-matrix) التي تحدد المهارات المطلوبة لكل مفردة. وتعتمد معظم تطبيقات التقييم الحالية لنماذج الشبكة البيزية (BN) في البداية على معلومات الخبراء في بنية الكفاية، والروابط من الكفايات للاستجابة على المهمة التي يمكن ملاحظتها. ويتم تعديل الهياكل المعتمدة على معلومات الخبراء بناءً على أدلة وبيانات الاستجابات. وتركز معايرة النموذج عادةً على تقدير معلمات الاحتمال الشرطي لهيكل الرسوم البيانية للشبكة البيزية المحددة، والاحتمالات المشروطة لكل من نماذج الكفاية والأدلة (Jenkins, 2001; Almond et al., 2007)، كما في الشكل (5).



شكل (5): نموذج كفاية قبل المعايرة ببيانات الاستجابات (Almond et al., 2007, p. 354).

ب-كيف نقيس = نموذج الأدلة / المهام (Evidence model): تؤدي المهام إلى عمل منتج يمكن إعطاؤه درجة بشكل لا لبس فيه على أنه صحيح/خطأ. وكل مهمة لها متغير نتيجة واحد يمكن ملاحظته، وتسمى المهام بالمفردات (Almond et al., 2015; Yan et al., 2021). ويمكن ترتيب كل من القيم الخمس للقدرة (θ) للحصول على قيم جداول الاحتمال الشرطية (CPT) كما في الجدول التالي في الشكل (6).

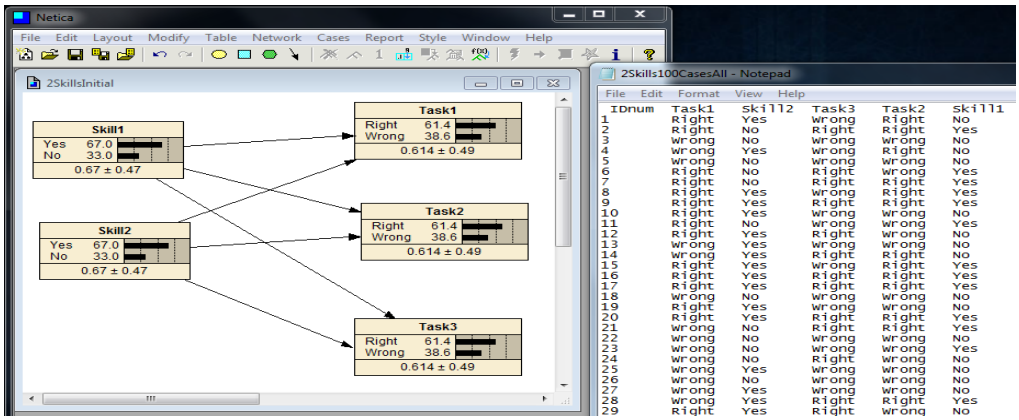
θ	Prior	Item 1	Item 2	Item 3	Item 4	Item 5
-2	0.1	0.3775	0.2227	0.1192	0.0601	0.0293
-1	0.2	0.6225	0.4378	0.2689	0.1480	0.0759
0	0.4	0.8176	0.6792	0.5000	0.3208	0.1824
1	0.2	0.9241	0.8520	0.7311	0.5622	0.3775
2	0.1	0.9707	0.9399	0.8088	0.7773	0.6225

شكل (6): قيم جداول الاحتمالية الشرطية (CPT) لمجموعة مفردات حسب مستوى القدرة (θ) (Almond et al., 2007, p. 160).

ويتم وضع المهام اعتمادًا على الكفايات المستهدفة ذات الصعوبة المنخفضة/ المتوسطة/ العالية. كما يتم بناء نماذج الأدلة على أساس مستوى الصعوبة وباستخدم مصفوفة كيو (Q) لتحديد العقدة الأصلية، وهي تفترض أن الأشياء القابلة للملاحظة مستقلة بشكل مشروط بين المهام، ولكن يمكن أن تعتمد على المهام،

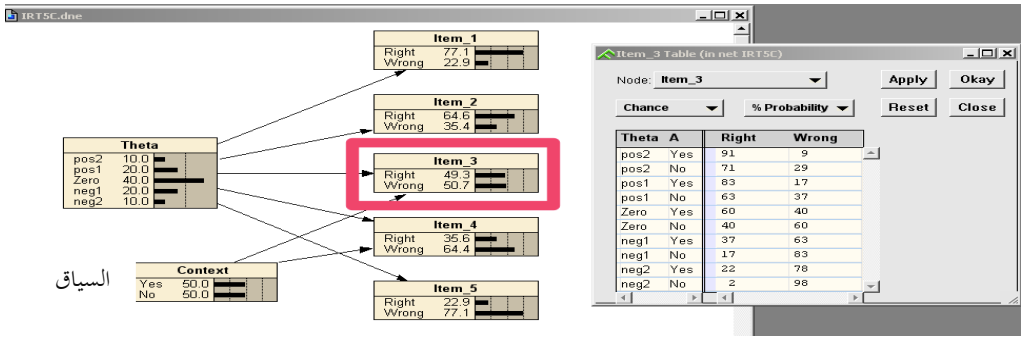
ويُشار إلى المهمة بالحرف (j). ويصف نموذج الأدلة كيفية استخلاص الأدلة مما يقوله الطالب أو يفعله في سياق المهمة (العمل المنتج). وينتج عن الأدلة قيم المتغيرات التي يمكن ملاحظتها. أما نموذج المهمة فيصف ميزات المهمة التي يلزم تحديدها. ويتضمن المواصفات الخاصة ببيئة العمل، والأدوات التي يستخدمها المفحوص ومنتجات العمل والمواد التحفيزية، والتفاعلات بين المفحوص والمهمة، بما يتفق مع متطلبات الإثبات لنموذج الأدلة. ويتم التعبير عن خصائص المهمة بواسطة متغيرات نموذج المهمة التي تم تحديدها من قبل مطور الاختبار أو الخبراء. ويتم تحديد مواصفات المهام الموجودة في المفردات وقيم الاحتمالية لكل منها (Almond et al., 2015).

ج-كم نقيس = نموذج التجميع (Assembly model): توضع المفردات وترتب حسب الصعوبة، ويتم وصف المهام التي تدخل في التقييم في اختبار ثابت أو ديناميكي باستخدام البرنامج المناسب (Mislevy et al., 2000; Almond et al., 2015)، كما هو موضح في الشكل (7).



شكل (7): المفردات والمهام التي تدخل في التقييم باستخدام أحد برامج الشبكة البيزية (نيتكا) (Yan et al., 2021).

د-كيف يتم التخصيص = نماذج العرض والتطبيق (Presentation & delivery models): يتم إضافة متغير السياق، مثل جمع المفردات التي تشترك في مهمة واحدة مع متغيرات نواتج يمكن ملاحظتهما، ويضاف متغير كامن خاص بالفرد ومهمة معينة تسمى "السياق" وبقيم احتمالية (Mislevy et al., 2000; Almond et al., 2015; Yan et al., 2021). والشكل (8) يوضح ذلك.



شكل (8): مثال جمع للمفردات (رقم 3 و4) في مهمة واحدة مع متغيري نواتج يمكن ملاحظتهما ومتغير كامن للفرد ومهمة معينة تسمى "السياق" (Yan et al., 2021).

رابعاً: مؤشرات مطابقة النموذج وطرق التقدير والمعايرة للشبكات البيزية (BN):

هناك عدد من المؤشرات المختلفة لمطابقة النموذج في الشبكات البيزية (BNs) المستخدمة في التقييم التربوي، ومنها ما يلي: 1) متوسط الفرق المطلق (Mean absolute difference) بين درجات الاختبار المتوقعة والدرجات الفعلية، حيث يتم تقدير درجة الاختبار المتوقعة من خلال الصدق التقاطعي بترك حالة واحدة (Leave-one-out cross validation)، واقترح هذا المؤشر باردوس وزملاؤه (Pardos et al., 2007)؛ 2) مؤشر درجة قود اللوغاريتمية (Good's Logarithmic score) الذي وضعه ويفر (Weaver, 1948)؛ و3) مؤشر درجة الاحتمالية الرتبية (Ranked probability score) (Epstein, 1969). وقام ويليامسون وزملاؤه (Williamson et al., 2000) بدراسة أداء هذه المؤشرات الثلاثة ووجدوا أن مؤشر درجة الاحتمالية الرتبية ومؤشر درجة قود هما الأفضل في اكتشاف أخطاء الرسوم في الشبكة البيزية (Culbertson, 2014).

ويتم "تقدير" (Estimation) هذه النماذج الشبكية البيزية (BNs) عادةً باستخدام خوارزمية تعظيم التوقعات (Expectation-Maximization Algorithm-EM) (Dempster et al., 1977) أو سلسلة ماركوف مونتي كارلو (Markov Chain Monte Carlo-MCMC) (Gilks et al., 1996)، أو خوارزمية متروبوليس-هاستينج روبنز-مونرو، المقترحة مؤخرًا (Metropolis-Hasting Robbins-Monro Algorithm-MH-RM) (Cai, 2010) التي نجحت في تقليل وقت الحساب مقارنة بالطرق الأخرى (Almond et al., 2015; Uglanova, 2021).

وفي إطار الشبكات البيزية، يستخدم التحقق من النموذج التنبؤي البعدي (Posterior Predictive Model Checking -PPMC) (Rubin, 1984) حيث تُقارن الملاحظة (الفعلية) والنموذج المتوقع. وتتم محاكاة الاستجابات للأفراد من النموذج المقترح باستخدام مجموعات من المَعلمات المستمدة من التوزيع البعدي. ويتم بعد ذلك حساب إحصائية التناقض (Discrepancy statistic) لكل مجموعة من مجموعات بيانات المحاكاة والبيانات الفعلية، ويتم الحصول على قيمة (P) التنبؤية البعدية من نسبة المجموعات المكررة التي يتجاوز فيها إحصاء التناقض البيانات الفعلية. وتشير قيمة (P) الصغيرة إلى عدم ملاءمة النموذج (Sinharay, 2006; Culbertson, 2014). كما يتم التحقق من مدى الارتباط مع المحكات الخارجية (Correlation with external criteria)، ومن صدق الاستنتاجات حول الطلاب، التي تم جمعها من خلال تطبيق الشبكة البيزية كنموذج للقياس (Uglanova, 2021).

ويمكن مقارنة تقديرات الاحتمالية (Likelihood estimates) مع النماذج بتطبيق محكات المعلومات (Information criteria)، مثل محك أكيك للمعلومات (Akaike Information Criterion-AIC) (Akaike, 1973) أو محك المعلومات البيزي (Bayesian Information Criterion- BIC) (Schwarz, 1978)، أو محك معلومات الانحراف (Deviance Information Criterion- DIC) (Spiegelhalter et al., 2002) لاختيار النموذج، كما يستخدم مؤشر الاتساق الهرمي (Hierarchy Consistency Index (HCI)) (Crawford, 2014).

وبالإضافة إلى المقاييس الكمية لمطابقة النموذج، يمكن أن توفر الرسوم التشخيصية (Diagnostic plots) توضيحًا لمدى مطابقة الشبكة البيزية للبيانات (Sinharay, 2006). وعند عرض البيانات، يتم رسم مصفوفة لاستجابات المفردات الثنائية أو المتدرجة. ويرسم عدد من مصفوفات المفردات بالمحاكاة من التوزيع البعدي للمعلمات، إلى جانب البيانات الفعلية، ويتم فحص الأنماط بحثًا عن الانحرافات. بالإضافة إلى ذلك، فإن مطابقة المنحنيات المميزة للمفردات (Item

characteristic curves) أو رسوم البواقي البيزية (Bayesian residuals) للمفردات أو الاختبار مقابل القدرات المقدرة، يمكن أن تحدد ما إذا كانت هناك حالات انحراف عن البيانات التي تنبأ بها النموذج المقترح. كما يمكن رسم التوزيعات القبلية مقابل التوزيعات البعدية لمَعلمات المفردات، مما يساعد في تحديد الاختلافات.

ويمكن تحديد الأفراد غير المطابقين، وتحليل أداء المفردات، بناءً على فحص جداول الاحتمالية الشرطية (CPTs). ويصمم مستوى صعوبة الاختبار لكل فرد، حيث يتم دمج المعلومات في التقييم، مثل محتوى المفردة، ومتغيرات نموذج المهمة. وينتهي الاختبار عند الوصول إلى دقة القياس المطلوبة أو العدد المحدد من المفردات. كما يتم اختبار ما إذا كانت المفردات تميز بين الطلاب من ذوي مستويات الكفاية المختلفة، ويتم تحليل البواقي لاختيار المفردات. كما يتم الحصول على تقديرات مَعلمات المفردات، ويستخدم تقدير نموذج بيز الشرطي وأقصى الاحتمالية Maximum likelihood (Mislevy et al., 2000; Culbertson, 2014)

خامساً: أهم البرامج الإحصائية لتحليلات الشبكات البيزية (BN):

هناك عدد من البرامج التي تستخدم عالميًا مع الشبكات البيزية (BNs) حيث أصبح هناك حزم برامج حاسوبية تجارية وأخرى مفتوحة المصدر، ومكتبات تسمح باستخدام هذه التحليلات وبطريقة مبسطة نسبيًا، ومن أهم هذه البرامج الإحصائية ما يلي:

1-برنامج "نيتكا Netica" (Norsys Software Crop): وهو برنامج يستحدم لتحليلات الشبكة البيزية (BN)، وبه أدوات تعليمية. ويستخدم للتشخيص والمحاكاة في مجالات التمويل والبيئة والطب والصناعة والتربية وغيرها من المجالات، والموقع الخاص به على الرابط (http://www.norsys.com). وهو يستخدم خوارزمية استدلال (junction tree). ويوفر واجهة رسوم بيانية لسهولة التشغيل، ويمكنه بسهولة تحويل الشبكة البيزية (BN) واستكشاف العلاقات بين المتغيرات في نموذج يقوم البرنامج بإنشائه من خلال ما يزود به من بيانات. ويتمتع هذا البرنامج بانتشار كبير في العالم نظرًا لبساطته وموثوقيته وجودة أدائه.

2- برنامج "آر R": ويستخدم لغة آر، والحزمة المستخدمة هي (bnlearn)، وتحتوي على خوارزميات مختلفة لتعلم بنية الشبكة البيزية (BN) وتقدير المَعلمات والاستدلال البيزي. والموقع الخاص به على الرابط (http://www.r-project.org).

3- برنامج "آر نيتكا RNetica" (Netica API for R): ويجمع بين لغة البرمجة الإحصائية مفتوحة المصدر (R) ومحرك الشبكة البيزية "نيتكا"، والغرض الأساسي منه هو توفير ربط لوظائف Netica API داخل (R) لإعداد البيانات وتنقيحها. وتتيح واجهة البرنامج استخدام الرسوم البيانية لتطوير النماذج. إضافة لسهولة البرمجة باستخدام هذا البرنامج، مما يجعل العمل عليه أسرع بكثير من استخدام واجهات لغات برمجية أخرى. والموقع الخاص به موجود على الرابط (http://pluto.coe.fsu.edu/RNetica).

4-برنامج "بقز BUGS": للاستدلال البيزي باستخدام عينات جيبس (Bayesian Inference Using Gibbs Sampling- BUGS) . و "بقز" عبارة عن حزمة برامج للتحليل البيزي للنماذج الإحصائية باستخدام سلسلة ماركوف مونت كارلو (MCMC). ويعتمد على أخذ عينات جيبس (Gibbs) لتحليل النموذج المحدد. وهناك إصداران رئيسان للبرنامج وهما وينبقز (WinBUGS) والموقع الخاص به على الرابط (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml) و "أوبنبقز" (OpenBUGS) والموقع الخاص به على الرابط (http://mathstat.helsinki.fi/openbug). ويمكن استخدام الأخير حيث إنه برنامج للشبكات البيزية،

ويحتوي على أوامر محدودة مشابهة لتلك الموجودة في برنامج ماتلاب (Matlab) بدون عرض رسوم شبكة بيز.

5-برنامج "جافا بيز JavaBayes": وهوعبارة عن مجموعة من أدوات جافا (Java) التي تعمل على إنشاء شبكات بيزية (BNs) ومعالجتها. ويشتمل على محرر رسوم بيانية، حيث يسمح محرر الرسوم بإنشاء وتعديل الشبكات البيزية. كما يتضمن خوارزميات للاستدلال (junction tree). وموقعه (https://www.cs.cmu.edu/~javabayes/Home).

6- برنامج "هيوجن Hugin": وهو واحد من أهم الأدوات المستخدمة في الشبكات البيزية (BNs). وهو برنامج تجاري، وكان أحد الحزم الأولى للنموذج داج (DAG). ويوفر بيئة رسوم بيانية تساعد في تحديد وبناء أسس المعرفة القائمة على الشبكات البيزية، كما يدعم خوارزمية الاستدلال (junction tree). وموقعه على الرابط (http://www.hugin.com).

7- برنامج "بيز نت تولبوكس BAYES NET TOOLBOX " (BNT) : وهو مكتبة مفتوحة المصدر تستخدم ماتلاب 2 (Matlab2)، وهي مدعومة من قبل عدد من الباحثين، وتشمل عدة خوارزميات متكاملة للاستدلال والتعلم الشبكي البيزي. ويستخدم أقصى الاحتمالية (Maximum likelihood) أو الأقصى البعدي (Maximum a posteriori) لتقدير المعلمات من البيانات الكاملة، و خوارزمية تعظيم التوقعات (EM) للبيانات غير المكتملة. كما يستخدم البرنامج عددًا من محكات مطابقة النموذج. والموقع له على الرابط (https://bayesnet.github.io/bnt/docs/usage.html).

8- برنامج "بيزيالاب BAYESIALAB": وهو أحد منتجات"بيزيا Bayesia " وهي شركة فرنسية مخصصة لاستخدام أساليب دعم القرار والتعلم من الذكاء الاصطناعي وتطبيقاتها في عدة مجالات، وهو بمثابة مختبر كامل لدراسة الشبكات البيزية (www.bayesia.com).

9-برنامج "جاقز JAGS": حيث يستخدم عينات جيبس (Just Another Gibbs Sampler) والموقع الخاص به (https://sourceforge.net/projects/mcmc-jags/).

10- برنامج "إف بي إم FBM ": وهو للنمذجة المرنة (Flexible Bayesian Modeling) والموقع الخاص به (http://www.cs.utoronto.ca/~radford/fbm.software.html)

11-برنامج "ستان Stan": ويعتمد على لغة ستان ويستخدم عينات (No-U-Turn)، وموقعه هو (http://mc-stan.org/Stan).

12-برنامج "إس بي إس إس موديلر SPSS Modeler": وهو برنامج تجاري يوفر شبكات بيزية. ويتوفر على الرابط (https://www.ibm.com/docs/en/cloud-paks/cp-data/4.8.x?topic=modeling-bayes-net-node). (Murphy, 2007; Korb & Nicholson, 2011; Mahjoub & Kalti, 2011; Almond et al., 2015).

سادساً: تطبيقات الشبكات البيزية (BN) في التقييم التربوي:

يمكن تطبيق الشبكات البيزية (BNs) في سياق التقييم التربوي للتنبؤ بمستوى كفايات الطلاب، وتصنيفهم بشكل صحيح حسب مستويات كفاياتهم. ومن أهم التطبيقات في مجال التقييم التربوي الشبكات البيزية المبنية على الأهداف (Object-oriented Bayes nets) والشبكات البيزية الديناميكية (Dynamic Bayes nets)، والتي تدعم أنظمة التقييم من أجل التعلم، والشبكات البيزية المبنية على المعرفة (Knowledge-based Bayes nets) ، والتقييم المرتكز على الأدلة (ECD) لربط التقييم بنظريات التعلم في هذا المجال، وهي أدوات لمساعدة مصممي التقييم للاستفادة من الشبكات البيزية (Mislevy et al., 2005).

وهناك بعض الدراسات التي طبقت الشبكات البيزية (BNs) للتقييم التربوي، ومن أهمها دراسة ميسلفي وزملائه (Mislevy et al., 2000) لتطوير نموذج شبكة بيزية (BN) في مجال الفيزياء التطبيقية بالجامعة، بهدف اختيار المفردات من مستويات مختلفة من المعلومات لمجالات فرعية مختلفة؛ حيث قاموا بدراسة خمسة عوامل وهي: حجم الشبكة، وعلاقات المجالات الفرعية، وعدد المفردات، وعدد المفحوصين، ومحك اختيار المفردة، وبنية معالم المفردة. وتكونت البيانات من استجابات (1215) طالبًا في جامعة إلينوي، على أسئلة ثنائية التدريج. وبمساعدة بعض خبراء الفيزياء، تم تحليل محتوى مفردات الاختبار وعددها (76) مفردة، وتحديد (13) مجالًا فرعيًا. وتم التحقق من استعادة مَعلمات النموذج للأفراد والمفردات من خلال المحاكاة، ثم وضع النموذج في سياق اختبار تكيفي محوسب (Computerized Adaptive Test-CAT) باستخدام نموذج شبكة بيزية لتحسين اختيار المفردات. وتم التحقق من أداء محك اختيار المفردات القائم على المعلومات لمختلف المجالات الفرعية. تم تطوير النموذج اعتمادًا على مرئيات الخبراء، كذلك تم تحديد العلاقات والنموذج الخطي الذي يربط المجالات الفرعية. وخلصت الدراسة إلى أن دقة معلمات الفرد والمسار تعتمد على مكان المعلمة في الشبكة، حيث استفادت المجالات الفرعية ذات الفروع العديدة من المعلومات الإضافية المقدمة من المفردات الخاصة بالمجالات الفرعية التابعة لها، مع زيادة قوة العلاقات بينها. واحتاجت المجالات الفرعية التي لديها عدد كبير من الأصول، وعدد قليل من الفروع أو لا توجد لها فروع، لعدد أكبر من المفردات لتحقيق نفس دقة القياس. كما أوضحت الدراسة أنه يمكن استعادة المَعلمات باستخدام اختبارات قصيرة (تصل إلى 3 مفردات لكل مجال فرعي)، وأحجام عينات صغيرة (تصل إلى 300 فرداً)، بالرغم من أن العينات الأكبر مطلوبة للحصول على مَعلمات مفردات وأفراد دقيقة. وانخفضت دقة التقدير مع زيادة حجم العينة، كما أدى زيادة عدد المفردات لكل بُعد إلى تحسين طفيف في دقة مَعلمات المفردات، ويرجع ذلك على الأرجح إلى التقديرات المحسنة لمعلمات الأفراد (Culbertson, 2014).

كما تضمنت دراسة ألموند وزملائه (Almond et al., 2007) تطبيق نموذج الشبكة البيزية (BN) على نماذج التقييم التشخيصي المعرفي (Cognitive diagnostic modeling assessment)، والاستدلال القائم على الأدلة، باستخدام مجموعات بيانات حقيقية ومحاكاة، وتطويرها ومعايرتها من البيانات، للوصول لاستنتاجات مبنية على الأدلة حول مستويات الكفايات للأفراد والمجموعات (Almond et al., 2007). واستخدموا التقييم المبني على الدليل (ECD) من خلال عدة خطوات وهي: 1) استخدام وسيلة تعليمية إرشادية لاختيار مجال تعليمي للاختبار، 2) تحديد جميع المهام المناسبة للمجال التعليمي المحدد، 3) حساب الوزن المتوقع للأدلة الخاصة بكل مهمة من المهام المحددة، 4)اختيار مهمة محددة ذات وزن أعلى للأدلة، 5) تطبيق المهمة، وجمع وتسجيل إجابات الطلاب، 6) تحديث نموذج كفاية الطالب، و7) التوقف أو العودة إلى الخطوة الأولى والتكرار. وكانت درجات الشبكة البيزية عالية الثبات حيث بلغت (0.88) وقدمت صدقًا مضافًا (Incremental validity) لدرجة الاختبار القبلي في التنبؤ بنتيجة الاختبار البعدي. بالإضافة إلى ذلك أدى النموذج بالوزن المتوقع للأدلة إلى زيادة تعلم الطلاب.

كما استخدمت دراسة موسوي وماكجيني (Moussavi & McGinny, 2009) الشبكة البيزية (BN) لتقييم الأداء المدرسي، وتشخيص أسباب انخفاض فعالية بعض المدارس، والمساعدة في صنع القرار. وطبقت الدراسة على عينة من بوتسوانا مكونة من (3322) طالبًا وطالبة في الصف السادس و(170)

معلمًا ومديرًا (أو ممثليهم) في (170) مدرسة. وقام الباحثون بوصف نواتج التعلم، وتم وضع محددات تحصيل الطلاب وفعالية المدرسة والظروف والعمليات داخل المدارس، واحتمالية أن يؤثر تدخل معين على تلك الظروف والعمليات ومن ثم النواتج. وتم وضع (14) متغيرًا في النموذج، ثم دُمج نموذج الطالب البيزي في نظام للاختبارات المحوسبة، ضمن مشروع تعليم للرياضيات. وتم تقييم هذا النموذج في البداية باستخدام بيانات محاكاة. ثم تم التأكد من قدرة النموذج على التشخيص من خلال بيانات (152) طالبًا، حيث أجرى كل منهم اختبارًا محوسبًا واختبارًا كتابيًا، صممت لقياس معرفة الطلاب في (12) مفهومًا. وتم التقييم من قبل ثلاثة خبراء للاختبار الكتابي، بينما استخدم النموذج البيزي للحصول على تقديرات معرفة كل طالب بالمفاهيم، وتم حساب الاتفاق بين التقييم لكل منهما. وأظهرت النتائج درجة عالية من الاتفاق بين الدرجات التي قدمها الخبراء وبين التشخيص الذي قدمته النماذج البيزية في الامتحان الكتابي.

كما طبقت دراسة تينجر وألموند (Tingir & Almond, 2017) نموذج الشبكة البيزية (BN) على بيانات من برنامج تقييم الطلاب الدوليين (Program for International Student Assessment -PISA) لعام (2012) لبيانات الرياضيات. وتم فحص العلاقة بين مفهوم الذات، والكفاءة الذاتية، والاهتمام بالرياضيات، والإنجاز في الرياضيات لكل من الولايات المتحدة وتركيا. واستخدمت الشبكات البيزية لوصف العلاقة المتبادلة بين هذه العوامل، ودرجة التحصيل في الرياضيات للطلاب من البلدين. ومن خلال بناء شبكات بيزية منفصلة لكل بلد، للتمكن من إجراء استنتاجات حول المتغيرات غير الملاحظة بناءً على المتغيرات الملاحظة، أمكن التنبؤ بدرجة التحصيل للمشاركين في برنامج التقييم الدولي للطلاب (PISA) من خلال اتجاهاتهم نحو الرياضيات، والكفاءة الذاتية ومفهوم الذات والاهتمام بالرياضيات، وأثبت النموذج دقة عالية في التنبؤ.

سابعاً: التعقيب والخلاصة والتوصيات:

يتضح من خلال ماتم استعراضه من مفاهيم ونظريات وإجراءات، أهمية الشبكات البيزية (BNs) للقياس والتقييم التربوي، حيث إنها توفر إطارًا لنمذجة العلاقات المعقدة؛ وبالتالي فهي الأنسب لنمذجة تعلم الطلاب وتقييمه ومتابعة تقدمهم والتنبؤ بسلوكهم وتقديم الدعم اللازم لهم وبشكل فردي، حيث تربط الشبكات البيزية (BNs) بين نظرية الاحتمالات ونماذج الرسم لتمثيل العلاقات الاحتمالية بين عدد كبير من المتغيرات. ورغم أن الشبكات البيزية (BNs) تتمتع بالقدرة على التعامل مع تعقيد نماذج القياس إلا أن استخدامها في مجال القياس والتقييم النفسي والتربوي يعتبر قليلاً مقارنة بالمجالات الأخرى ذات العلاقة.

ومن خلال استعراض الدراسات التي استخدمت الشبكات البيزية (BNs) تتضح محدودية تطبيقها في مجال القياس النفسي والتقييم التربوي. وبالرغم من التقدم النظري لاستخدامها في هذا المجال، إلا أن تطبيقها كنموذج قياس جديد ما يزال محدودًا، وبشكل أخص في العالم العربي. وفي حين أن مختصي القياس النفسي على دراية جيدة بتقنيات تطوير وتحليل الاختبارات بناء على نظرية الاستجابة للمفردة (IRT) ونماذج التصنيف التشخيصي (DCM)، وغيرها من الأساليب المتقدمة، إلا أنه لا يُعرف حاليًا سوى القليل جدًا عن خصائص القياس للاختبارات باستخدام نماذج الشبكات البيزية (BN). ودراسات القياس النفسي التي طبقت هذه الشبكات كنموذج قياس واستخدمتها في التقييم التربوي والاختبارات التكيفية المحوسبة (CAT) محدودة للغاية، واقتصرت على دراسة كل من ميسلفي وزملائه (Mislevy et al., 2000) ودراسة ألموند وزملائه (Almond et al., 2007)

ودراسة موسوي وماكجيني (Moussavi & McGinny, 2009) ودراسة تينجر وألموند (Tingir & Almond, 2017)، وهم رواد القياس في استخدام الشبكات البيزية للتقييم التربوي. وبعض الدراسات ركزت على مقارنة النماذج، وجودة الاستدلال حول الطلاب، وملاءمة المفردات، والملاءمة الكلية، وبعضها على تحليل ملاءمة الأفراد (Mislevy et al., 2002; Crawford, 2014; Almond et al., 2015; Levy & Mislevy, 2016).

كما أن تحليل المهام المعرفية وتحديد المتغيرات الملاحظة والكامنة، لمهام معقدة ذات أبعاد كثيرة ما يزال يطرح تحديات تطبيقية. وهناك عدد قليل من الدراسات التي أجريت على الأداء التفاضلي للمفردة (Differential Item Function-DIF) ومطابقة الأفراد، مما يستدعي إجراء المزيد من البحوث المنهجية في هذه المجالات (Almond et al., 2015; Uglanova, 2021). وهناك امتدادات واعدة مستقبليًا للشبكات البيزية (BN)، في القياس النفسي الحديث مثل الشبكة البيزية الديناميكية (DBN)، التي تسمح بإجراء تحليل مع الأخذ في الاعتبار التبعيات بين المجالات الزمنية، وتتابع محاولات الطلاب لحل المهمة وتقدم تغذية راجعة فورية أثناء إجراء الاختبار المجالات وقد تم تطبيقها في التقييم التربوي (Reye, 2004; Almond et al., 2015; Levy, 2019; Xing et al., 2020; Uglanova, 2021). كما أن هناك نماذج هجينة (Hybridized models)، تتكون من كل من نموذج نظرية الاستجابة للمفردة متعددة الأبعاد (MIRT) ونموذج الشبكة البيزية (BN) للحصول على معلومات جديدة من أنظمة التقييم وإجراء استنتاجات أكثر دقة حول الطلاب (Mislevy et al., 2002; Uglanova, 2021).

وقد أشار ديكليرك وزملاؤه (De Klerk et al., 2015) إلى أن الشبكة البيزية تعتبر الإطار الأكثر استخدامًا على نطاق واسع لتحليل القياس النفسي في مجال التقييم القائم على المحاكاة (Simulation-based assessment) والتقييم القائم على الألعاب (Game-based assessment). وعادةً ما تكون الخصائص الكامنة، التي يتم تقييمها من خلال هذين النوعيين ذات بنية معقدة للغاية؛ وبالتالي، يجب أن يكون نموذج القياس مرنًا بدرجة كافية ليأخذ في الاعتبار نظام العلاقات المعقدة بين المهارات وأفعال الطلاب. لذلك، فإن أساليب النمذجة النفسية المبنية على تطبيق الشبكات البيزية مفيدة لها (Xing et al., 2020).

ويخلص من كل ماسبق إلى التوصيات التالية:

1. أهمية توفير إطار نظري للشبكات البيزية وتطبيقاتها في مجال القياس والتقييم النفسي والتربوي باللغة العربية للمختصين والطلبة حتى يتمكنوا من مسايرة التقدم العالمي في هذا المجال.
2. حث المختصين وطلبة الدراسات العليا على تناول النماذج الشبكية البيزية وامتداداتها في أبحاثهم، ومنها دراسة الأداء التفاضلي للمفردات، ودقة وحساسية مؤشرات المطابقة تحت الظروف المختلفة (مثل: طبيعة مجال التعلم، عدد المفردات، وعدد الأفراد).
3. إجراء المزيد من الدراسات على الخصائص السيكومترية لنماذج الشبكات البيزية عند استخدامها في سياقات القياس والتقييم النفسي والتربوي.
4. تدريب المختصين على استخدام البرامج الحاسوبية الواعدة في مجال الشبكات البيزية، مما يساعد على التطبيق وإجراء الأبحاث وإصدار القرارات الدقيقة.
5. ضرورة تبني مراكز القياس والتقييم النفسي والتربوي بالعالم العربي لهذه النماذج عند بناء أدوات التقييم بغرض التشخيص والتنبؤ ومتابعة وتحسين التعلم، وعند تطبيق أساليب

التعلم والتقييم بالألعاب الإلكترونية، والاختبارات التكيفية لفعاليتها ومرونتها العالية.

6. إجراء المزيد من الأبحاث لربط الشبكات البيزية بالذكاء الاصطناعي وتطبيقاته في مجال تقييم تعلم الطلبة ومتابعته.

المراجع العربية

علام، صلاح الدين محمود (2005). *نماذج الاستجابة للمفردات الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي* (ط1). دار الفكر العربي.

المراجع الأجنبية

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & B. F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Academiai Kiado.

Almond, R., DiBello, L., Moulder, B., & Zapata-Rivera, J. (2007). Modeling diagnostic assessment with Bayesian networks. *Journal of Educational Measurement*, 44(4), 341-359.

Almond, R., Mislevy, R., Steinberg, L., Williamson, D., & Yan, D. (2015). *Bayesian networks in educational assessment*. Springer.

Allam, S. (2005). *Unidimensional and multidimensional item response theory models and their applications in psychological and educational measurements*. Daralfikr Alarabi.

Bekele. R. & Menzel, W. (2005, February). *A Bayesian approach to predict performance of a student (BAPPS): A case with Ethiopian students* [Paper presentation]. IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria.
http://www.actapress.com/Abstract.aspx?paperId=18917

Cai, L. (2010). Metropolis-Hastings Robbins-Monro Algorithm for confirmatory item factor analysis. *Journal of Educational and Behavioral Statistics, 35*, 307–335. doi: 10.3102/1076998609353115

Chanthiran, M., Ibrahim, A., Rahman, M., & Mariappan, P. (2022) Bayesian network approach in education: A bibliometric review using R-tool and future research directions. *The Eurasia Proceedings of Educational & Social Sciences (EPESS), 25*, 17-25. https://doi.org/10.55549/epess.1191900

Crawford, A. (2014). *Posterior predictive model checking in Bayesian networks* (Unpublished Doctoral Dissertation). Arizona State University.

Cruz, N., Desai, S., Dewitt, S., Hahn, U., Lagnado, D., Liefgreen, A., Phillips, K., Pilditch, T., & Tešić, M. (2020). Widening access to Bayesian problem solving. *Frontiers in Psychology, 11*, 1-8. https://doi.org/10.3389/fpsyg.2020.00660

Culbertson, M. J. (2014). *Graphical models for student knowledge: Networks, parameters, and item selection* (Unpublished Doctoral Dissertation). University of Illinois Urbana-Champaign. https://www.ideals.illinois.edu/handle/2142/49372

Culbertson, M. J. (2016). Bayesian networks in educational assessment: The state of the field. *Applied Psychological Measurement, 40*(1), 3–21. https://doi.org/10.1177/0146621615590401

De Klerk, S., Veldkamp, B., & Eggen, T. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education, 85*, 23–34. https://doi.org/10.1016/j.compedu.2014.12.020

Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B, 39*, 1–38.

Desmarais, M. & Gagnon, M. (2006, October). *Bayesian student models based on item to item knowledge structures* [Paper presentation]. First European Conference on Technology Enhanced Learning, Crete, Greece. doi: 10.1007/11876663_11

Epstein, E. S. (1969). A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology, 8*(6), 985–987. doi: 10.1175/1520-0450(1969)008<0985:ASSFPF>2.0.CO;2

Gilks, W., Richardson, S., & Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in practice*. Chapman & Hall.

Jenkins, F., Mislevy, R., Senturk, D., Steinberg, & Yan, D. (2001). Models for conditional probability tables in educational assessment. In T. Jaakkola & T. Richardson (Eds.), *Artificial Intelligence and Statistics* (pp. 137-143). Morgan Kaufmann.

Korb, K., & Nicholson, A. (2011). *Bayesian artificial intelligence* (2nd ed.). CRC Press.

Levy, R. (2009). The rise of Markov Chain Monte Carlo estimation for psychometric modeling. *Journal of Probability and Statistics.* Article 537139. doi:10.1155/2009/537139

Levy, R. (2019). Dynamic Bayesian network modeling of game-based diagnostic assessments. *Multivariate Behavioral Research, 54* (6), 771-

Schultz, M., Borrowman, D., & Small, M. (2011). *Bayesian networks for modeling dredging decisions*. U.S. Army Corps of Engineers.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Sinharay, S. (2006). Model diagnostics for Bayesian networks. *Journal of Educational and Behavioral Statistics, 31*, 1–33. doi: 10.3102/10769986031001001

Spiegelhalter, D., Best, N., Carlin, B., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B, 64*, 583–639. doi: 10.1111/1467-9868.00353

Tingir, S. & Almond, R. (2017). Using Bayesian networks to visually compare the countries: An example from PISA. *Journal of Education & Social Policy, 4(3),* 13-23.

Uglanova, I. (2021). Model criticism of Bayesian networks in educational assessment: A systematic review. *Practical Assessment, Research, and Evaluation, 26* (22), 1-17. doi: https://doi.org/10.7275/21220899

Wang, L., Jian, S., Liu, Y., & Xin, T. (2023) Using Bayesian networks for cognitive assessment of student understanding of buoyancy: A granular hierarchy model. *Applied Measurement in Education, 36*(1), 45-59. doi:10.1080/08957347.2023.2172014

Weaver, W. (1948). Probability, rarity, interest, and surprise. *The Scientific Monthly, 67*(6), 390–392. doi:10.2307/22339

West, P., Rutstein, D., Mislevy, R., Liu, J., Choi, Y., Levy, R., Crawford, A., DiCerbo, K., Chappel, K., & Behrens, J. (2010) *A Bayesian network approach to modeling learning progressions and task performance*. CRESST Report 776. National Center for Research on Evaluation, Standards, and Student Testing (CRESST). https://files.eric.ed.gov/fulltext/ED512650.pdf

Williamson, D., Almond, R., & Mislevy, R. (2000). Model criticism of Bayesian networks with latent variables. *Uncertainty in Artificial Intelligence Proceedings* (pp. 634–643). Morgan Kaufman Publishers.

Xing, W., Li, C., Chen, G., Huang, X., Chao, J., Massicotte, J., & Xie, C. (2020). Automatic assessment of students' engineering design performance using a Bayesian network model. *Journal of Educational Computing Research, 1-18*. https://doi.org/10.1177/0735633120960422

Yan, D., Zapata, D., & Almond, R. (2021, May). *Bayesian networks in educational assessment tutorial* [Workshop]. The National Council on Measurement in Education, Baltimore, USA.

794. https://doi.org/10.1080/00273171.2019.1590794

Levy, R., & Mislevy, R. (2016). *Bayesian psychometric modeling*. CRC Press.

Mahjoub, M. & Kalti, K. (2011). Software comparison dealing with Bayesian networks. *Advances in Neural Networks, 6677*, 168-177.

Martin, J., & Vanlehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human-Computer Studies, 42*, 575-591.

Mislevy, R. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*(4), 439–483. https://doi.org/10.1007/BF02294388

Mislevy, R. J., Almond, R., DiBello, L., Jenkins, F., Steinberg, L., Yan, D., & Senturk, D. (2002). *Modeling conditional probabilities in complex educational assessments* (Tech. Rep. No. 580). National Center for Research on Evaluation, Standards, and Student Testing.

Mislevy R. J., Almond R. G., Yan, D., & Steinberg L. (2000). *Bayes nets in educational assessment: Where do the numbers come from?* (Tech. Rep. No. 518). National Center for Research on Evaluation, Standards, and Student Testing.

Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered design: Layers, structures, and terminology.* SRI International.

Moussavi, M. & McGinny, N. (2009). *A Bayesian network for school performance*. Unpublished paper.

https://www.causallinks.com/wp-content/uploads/2019/07/A_Bayesian_Network_for_School.pdf

Murphy, K. (2007). *Software for graphical models: A review*. ISBA Bulletin.

Neapolitan, R. E. (2004). *Learning Bayesian networks*. Prentice Hall.

Pardos, Z., Feng, M., Heffernan, N. & Heffernan-Lindquist, C. (2007, July). *Analyzing fine-grained skill models using Bayesian and mixed effect methods* [Paper presentation]. Proceedings of the Educational Data Mining Workshop at the 13th International Conference on Artificial Intelligence in Education, Los Angeles, USA.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* Morgan Kaufmann Publishers, Inc.

Reckase, M. D. (2009). *Multidimensional item response theory models in multidimensional item response theory.* Springer

Reye, J. (2004). Student modeling based on belief networks. *International Journal of Artificial Intelligence in Education, 14*(1), 63–96.

Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guliford Press.