



DOR: 20.1001.1.24237906.1401.10.18.1.4

Research Article

Page 1-34

Evaluation of the Power of Concept Map-Based Tests in Replacing Traditional Tests**Sadegh Sayadi¹, Javad Mesrabadi^{2*}, Abolfazl Farid³, Ramin HabibiKaleybar⁴**

1. PhD Educational Psychology, Azarbaijan Shahid Madani University, Province Azarbaijan Sharqi, Iran.
2. Professor of Educational Psychology, Department of Educational Sciences, Faculty of Azarbaijan Shahid Madani University, Province Azarbaijan Sharqi, Iran
3. Associate Professor of Educational psychology, Department of Educational Sciences, Faculty of Azarbaijan Shahid Madani University, Province Azarbaijan Sharqi, Iran
4. Associate Professor of Educational Psychology, Department of Educational Sciences, Faculty of Azarbaijan Shahid Madani University, Province Azarbaijan Sharqi, Iran

Submit Date: 19 January 2021**Revise Date:** 18 June 2021**Accept Date:** 27 June 2021**Publication Date:** 21 June 2022**Abstract**

Objective: Traditional tests have been criticized by educational evaluation researchers for many years. In this regard, concept maps have shown that in addition to educational goals, they can also be used as educational evaluation tools, so the purpose of this research is to evaluate the power of tests based on concept maps as an alternative approach for tests. It was traditional.

Method: The statistical population of the research consisted of all first-year students of experimental sciences in Eslamabad-e-Gharb city in the academic year 2017-2018. The sample of the research included students of four classes (109 people) of the first year of the biology course of the second secondary level, who were selected by multi-stage cluster random sampling. The research tools included a four-choice educational achievement test, a hierarchical conceptual map completion test, a network conceptual map completion test, a hierarchical conceptual map drawing test, and a network conceptual map drawing test. The research design was an evaluation design. Statistical methods included Pearson's correlation coefficient, Kuder-Richardson coefficient, independent two-sample t-test and one-way analysis of variance (Kruskal-Wallis test and Benferroni's post hoc test).

Findings: The research results showed that four tests based on concept maps have high content validity and reliability. In terms of criterion validity, four concept map-based tests had different correlation values with a four-choice test. In terms of differential validity, with the exception of the conceptual network map drawing test, the other three tests were able to detect the difference between the scores of the two groups of high and low academic achievement.

Conclusion: The results of the present study showed that it is possible to use concept maps as evaluation tools, at least from a technical point of view; Therefore, it is necessary that the relevant officials in the country's education and evaluation system have a special and different view on the practical role of concept maps in the context of being used as a new and alternative evaluation tool.

Keywords: Evaluation, Concept map-Based Tests, Validity, Reliability.

Citation: Sayadi, s., Mesrabadi, J., Farid, A., & HabibiKaleybar, R. (2022). Evaluation of the Power of Concept Map-Based Tests in Replacing Traditional Tests. *Biquarterly Journal of Cognitive Strategies in Learning*, 10(18), 1-34.

***Corresponding Author:** Javad Mesrabadi

E-mail: mesrabadi@azaruniv.ac.ir

Extended Abstract

1. Introduction

Although voices are always heard about removing, modifying and modernizing the existing evaluation system, no fundamental action has ever been taken because there are always numerous obstacles that protect the existing evaluation system. Therefore, creating alternative evaluation methods has always been a necessity. In this regard, educational researchers have made efforts to discover effective forms of evaluation due to dissatisfaction with traditional evaluation methods, which one of the pioneers in changing this situation is constructivism theory (Arneson, 2005).

One of the evaluation methods consistent with constructivist theory is the concept map-based evaluation method, which is considered as one of the alternative evaluation methods to traditional evaluation methods. Ruiz-Primo and Shavelson (1996) pointed out that reliability and validity information about different mapping techniques should be supplied before concept maps are used for assessment (Yin & Shavelson, 2008). Because demonstrating the reliability and validity of any assessment measure is critical (Schmidt, 2006). Therefore, the main question in the present study is: Are various concept map-based tests effective in assessing learners' academic achievement? And whether these tests have the ability to replace traditional assessment tests?

2. Materials and Methods

The research design was an evaluation design. The statistical population of the study consisted of all students of the first year of the empirical science field in Eslamabad-e-Gharb city in academic year 2017-2018. Among these students, four classes (109 students) were selected by multi-stage random cluster sampling as the subjects of the study. For this purpose, four high schools were randomly selected from all high schools in Eslamabad-e-Gharb, and from each of these high schools, a first year class in empirical sciences was selected. Then, each of these classes was randomly placed in one of the four groups of drawing a hierarchical conceptual map, drawing a network conceptual map, completing a hierarchical conceptual map, and completing a network conceptual map. After selecting the biology textbook of the first year of high school as the educational text, the contents of different parts of this book were turned into concept maps by Cmap software. The following five tests (all of which are researcher-made) were also used to measure students' academic achievement:

1. Multiple choice test
2. Hierarchy concept map completion test
3. Network concept map completion test
4. Hierarchical concept map construction test
5. Network concept map construction test

Pearson's correlation coefficient, independent two-sample t-test, Kuder-Richardson method, inter-rater reliability, Kruskal-Wallis test and Benferroni's post hoc test were used for data analysis.

3. Results

In the present study, validity was evaluated in the forms of three types of content, criterion and differential validities and reliability was evaluated by two methods: Kuder-Richardson and inter-rater reliability.

To evaluate the content validity of the tests, the judgment of 8 biology teachers was used as a judge. The judges had to determine the extent to which the tests were able to cover the content and educational objectives of the topic. The results showed that among the concept map-based tests, the fill-in a hierarchical concept map test has a higher content validity (0.88 coverage of topics). The results of Kruskal-Wallis test to examine the difference between the judges' rankings on the content validity of the five tests showed that the multiple-choice test has the highest average score and the construct a network concept map test has the lowest average score. The calculated chi-square is equal to 10.41, which is statistically significant at the level of 0.05; Therefore, it can be concluded that five types of tests have different content validity. Also, the results of Benferoni post hoc test showed that out of ten comparisons made between five different tests, there is a significant difference between only multiple-choice test and the construct a network concept map test in terms of judges' ranks ($P < 0.05$). In other words, the average rank of the judges in the multiple-choice test (28.13) is significantly higher than the average rank of the judges in the construct a network concept map test (12.69). To know the desirability of the criterion validity of the concept map-based tests, the correlation coefficient between all four concept map-based tests with the scores of a multiple-choice test was calculated. The results showed that the construct a hierarchical concept map test has the highest Pearson correlation coefficient with the multiple-choice test (0.68). The lowest value of correlation coefficient belongs to the construct a network concept map test (0.30), which is not statistically significant.

In order to ensure the differential validity of different concept map-based tests in the present study, the ability of these tests to distinguish the upper and lower groups of academic achievement (based on the average scores of students in the multiple-choice test) was examined. An independent two-sample t-test was used to examine the difference between the mean of the upper and lower groups in each of the concept map-based tests. The results showed that the highest t value belongs to the the fill-in a hierarchical concept map test ($t = 3.44$) and the lowest t value belongs to the construct a network concept map test ($t = 1.61$), Which is the only test that failed to distinguish between two groups with high and low academic achievement.

The Kuder-Richardson 21 method was used to evaluate the reliability of the fill-in concept map tests; The obtained coefficient for the fill-in a hierarchical concept map test was 0.85 and for the fill-in a network concept map test was 0.79. In order to evaluate the reliability of construct concept map tests, the inter-rater reliability method was used. For this purpose, the correlation coefficient between the scores of eight rater was calculated. As the results showed, the construct a hierarchical concept map test had a 0.95 correlation coefficient and the construct a network concept map test had a 0.98 correlation coefficient, that showed these tests have high and significant inter-rater reliability.

To select the most appropriate concept map-based test from four tests, the test that has higher coefficients in terms of content validity, criterion validity, differential validity and reliability was selected as the most appropriate test.

Table 1. Values of content, criterion, differential validity and reliability of concept map-based tests

Test type	Reliability	Differential validity	Criterion validity	Content validity
Fill-in a hierarchical C-map	0.85	3/44	0.61	88/13
Fill-in a network C-map	0.79	2/73	0.54	87/50
Construct a hierarchical C-map	0.95	3/37	0.68	84/38
Construct a network C-map	0.98	1/61	0.30	83/13

As shown in Table 8, the various concept map-based tests have different values in terms of content validity, criterion validity, differential validity and reliability. By comparing these four tests, we find that the Fill-in a hierarchical concept map test in the two indicators of content validity and differential validity is higher than other tests. The weakest indicators are related to the construct a network concept map test. This test has the lowest coefficients in all three indicators of content validity, criterion validity and differential validity.

4. Discussion and Conclusion

The aim of this study was to investigate the validity and reliability of concept map-based tests. The results showed that these tests have different values of these indicators. The findings of the present study generally showed that the four concept map-based tests have high content validity. Regarding the criterion validity, the results showed that the four concept map-based tests have different correlation values from 0.30 to 0.68 with a multiple-choice test of academic achievement. Regarding the differential validity, the three tests of the Fill-in a hierarchical concept map, Construct a hierarchical concept map and Fill-in a network concept map were able to distinguish well between the scores of the two

groups of high and low academic achievement, but the construct a network concept map did not have this distinguishing power.

5. Ethical Considerations

All ethical principles are considered in this article. The participants were informed about the purpose of the study and its implementation stages. They were also assured about the confidentiality of their information and were free to leave the study.

Authors' contributions: All authors have participated in the design, implementation and writing of all sections of the present study.

Conflicts of interest: The authors declared no conflict of interest.

ارزشیابی توان آزمون‌های مبتنی بر نقشه مفهومی در جایگزینی آزمون‌های سنتی

Evaluation of the Power of Concept Map-Based Tests in Replacing Traditional Tests

صادق صیادی^۱، جواد مصرآبادی^{۲*}، ابوالفضل فرید^۳ و رامین حبیبی کلپیر^۴

دریافت مقاله: ۱۳۹۹/۱۰/۳۰ بازنگری مقاله: ۱۴۰۰/۳/۲۸
پذیرش مقاله: ۱۴۰۰/۰۴/۰۶ انتشار مقاله: ۱۴۰۱/۳/۳۶

چکیده

هدف: آزمون‌های سنتی برای سالیان متمادی موردانتقاد پژوهشگران حوزه ارزشیابی آموزشی بوده‌اند. در همین راستا نقشه‌های مفهومی نشان داده‌اند که علاوه بر اهداف آموزشی می‌توانند به‌عنوان ابزارهای ارزشیابی آموزشی نیز به‌کار روند، لذا هدف پژوهش حاضر ارزشیابی توان آزمون‌های مبتنی بر نقشه مفهومی به‌عنوان یک رویکرد جایگزین برای آزمون‌های سنتی بود.

روش: جامعه آماری پژوهش را تمامی دانش‌آموزان سال اول رشته علوم تجربی شهرستان اسلام‌آباد غرب در سال تحصیلی ۱۳۹۶-۱۳۹۷ تشکیل می‌دادند. نمونه پژوهش نیز شامل دانش‌آموزان چهار کلاس (۱۰۹ نفر) درس زیست شناسی سال اول مقطع متوسطه دوم بود که به روش نمونه‌گیری تصادفی خوشه‌ای چندمرحله‌ای انتخاب شدند. ابزارهای پژوهش شامل آزمون چهارگزینه‌ای پیشرفت تحصیلی، آزمون تکمیل نقشه مفهومی سلسله‌مراتبی، آزمون تکمیل نقشه مفهومی شبکه‌ای، آزمون ترسیم نقشه مفهومی سلسله‌مراتبی و آزمون ترسیم نقشه مفهومی شبکه‌ای بود. طرح پژوهش یک طرح ارزشیابی بود. روش‌های آماری شامل ضریب همبستگی پیرسون، ضریب کودر-ریچاردسون، آزمون t دو نمونه‌ای مستقل و تحلیل واریانس یک راهه (آزمون کروسکال-والیس و آزمون تعقیبی بنفرونی) بود.

یافته‌ها: نتایج پژوهش نشان داد که چهار آزمون مبتنی بر نقشه مفهومی از روایی محتوایی و پایایی بالایی برخوردارند. در زمینه روایی ملاکی چهار آزمون مبتنی بر نقشه مفهومی دارای مقادیر متفاوت همبستگی با یک آزمون چهارگزینه‌ای بودند. در زمینه روایی افتراقی، به‌استثنای آزمون ترسیم نقشه مفهومی شبکه‌ای، سه آزمون دیگر به‌خوبی از عهده تشخیص تفاوت بین نمرات دو گروه پیشرفت تحصیلی بالا و پایین برآمدند.

نتیجه‌گیری: نتایج پژوهش حاضر نشان داد که امکان استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی حداقل از لحاظ فنی وجود دارد؛ بنابراین ضروری است که مسئولان مربوطه در سیستم آموزش و ارزشیابی کشور نیز نگاه ویژه و متفاوتی به نقش کاربردی نقشه‌های مفهومی در زمینه استفاده شدن به‌عنوان یک ابزار ارزشیابی جدید و جایگزین داشته باشند.

کلید واژه‌ها: ارزشیابی، آزمون‌های مبتنی بر نقشه مفهومی، روایی، پایایی.

۱. دکتری روانشناسی تربیتی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران
۲. استاد، گروه علوم تربیتی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران
۳. دانشیار، گروه علوم تربیتی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران
۴. دانشیار، گروه علوم تربیتی، دانشکده علوم تربیتی و روانشناسی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران

۱. مقدمه

فهم دلایل تفاوت‌های فردی فراگیران در یادگیری و عملکرد تحصیلی همواره از دغدغه‌های معلمان، مربیان و پژوهشگران آموزشی در سطوح مختلف تحصیلی بوده است. در فرایند یادگیری و آموزش از سطح دبستان تا دانشگاه و آموزش عالی همواره نقش ارزشیابی و بازخورد^۱ مورد توجه بوده است. برای این منظور همواره ارزشیابی‌های گوناگونی از عملکرد فراگیران در سطوح مختلف به عمل می‌آید چراکه فرایند ارزشیابی از ارکان اساسی هر نظام آموزشی به شمار می‌رود. به صورت سنتی، دانش همواره با استفاده از آزمون‌های سنتی (مانند آزمون‌های چندگزینه‌ای، کوتاه‌پاسخ یا صحیح/غلط و ...) مورد ارزشیابی قرار گرفته است که در بیشتر مواقع این ارزشیابی‌ها، بینشی را درباره اینکه فراگیران چگونه مفاهیم را به یکدیگر ربط داده و چه معنایی به آن روابط می‌دهند، فراهم نمی‌آورد (جوزف، کنرادسون، ویکمار و رو، ۲۰۱۷). تجربه عملی از ارزشیابی نشان‌دهنده این است که اشکال آزمون‌های سنتی اگرچه ممکن است که عینی بوده و یا برای استفاده در جریان ارزشیابی آسان باشند، اما پاسخ‌گویی به آن‌ها اغلب مبتنی بر حافظه است؛ بنابراین چنین آزمون‌هایی تمایل دارند تا عوامل اصلی دانش را مورد ارزشیابی قرار دهند اما در تشخیص ساختار دانش شاگردان با مشکل روبرو می‌شوند (لیپینگ، یان، بوکینگ و زنگیاو^۳، ۲۰۰۹). اگرچه آزمون‌های سنتی جنبه‌های مثبتی دارند، اما استفاده از آن‌ها جهت اندازه‌گیری و ارزشیابی یادگیری فراگیران برای سال‌های متمادی شدیداً مورد انتقاد قرار گرفته است. این روش‌های سنتی عموماً کارایی لازم در زمینه ارزشیابی را ندارند. برای مثال، آزمون‌های چندگزینه‌ای معمولاً روابط پیچیده بین مفاهیم را نشان نمی‌دهند. علاوه بر این، آن‌ها به میزان زیادی پاسخ‌های شاگردان را محدود می‌سازند و در شناسایی و کشف دلایل پشت پاسخ‌ها با مشکل مواجه هستند. نهایتاً این‌که آزمون‌های چندگزینه‌ای اگرچه می‌توانند مفاهیم مهمی را که طراحان آزمون، قصد ارزشیابی آن‌ها را دارند مورد ارزشیابی قرار دهند اما برای ردیابی کج‌فهمی‌هایی^۴ (اطلاعاتی که قبلاً به صورت نادرست آموخته شده‌اند) که هنوز کشف نشده‌اند ابزارهای مفیدی نیستند؛ بنابراین آزمون‌های سنتی در تشخیص مؤثر اطلاعات ارزشمند درباره جنبه‌های اساسی ساختار دانش آزمودنی‌ها شکست می‌خورند (لیو، لین و تسای^۵، ۲۰۰۹). چراکه بیشتر روی یادآوری ایده‌ها و مفاهیم پراکنده و مجزا تمرکز می‌کنند (شوندیمن^۶، ۲۰۱۵).

آزمون‌های سنتی اغلب بازنشاسی یا یادآوری حقایق را مورد سنجش قرار می‌دهند و توجهی به اینکه فراگیران این حقایق یا مفاهیم را در یک چارچوب مفهومی بزرگ‌تر سازمان‌دهی کنند، ندارند. پرسش‌های فردی و اختصاصی در یک آزمون چندگزینه‌ای یا صحیح - غلط شبیه به تکه‌ها و ذرات

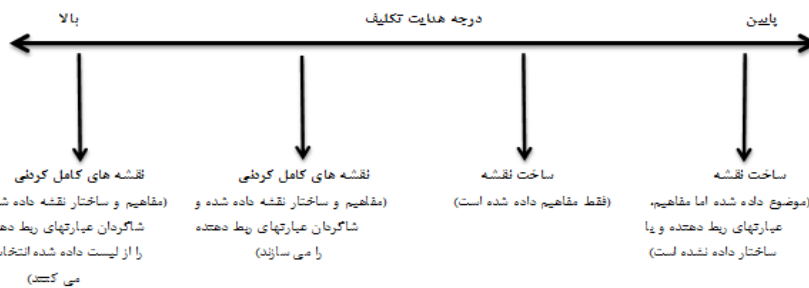
1. Evaluation & feedback
2. Joseph, Conradsson, Wikmar, & Rowe
3. Liping, Yan, Boqing, & Zengyao
4. misconceptions
5. Liu, Lin, & Tsai
6. Schwendimann

مجزا و از هم پراکنده دانش می‌باشند. هر پرسش اختصاصی ممکن است روی یک قسمت مهم از یک کل بزرگ‌تر متمرکز باشد اما تا زمانی که این اجزای فردی به هم مرتبط نشده و پیوند نیابند، تصویر ارائه‌شده توسط پازل تکمیل‌شده واضح نیست. ضعف ذاتی چنین آزمون‌هایی این است که یادگیرنده ممکن است همه یا بیشتر سؤالات مجزا را درست پاسخ دهد اما در به دست آوردن یک دیدگاه منسجم یا شبکه‌ای از ایده‌ها و مفاهیم منسجم ناتوان باشد. نارضایتی از شکل‌های سنتی ارزشیابی منجر به این شده است که تعدادی از مربیان و ارزیابان به فکر جستجوی روش‌های جایگزینی برای ارزشیابی یادگیری فراگیران باشند (اسنسون^۱، ۲۰۱۰). محققان عنوان کرده‌اند که ابداع راهبردهای ارزشیابی نوین به منظور پیشبرد یادگیری فردی فراگیران و آگاهی نسبت به ارزشیابی دوره‌های تحصیلی ضروری است (واتسون، پلکی، نویس و راجرز^۲، ۲۰۱۶). ارزشیابی در مدارس همواره موضوع چالش‌انگیزی بوده است، مخصوصاً با در نظر گرفتن این نکته که یک روش ارزشیابی می‌بایست ظرفیت واقعی شاگردان را منعکس سازد (نصری، روسلان، سکوان، ابوباکار و پوته^۳، ۲۰۱۰). اگرچه همواره صداهایی در زمینه برداشتن، اصلاح و نوسازی نظام ارزشیابی موجود شنیده می‌شود اما هیچ‌وقت اقدامی به‌صورت اساسی صورت نگرفته است چراکه همواره موانع متعددی از نظام ارزشیابی موجود محافظت می‌کنند؛ بنابراین، ایجاد روش‌های ارزشیابی جانشین همواره از ضروریات بوده است. در همین راستا پژوهشگران آموزشی تلاش‌هایی را در جهت کشف اشکال مؤثر ارزشیابی به دلیل نارضایتی از روش‌های ارزشیابی سنتی انجام داده‌اند که یکی از پیشگامان تغییر این وضعیت نظریه سازنده‌گرایی^۴ می‌باشد (آرنسون^۵، ۲۰۰۵).

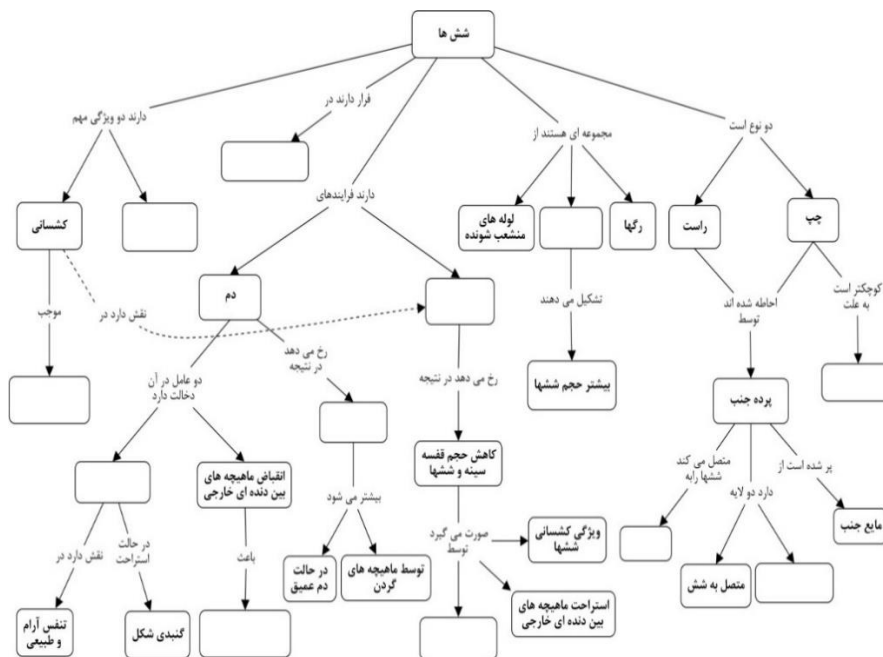
در اواخر دهه ۱۹۸۰ دیدگاه سازنده‌گرایی به‌عنوان یک نظریه آموزشی هدایت‌گر پدید آمد. در دیدگاه سازنده‌گرایی اعتقاد بر این است که فراگیران می‌توانند فعالانه دانششان را مبتنی بر تجارب قبلی‌شان بسازند و دانش جدید باید به ساختارهای موجود ربط داده شود تا به یاد آورده شده و معنادار شوند (ولدیمانوتل، آباته و برهین^۶، ۲۰۲۰). بر این اساس، دیدگاه سازنده‌گرایی نسبت به ارزشیابی نیز متفاوت از دیدگاه‌های سنتی ارزشیابی خواهد بود. یکی از روش‌های ارزشیابی هم‌خوان با نظریه سازنده‌گرایی، روش ارزشیابی مبتنی بر نقشه مفهومی^۷ می‌باشد که به‌عنوان یکی از روش‌های ارزشیابی جایگزین برای روش‌های ارزشیابی سنتی در نظر گرفته شده است. این آزمون‌ها از جنبه‌های گوناگون انواع متفاوتی دارند. تکالیف نقشه مفهومی از لحاظ درجه هدایت (درجه‌ای که مفاهیم، عبارت‌های ربط‌دهنده و کل ساختار برای فراگیران فراهم شده است) تنوع زیادی دارند. در

-
1. Snelson
 2. Watson, Pelkey, Noyes, & Rodgers
 3. Nasri, Roslan, Sekuan, Abu Bakar, & Puteh
 4. Constructivism
 5. Arneson
 6. Woldeamanuel, Abate & Berhane
 7. Concept map-based evaluation

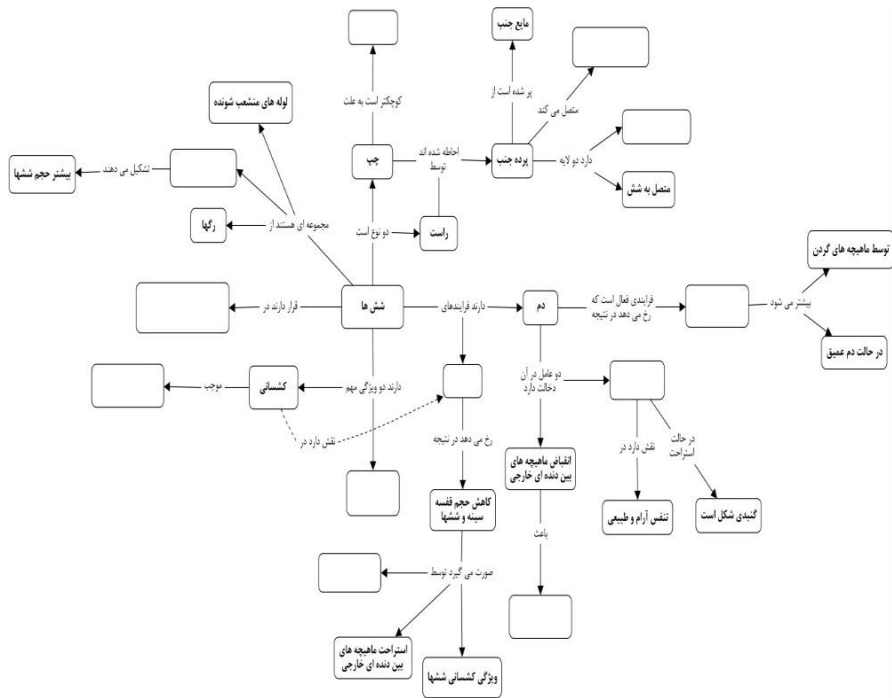
تکالیفی که کمترین میزان هدایت را دارند فقط یک موضوع یا سؤال اصلی برای فراگیران فراهم می‌شود و خود فراگیران کل نقشه را می‌سازند. در دیگر تکالیف علاوه بر موضوع اصلی لیستی از مفاهیم نیز در اختیار فراگیران قرار داده می‌شود. در تکالیفی که کمترین درجه هدایت‌شدگی را دارند تمام عناصر نقشه مفهومی در اختیار فراگیران قرار داده شده است و آن‌ها فقط باید نقشه‌های مفهومی را تکمیل کنند (تصویر ۱). از لحاظ شکلی نیز آزمون‌های مبتنی بر نقشه مفهومی تفاوت‌هایی اساسی دارند. دو نوع اصلی آن‌ها شامل آزمون‌های مبتنی بر نقشه‌های مفهومی سلسله‌مراتبی که مبتنی بر نظریه یادگیری معنادار می‌باشند (تصویر ۲) و آزمون‌های مبتنی بر نقشه‌های مفهومی شبکه‌ای که مبتنی بر نظریه تداومی‌گرا می‌باشند (تصویر ۳).



تصویر ۱. درجه هدایت تکالیف در آزمون‌های مبتنی بر نقشه مفهومی



تصویر ۲. یک نمونه از آزمون نقشه مفهومی سلسله‌مراتبی



تصویر ۳. یک نمونه از آزمون نقشه مفهومی شبکه‌ای

منطق استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی این است که اولاً آن‌ها پهنه‌ای را به ذهن شاگردان باز می‌کنند و ثانیاً آن‌ها برتر از روش‌های ارزشیابی هستند که فقط انتخاب یک گزینه را در بر می‌گیرند (مانند آزمون‌های چندگزینه‌ای، جور کردن پاسخ‌ها، آزمون‌های صحیح - غلط) چراکه نقشه‌های مفهومی برای ارزشیابی فهم شاگردان از اینکه چگونه مجموعه‌ای از مفاهیم مربوطه در داخل یک کل سازمان یافته جای می‌گیرند، ابزارهای مناسبی هستند (اسنلسون، ۲۰۱۰). ثالثاً آن‌ها می‌توانند جنبه‌هایی از یادگیری را اندازه‌گیری کنند که آزمون‌های سنتی قادر به اندازه‌گیری آن‌ها نیستند (استودارت، أبرامس، گاسپر و کانادای^۱، ۲۰۰۰). علاوه بر موارد یاد شده توانایی توصیف ساخت شناختی یادگیرنده از طریق ترسیم نقشه مفهومی منجر به استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی شده است (آرسون، ۲۰۰۵). این آزمون‌ها سه نوع تفکر تحلیلی، خلاق و عملی^۲ فراگیران را فعال می‌سازند (میکر و زیمرمن^۳، ۲۰۲۰).

علی‌رغم مزایای یاد شده روئیز-پریمو و شیولسون^۴ (۱۹۹۶) خاطرنشان ساختند که قبل از اینکه بتوان از تکالیف مختلف نقشه مفهومی برای اهداف ارزشیابی استفاده کرد می‌بایست اطلاعات موثقی

1. Stoddart, Abrams, Gasper & Canaday
 2. Analytical, Creative and Practical Thinking
 3. Maker & Zimmerman
 4. Ruiz-Primo & Shavelson

را درباره *روایی و پایایی*^۱ آن‌ها فراهم ساخت (بین^۲ و شیولسون، ۲۰۰۸) چراکه شرح دادن روایی و پایایی هر ابزار ارزشیابی ضروری است (اشمیت^۳، ۲۰۰۶). روایی و پایایی امتیازات به‌دست‌آمده از یک ارزشیابی آنچه را که شاخص‌های روانسنجی^۴ ارزشیابی نامیده می‌شود، تشکیل می‌دهند (پلومر^۵، ۲۰۰۸). علاوه بر این، ساختار نقشه‌های مفهومی نیز تفاوت‌هایی را در ارزشیابی مبتنی بر نقشه مفهومی رقم می‌زند. برای نمونه آمادیو، تریکوت و مارین^۶ (۲۰۰۹) نشان دادند که نقشه‌های با ساختار سلسله‌مراتبی به خوانندگان یک متن کمک می‌کنند تا در حین خواندن، انسجام و یکپارچگی فهم‌شان را حفظ کرده و احساس گم‌گشتگی مفهومی‌شان را کاهش دهند درحالی‌که نقشه‌های مفهومی با ساختار شبکه‌ای، الزامات و خواست‌های بیشتری را تحمیل کرده و درک و فهم مطالب را مختل می‌کنند.

در مجموع ارزشیابی دقیق اطلاعات و یادگیری شاگردان امروزه با مشکلات زیادی همراه است چراکه انجام ارزشیابی دقیقی که به عمق دانش و درک یادگیرندگان نفوذ کند، مشکل به نظر می‌رسد. ارزشیابی مبتنی بر نقشه مفهومی نیز به‌عنوان یکی از روش‌های نوین جهت ارزشیابی دقیق معلومات و توانایی‌های یادگیرندگان علی‌رغم مزایای عنوان شده با ابهامات و مشکلاتی نیز همراه است. تکالیف ارزشیابی (آزمون‌های) مبتنی بر نقشه مفهومی از لحاظ ساختار تکلیف و الزامات تکلیف تنوع بسیار زیادی دارند که هر یک از آن‌ها مزایا و معایب خاص خود را دارد. مسأله دیگر مربوط به این نوع ارزشیابی، نگرانی‌هایی است که در مورد روایی و پایایی آن‌ها وجود دارد چراکه لازمه انجام یک ارزشیابی دقیق و مؤثر داشتن روایی و پایایی مناسب است. بنا بر آنچه ذکر شد پژوهش حاضر در زمینه امکان‌سنجی استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی پیشرفت تحصیلی بر روی چند مؤلفه اساسی تمرکز دارد. اولین مؤلفه توجه به ساختار تکالیف ارزشیابی مبتنی بر نقشه مفهومی (نقشه‌های سلسله‌مراتبی در مقابل نقشه‌های شبکه‌ای) است، دومین مؤلفه توجه به الزامات تکالیف ارزشیابی مبتنی بر نقشه مفهومی (ترسیم نقشه در مقابل تکمیل نقشه) است و نهایتاً سومین مؤلفه مربوط به ویژگی‌های روان‌سنجی (روایی و پایایی) تکالیف ارزشیابی مبتنی بر نقشه مفهومی است. در واقع سؤال اساسی در پژوهش حاضر آن است که: آیا آزمون‌های مختلف مبتنی بر نقشه مفهومی در ارزشیابی پیشرفت تحصیلی فراگیران کارایی لازم را دارند؟ و اینکه آیا این آزمون‌ها توانایی لازم برای جایگزینی آزمون‌های سنتی ارزشیابی را دارند یا خیر؟

1. Validity & reliability

2. Yin

3 Schmidt

4. Psychometric Properties

5. Plummer

6. Amadiou, Tricot & Mariné

جریان یاددهی - یادگیری در محیط‌های تحصیلی نیاز فزاینده‌ای به استفاده از روش‌های ارزشیابی متنوع و معتبر دارد. یافته‌های پژوهش حاضر می‌تواند از اعتبار نقشه‌های مفهومی در ارزشیابی سطوح بالاتر توانایی‌های شناختی شاگردان و ایجاد یادگیری معنادار حمایت کند. با توجه به اینکه روش‌های مختلفی در طراحی تکالیف ارزشیابی مبتنی بر نقشه مفهومی وجود دارد که هر کدام جنبه‌های خاصی از توانایی‌های شناختی و یادگیری شاگردان را مورد اندازه‌گیری قرار می‌دهند بنابراین یافته‌های پژوهش حاضر می‌تواند در ارزشیابی نقشه‌های مفهومی گوناگون و میزان اثربخشی هر کدام از آن‌ها نیز کاربرد داشته باشد.

سؤالات آزمون

۱. آزمون‌های مبتنی بر نقشه مفهومی تا چه اندازه از روایی محتوایی برخوردارند؟
۲. آزمون‌های مبتنی بر نقشه مفهومی تا چه اندازه از روایی ملاکی برخوردارند؟
۳. آزمون‌های مبتنی بر نقشه مفهومی تا چه اندازه از روایی تمییزی (افتراقی) برخوردارند؟
۴. آزمون‌های مبتنی بر نقشه مفهومی تا چه اندازه از پایایی برخوردارند؟

۲. پیشینه پژوهش

در اوایل سال ۱۹۷۰ میلادی جوزف نواک و همکارانش سؤالاتی درباره اکتساب و استفاده کودکان از مفاهیم علمی مطرح و پیگیری نمودند. آن‌ها از سؤالات آزمون‌های سنتی (مانند آزمون‌های چندگزینه‌ای و کوتاه پاسخ) که در اندازه‌گیری مؤثر دانش یادگیرنده محدود و ناتوان بودند رضایت نداشتند. آن‌ها یافتند که در جریان پاسخ‌دهی دانش‌آموز به این سؤالات سنتی، در تعدادی از موارد بین پاسخ‌های درستی که دانش‌آموز انتخاب می‌کند با مفاهیمی که آن‌ها در مصاحبه‌های بعدی بیان می‌کنند، مطابقت اندکی وجود دارد. در حین بررسی نحوه اکتساب مفاهیم در کودکان، به‌وضوح آشکار شد که کودکان درک از مفاهیم را به‌صورت درجه‌ای، اکتساب می‌کنند. گروه نواک فرض گرفتند که کودکان درک از معنای یک مفهوم را زمانی که آن‌ها یک مجموعه از ارتباط‌های گزاره‌ای بین مفهوم مرکزی و دیگر مفاهیم مرتبط را اکتساب می‌کنند، به‌صورت درجه به درجه توسعه می‌دهند. آن‌ها نتیجه گرفتند که نمی‌توانند به‌صورت مستند تغییرات در درک مفهومی دانش‌آموزان را در طول زمان نشان دهند مگر اینکه بتوانند راهی را برای به‌دست آوردن تغییرات در ساختار شناختی یک دانش‌آموز - که آن‌ها مدعی بودند که یک مجموعه از ارتباطات گزاره‌ای وابسته به هم را در بر می‌گیرد - ابداع کنند. یکی از دستیاران تحقیقاتی قبلی نواک به نام ریچارد راول^۱، از دست‌نوشته‌های مربوط به مصاحبه‌های بالینی که انجام داده بود، نقشه‌های مفهومی را استخراج کرد. از شروع ساخت نقشه‌های مفهومی، راول آن‌ها را به‌منظور یافتن ارتباط‌های معنادار بین مفاهیم مرتبط مورد تجزیه و تحلیل قرار می‌داد. از این طریق و از طریق دیگر تلاش‌های

تحقیقاتی، استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای آموزشی و ارزشیابی آغاز شد. از آن زمان به بعد تحقیقات، استفاده‌های گوناگون از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی را به‌منظور اندازه‌گیری چارچوب مفهومی دانش‌آموزان ادامه داده‌اند (پلومر، ۲۰۰۸). استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی سابقه‌ای طولانی در تحقیقات خارجی دارد اما در ایران تحقیقات چندانی در این زمینه صورت نگرفته است. تعدادی از این پژوهش‌ها در بخش زیر آورده شده است:

مصرآبادی (۱۳۸۸) در پژوهشی به بررسی روایی و پایایی چهار نوع آزمون مبتنی بر نقشه مفهومی (ترسیم آزاد، ترسیم هدایت‌شده، تکمیل آزاد و تکمیل هدایت‌شده) و همچنین سه نظام نمره‌گذاری رابطه‌ای، ساختاری و شباهت به نقشه متخصص پرداخت. روایی محتوایی در این پژوهش برای نقشه‌های نوع تکمیل شدنی برابر با $0/84$ و برای نقشه‌های مفهومی ترسیمی برابر با $0/75$ به دست آمد. روایی ملاکی برای آزمون تکمیل آزاد نقشه مفهومی $r = 0/34$ ، تکمیل هدایت‌شده نقشه مفهومی $r = 0/42$ ، ترسیم آزاد نقشه مفهومی $r = 0/75$ و ترسیم هدایت‌شده نقشه مفهومی $r = 0/40$ بود. روایی افتراقی (براساس آزمون t مستقل) برای آزمون تکمیل آزاد نقشه مفهومی $t = 2/74$ ، تکمیل هدایت‌شده نقشه مفهومی $t = 3/24$ ، ترسیم آزاد نقشه مفهومی $t = 5/33$ و ترسیم هدایت‌شده نقشه مفهومی $t = 2/99$ به دست آمد. همچنین پایایی بین نمره‌گذاران^۱ نیز برای آزمون تکمیل آزاد نقشه مفهومی $0/91$ ، تکمیل هدایت‌شده نقشه مفهومی $0/96$ ، ترسیم آزاد نقشه مفهومی $0/82$ و ترسیم هدایت‌شده نقشه مفهومی $0/83$ به دست آمد. در زمینه روایی و پایایی نظام‌های نمره‌گذاری آزمون‌های ترسیم نقشه مفهومی نیز روایی ملاکی نظام رابطه‌ای $0/68$ ، نظام ساختاری $0/61$ و شباهت به نقشه متخصص $0/49$ به دست آمد. روایی افتراقی نظام رابطه‌ای $5/36$ ، نظام ساختاری $5/05$ و نظام شباهت $2/12$ بود. پایایی بین نمره‌گذاران (مصححان) نیز در نظام رابطه‌ای $0/91$ ، در نظام ساختاری $0/85$ و در نظام شباهت به نقشه متخصص $0/72$ بود. اکین، اولسینار-ساگیر و سالتان^۲ (۲۰۱۶) به برآورد روایی و پایایی نقشه‌های مفهومی کامل کردنی و ساخت یادگیرنده پرداختند. نتایج آن‌ها نشان داد که پایایی بین نمره‌گذاران در نقشه‌های مفهومی کامل کردنی $0/89$ و در نقشه‌های مفهومی ساخت یادگیرنده $0/80$ بود. به‌منظور برآورد روایی ملاکی نیز به محاسبه ضریب همبستگی بین نمره‌های به‌دست‌آمده از نقشه‌های مفهومی (نقشه‌های ساخت یادگیرنده با استفاده از نظام نمره‌گذاری ساختاری) با نمره‌های به‌دست‌آمده از آزمون‌های چندگزینه‌ای و آزمون‌های شبکه‌های ساختاریافته^۳ پرداخته شد. نتایج نشان داد که ضریب همبستگی بین آزمون نقشه مفهومی ساخت یادگیرنده با آزمون چندگزینه‌ای $0/51$ ، با آزمون شبکه ساختاریافته $0/27$ و با آزمون نقشه مفهومی کامل کردنی $0/46$ بود. ضریب همبستگی بین آزمون کامل کردنی نقشه

1. Inter-rater Reliability
2. Ekin, Ulcinar Sagir, & Saltan
3. Structured Grids

مفهومی نیز با آزمون چندگزینه‌ای ۰/۸۱ و با آزمون شبکه ساختاریافته ۰/۲۲ بود. هانگ و لین^۱ (۲۰۱۵) نیز به منظور بررسی روایی نقشه‌های مفهومی‌شان از روایی محتوایی استفاده کردند. برای این منظور آن‌ها از نظرات سه متخصص استفاده کردند و به ضریب همبستگی پیرسون ۰/۹۲ دست یافتند. این پژوهشگران برای بررسی پایایی نقشه‌ها به محاسبه ضریب همبستگی درون طبقه‌ای (ICC)^۲ بین نمره‌گذاران پرداختند. نتایج آن‌ها نشان داد که ضریب همبستگی درون طبقه‌ای برای روابط بین مفاهیم ۰/۹۹، سطوح سلسله‌مراتب ۰/۹۶، همبرها ۰/۹۲ و مثال‌ها (نمونه‌ها) ۰/۹۵ بود. کراوالیو^۳ (۲۰۱۰) نیز در پژوهش خود به بررسی پایایی و روایی نقشه‌های مفهومی در درس آمار پرداخت. برای این منظور وی چهار گروه ترسیم نقشه مفهومی را مورد استفاده قرار داد که در آن، گروه یک در مورد مبحث آزمون فرضیه، گروه دو در مورد مبحث شاخص t، گروه سه در مورد مبحث برآورد و گروه چهار در مورد مبحث ANOVA به ساختن نقشه‌های مفهومی پرداختند. سپس نقشه‌های مفهومی یادگیرندگان چهار گروه، به دو روش متفاوت کمی و کیفی نمره‌گذاری شدند. به منظور برآورد پایایی نقشه‌های مفهومی از روش پایایی بین نمره‌گذاران استفاده شد. برای این منظور دو نمره‌گذار به صورت مستقل به نمره‌گذاری نقشه‌های مفهومی شاگردان به دو نظام نمره‌گذاری کمی و کیفی پرداختند و سپس نمره‌های حاصل از دو نظام نمره‌گذاری را با هم جمع نموده (برای هر گروه به صورت جداگانه) و به یک نمره کلی تبدیل کردند. سپس به محاسبه ضریب همبستگی پیرسون بین نمرات تولیدشده توسط هر دو نمره‌گذار پرداخته شد. ضریب همبستگی بین نمرات دو نمره‌گذار در گروه یک (مبحث آزمون فرضیه) ۰/۹۸، در گروه دو (مبحث شاخص t) ۰/۹۷، در گروه سه (مبحث برآورد) ۰/۹۸ و در گروه چهار (مبحث ANOVA) ۰/۹۹ به دست آمد. همچنین به منظور برآورد روایی (از نوع همزمان) به محاسبه ضریب همبستگی بین دو نظام نمره‌گذاری کمی و کیفی پرداخته شد. نتایج نشان داد که ضریب همبستگی بین دو نظام نمره‌گذاری کمی و کیفی در گروه یک ۰/۵۸، در گروه دو ۰/۴۷، در گروه سه ۰/۳۹ و در گروه چهار ۰/۴۴ بود. تمامی این ضرایب به استثنای ضریب همبستگی بین دو نظام نمره‌گذاری کمی و کیفی در گروه سه (۰/۳۹) معنادار بودند. همچنین این ضرایب همبستگی بین دو نظام نمره‌گذاری کمی و کیفی در پژوهش آنان به عنوان شاهی برای روایی همگرا^۴ نیز مورد استفاده قرار گرفت (چراکه دو نظام متفاوت نمره‌گذاری سازه یکسانی را مورد اندازه‌گیری قرار می‌دادند). به منظور برآورد روایی تمیزی (افتراقی) یک نقطه برش بین نمرات به دست آمده توسط شاگردان در هر یک از گروه‌های ترسیم نقشه مفهومی ایجاد شد که شاگردان را به دو گروه ماهر و غیرماهر در ترسیم نقشه مفهومی تقسیم می‌کرد و سپس عملکرد این دو گروه ماهر و غیرماهر در دو آزمون (با استفاده از یک آزمون t

1. Hung & Lin
2. Intra-Class Correlation
3. Cravalho
4. Convergent validity

دونمونه‌ای مستقل) مورد مقایسه قرار می‌گرفت؛ بنابراین گروه‌های ماهر و غیرماهر در ترسیم نقشه مفهومی در گروه دو (مبحث شاخص t) با آزمون اول و گروه‌های ماهر و غیرماهر در گروه چهار (مبحث ANOVA) با آزمون دوم مورد مقایسه قرار گرفتند به این علت که ایجاد این نقشه‌ها در نزدیک‌ترین زمان به دو امتحان داده‌شده به آن‌ها قرار داشت (هر کدام از دو آزمون نیز سؤالات محاسبه‌ای و مفهومی را در بر می‌گرفت). آزمون t دونمونه‌ای مستقل برای مقایسه عملکرد امتحان گروه‌های ماهر و غیرماهر در ترسیم نقشه مفهومی (مبحث شاخص t) در آزمون اول نشان داد که شاخص t برابر با ۱/۶۱ (غیرمعنادار) می‌باشد. آزمون t دونمونه‌ای مستقل در آزمون دوم (برای مقایسه عملکرد گروه‌های ماهر و غیرماهر ترسیم نقشه مفهومی در مبحث ANOVA) نیز نشان داد که شاخص t برابر با ۲/۱۸ (معنادار) می‌باشد. ون، کرابه، لی، تریگاست و فیشر^۱ (۲۰۱۷) نیز در زمینه پایایی آزمون‌های مبتنی بر نقشه مفهومی، ضریب همسانی درونی نقشه‌های مفهومی را ۰/۸۰ و پایایی بین نمره‌گذاران را ۰/۵۴ گزارش کردند. نتایج آن‌ها در زمینه روایی ملاکی نیز نشان داد که ضریب همبستگی بین نمرات آزمون نقشه مفهومی و نمرات یک آزمون چندگزینه‌ای برابر با ۰/۲۱ بود.

۳. روش پژوهش

طرح پژوهشی مورد استفاده در این پژوهش یک طرح ارزشیابی بود. طبق نظر روسی (۱۹۸۲؛ نقل از سیف، ۱۳۸۷) این دسته از پژوهش‌ها نوعی ارزشیابی هستند که مبتنی بر شواهد تجربی جمع‌آوری شده به راه‌هایی که قابل تکرار است می‌باشند و در آن‌ها مسائل مربوط به روایی درونی، روایی بیرونی و روایی سازه مورد توجه قرار می‌گیرد. در پژوهش حاضر نیز شاخص‌های مربوط به روایی و پایایی آزمون‌های مبتنی بر نقشه مفهومی مورد بررسی قرار گرفتند.

۳-۱. جامعه، نمونه آماری و روش نمونه‌گیری

جامعه آماری^۲ پژوهش حاضر را تمامی دانش‌آموزان سال اول رشته علوم تجربی مقطع متوسطه دوم شهرستان اسلام‌آباد غرب در سال تحصیلی ۱۳۹۶-۱۳۹۷ تشکیل دادند. دانش‌آموزان چهار کلاس (۱۰۹ نفر) به روش نمونه‌گیری تصادفی خوشه‌ای چندمرحله‌ای^۳ به‌عنوان آزمودنی‌های پژوهش انتخاب شدند. برای این منظور از بین تمام دبیرستان‌های شهرستان اسلام‌آباد غرب چهار دبیرستان به‌صورت تصادفی انتخاب گردید که از هر یک از این دبیرستان‌ها یک کلاس سال اول رشته علوم تجربی انتخاب شد. سپس هر کدام از این کلاس‌ها به‌صورت تصادفی در یکی از گروه‌های چهارگانه

1. Won, Krabbe, Ley, Treagust & Fischer

2. Statistical population

3. Multi-stage cluster sampling method

ترسیم نقشه مفهومی سلسله‌مراتبی، ترسیم نقشه مفهومی شبکه‌ای، تکمیل نقشه مفهومی سلسله‌مراتبی و تکمیل نقشه مفهومی شبکه‌ای قرار گرفتند.

۳-۲. ابزارهای پژوهش

پس از انتخاب کتاب درسی زیست‌شناسی سال اول دبیرستان به‌عنوان متن آموزشی، مطالب قسمت‌های مختلف این کتاب به‌وسیله نرم‌افزار *Cmap* به نقشه‌های مفهومی تبدیل شدند. همچنین برای اندازه‌گیری میزان پیشرفت تحصیلی دانش‌آموزان از پنج آزمون (تمامی آن‌ها محقق ساخته هستند) زیر استفاده شد:

(۱) **آزمون چهارگزینه‌ای:** در این آزمون محقق ساخته، اهداف آموزشی محتوای درسی به شکل ۳۳ سؤال چهارگزینه‌ای و در دو سطح دانش و درک (طبقه‌بندی بلوم از حیطه شناختی) موردسنجش قرار گرفت. این آزمون به‌عنوان مقیاس ملاک مورد استفاده قرار گرفت.

(۲) **آزمون تکمیل نقشه مفهومی سلسله‌مراتبی:** در این آزمون یک نقشه مفهومی سلسله‌مراتبی محقق ساخته که قسمتی از مفاهیم آن حذف شده بود در اختیار آزمودنی‌ها قرار گرفت و از آن‌ها خواسته شد تا نقشه را کامل کنند.

(۳) **آزمون تکمیل نقشه مفهومی شبکه‌ای:** در این آزمون یک نقشه مفهومی شبکه‌ای محقق ساخته که قسمتی از مفاهیم آن حذف شده بود در اختیار آزمودنی‌ها قرار گرفت و از آن‌ها خواسته شد تا نقشه را کامل کنند.

(۴) **آزمون ترسیم نقشه مفهومی سلسله‌مراتبی:** در این آزمون مجموعه‌ای از مفاهیم و عبارت‌های ربط‌دهنده (که توسط محقق فراهم شده بود) در اختیار آزمودنی‌ها قرار گرفت و از آن‌ها خواسته شد تا با استفاده از آن‌ها به ترسیم نقشه‌های مفهومی سلسله‌مراتبی اقدام نمایند.

(۵) **آزمون ترسیم نقشه مفهومی شبکه‌ای:** در این آزمون نیز مانند آزمون سلسله‌مراتبی مجموعه‌ای از مفاهیم و عبارت‌های ربط‌دهنده در اختیار آزمودنی‌ها قرار گرفت و از آن‌ها خواسته شد تا با استفاده از آن‌ها به ترسیم نقشه‌های مفهومی شبکه‌ای اقدام نمایند.

۴. روش اجرا

جزئیات اجرای پژوهش در قالب شش مرحله به شرح زیر بود:

۱- آماده‌سازی:

در این مرحله مقدمات اجرای پژوهش فراهم شد. فعالیت‌های این مرحله عبارت بودند از: انتخاب مدارس و کلاس‌های نمونه، هماهنگی با آموزش و پرورش استان و شهرستان و مدارس، آماده‌سازی ابزار و وسایل لازم، آموزش مدرسین برای آموزش نقشه‌کشی مفهومی به آزمودنی‌ها.

لازم به ذکر است که در فرایند آموزش مدرسین چهار معلم زیست‌شناسی در یک جلسه دو ساعته با نحوه ترسیم و تکمیل نقشه‌های مفهومی آشنا شدند.

۲- آموزش

قبل از اجرای آزمون‌ها، دو جلسه ۶۰ دقیقه‌ای آموزشی برای یاد دادن نحوه تکمیل یا ترسیم نقشه مفهومی بر روی چهار گروه اجرا شد. در طول این دو جلسه آموزش، فراگیران چهار گروه آزمایشی با شیوه‌های متفاوت تحت آموزش مطالب آموزشی یکسانی قرار گرفتند. جزئیات اجرا در این چهار گروه به شرح زیر بود:

رویه آموزش دو گروه ترسیم نقشه مفهومی (سلسله‌مراتبی و شبکه‌ای)

آموزش نقشه مفهومی در این دو گروه، سه مرحله اصلی که هر کدام شامل چند زیر مرحله بود را در بر گرفت:

مرحله اول- تعریف و توصیف نقشه مفهومی و فواید نقشه مفهومی

مرحله دوم- طراحی نقشه مفهومی

فرایندی که نقشه‌های مفهومی تولید می‌شوند به شرح زیر است:

گام اول: در حین (بعد از) خواندن یک متن، روی آن قسمت از متن که به آن علاقمند هستید تمرکز کنید و از خودتان بپرسید که نقشه مفهومی من درباره چیست؟
گام دوم: مفاهیم مناسب و مربوطه متن را شناسایی کنید.

گام سوم: شما می‌بایست با ساخت سلسله‌مراتبی از مفاهیم شناسایی شده شروع کنید. برای این منظور مفاهیم را لیست کرده و درباره مفاهیم اصلی و عمده و مفاهیم زیرمجموعه و فرعی فکر کنید. کلی‌ترین و جامع‌ترین مفاهیم را مشخص کنید؟ کدام یک از مفاهیم جامعیت و اهمیت کمتری دارند؟ پس‌از آن، مفاهیم کلی را در بالا (در نقشه‌های مفهومی سلسله‌مراتبی) و یا در مرکز (در نقشه‌های مفهومی شبکه‌ای) نقشه قرار داده و مفاهیم با جامعیت و اهمیت کمتر را در پایین نقشه (در نقشه‌های مفهومی سلسله‌مراتبی) و یا در حاشیه‌های نقشه (در نقشه‌های مفهومی شبکه‌ای) قرار دهید.

گام چهارم: روابط میان مفاهیم را برقرار کنید. با مفاهیم کلی شروع کنید.

گام پنجم: عبارت‌های ربط‌دهنده مناسب را انتخاب کنید و بررسی کنید آیا روابط بین مفاهیم معنادار است؟

گام ششم: با استفاده از فلش‌ها (پیکان‌ها) جهت روابط را تعیین کنید تا روابط به گزاره (جملات معناداری با مفاهیم، عبارت‌های ربط‌دهنده و جهت) تبدیل شوند.

گام هفتم: تمام روابط را با مفاهیمی که شناسایی کردید به صورت یک نقشه ترسیم کنید.

گام هشتم: ساختار نقشه را مجدداً مورد بازبینی و اصلاح قرار دهید.

گام نهم: نمونه‌هایی از نقشه‌های مفهومی سلسله‌مراتبی یا شبکه‌ای را مشاهده و بررسی کنید.

مرحله سوم- تمرین طراحی نقشه مفهومی

در این مرحله شاگردان پس از آشنایی با نقشه مفهومی و نحوه طراحی آن، به ترسیم نقشه‌های مفهومی درباره موضوعات درس زیست‌شناسی پرداختند.

دو نکته لازم به توجه در این مرحله عبارت بودند از:

(۱) شروع از وسط نقشه و حرکت تدریجی به کناره‌ها (در نقشه‌های مفهومی شبکه‌ای) و یا شروع از بالای نقشه و حرکت تدریجی به پایین (در نقشه‌های مفهومی سلسله‌مراتبی) و (۲) حرکت مطالب در جهت عقربه‌های ساعت در نقشه‌های مفهومی شبکه‌ای و حرکت مطالب از سمت راست نقشه به سمت چپ نقشه در نقشه‌های مفهومی سلسله‌مراتبی.

رویه دو گروه تکمیل نقشه مفهومی (سلسله‌مراتبی و شبکه‌ای)

در این دو گروه نیز شاگردان دو مرحله اصلی آموزش دو گروه قبلی را دریافت کردند؛ اما در مرحله سه، به جای تمرین ترسیم نقشه، آزمودنی‌های نقشه‌های مفهومی ناقص از قبل تهیه‌شده را دریافت و نحوه تکمیل کردن آن‌ها را تمرین کردند.

۳- ارزشیابی

در مرحله سوم آزمودنی‌های تمام گروه‌ها به یک آزمون چهارگزینه‌ای و یک آزمون مبتنی بر نقشه مفهومی پاسخ دادند. مدت زمان اجرای این مرحله ۸۵ دقیقه بود. در این مرحله ابتدا هر گروه آزمون چهارگزینه‌ای پیشرفت تحصیلی را ظرف مدت ۱۵ دقیقه پاسخ دادند. بعد از چند دقیقه استراحت هر کدام از گروه‌ها آزمون نقشه مفهومی مربوط به خود را انجام دادند. به این صورت که از یک گروه خواسته شد تا نقشه مفهومی سلسله‌مراتبی مبحث (گوارش غذا) را ترسیم کنند. از گروه دوم خواسته شد تا نقشه مفهومی شبکه‌ای مبحث را ترسیم کنند. از گروه سوم خواسته شد تا نقشه مفهومی سلسله‌مراتبی ناقص مبحث را تکمیل کنند و نهایتاً از گروه چهارم خواسته شد تا نقشه مفهومی شبکه‌ای ناقص مبحث را تکمیل کنند. مدت زمان لازم برای انجام آزمون‌های مبتنی بر نقشه مفهومی ۶۰ دقیقه بود.

۴- آموزش نمره‌گذاران

در پژوهش حاضر از هشت داور مستقل برای نمره‌گذاری آزمون‌های مبتنی بر نقشه مفهومی استفاده شد. این هشت داور از معلمان زیست‌شناسی استان کرمانشاه بودند که طی یک جلسه آموزش ۹۰ دقیقه‌ای با روش‌های مختلف نمره‌دهی آزمون‌های مبتنی بر نقشه مفهومی آشنا شدند.

۵- نمره‌دهی آزمون‌ها

آزمون چهارگزینه‌ای: برای نمره‌دهی این آزمون به شمارش پاسخ‌های درست از کل ۳۳ سؤال براساس یک کلید پاسخ سؤالات اقدام و سپس نسبت نمرات نهایی از ۲۰ محاسبه و نمره نهایی هر آزمودنی مشخص شد.

دو آزمون تکمیل نقشه مفهومی سلسله‌مراتبی و شبکه‌ای: این نمره‌دهی براساس یک نمونه کلید از نقشه مفهومی تهیه‌شده توسط گروه پژوهشی با توجه به میزان تکمیل بخش‌های مختلف نقشه

مفهومی ناقص صورت گرفت که در آن به هر جای خالی صحیح پر شده یک نمره داده شد و سپس نسبت نمرات از ۲۰ نمره محاسبه و نمره نهایی هر آزمودنی مشخص شد.

دو آزمون ترسیم نقشه مفهومی سلسله‌مراتبی و شبکه‌ای: برای نمره‌دهی این آزمون‌ها از سه روش نمره‌دهی رابطه‌ای، ساختاری و شباهت به نقشه متخصص استفاده شد. در روش رابطه‌ای به هر رابطه بین مفاهیم به‌طور جداگانه (بدون در نظر گرفتن دیگر رابطه‌ها) نمره‌ای اختصاص داده می‌شود. در روش ساختاری به دانش ساختاری سلسله‌مراتبی اهمیت داده می‌شود. در روش شباهت به نقشه متخصص نقشه مفهومی فراگیر با یک نقشه مفهومی معیار (تهیه شده توسط ارزیاب) مورد مقایسه قرار می‌گیرد و در آن همخوانی هر گزاره در نقشه مفهومی آزمودنی با گزاره موجود در نقشه مفهومی معیار یک نمره برای آزمودنی به همراه دارد. در نهایت میانگین هر سه روش نمره‌دهی محاسبه و نسبت نمرات از ۲۰ نمره محاسبه و نمره نهایی هر آزمودنی مشخص شد.

۵. روش‌های آماری

برای تحلیل داده‌ها از ضریب همبستگی پیرسون^۱، آزمون *t* دونمونه‌ای مستقل^۲، روش کودر - ریچاردسون^۳، پایایی بین نمره‌گذاران و آزمون کروسکال-والیس^۴ و آزمون تعقیبی بنفرونی^۵ استفاده شد.

۶. یافته‌های پژوهش

در پژوهش حاضر روایی در قالب سه نوع روایی محتوایی، ملاکی و افتراقی و پایایی نیز به دو روش کودر-ریچاردسون و پایایی بین نمره‌گذاران مورد ارزشیابی قرار گرفتند.

برای بررسی روایی محتوایی آزمون‌ها از قضاوت ۸ نفر از مدرسین درس زیست‌شناسی به‌عنوان داور استفاده شد. داوران بایستی مشخص می‌کردند که تا چه حدی آزمون‌ها توانسته‌اند محتوا و اهداف آموزشی مبحث مربوطه را پوشش دهند.

1. Pearson correlation coefficient
 2. Independent t test
 3. Kuder-Richardson
 4. Kruskal-Wallis
 5. Bonferroni post hoc

جدول ۱. میزان انطباق محتوای آزمون‌های مختلف با اهداف و محتوای آموزشی

نوع آزمون	دور ۱	دور ۲	دور ۳	دور ۴	دور ۵	دور ۶	دور ۷	دور ۸	میانگین
چهارگزینه‌ای	۹۰	۸۵	۸۵	۹۰	۹۵	۹۰	۹۵	۹۰	٪۹۰
تکمیل نقشه مفهومی سلسله‌مراتبی	۹۰	۹۰	۹۵	۹۰	۹۰	۸۰	۸۵	۸۵	٪۸۸/۱۳
تکمیل نقشه مفهومی شبکه‌ای	۸۰	۹۰	۸۵	۸۵	۹۰	۸۵	۹۵	۹۰	٪۸۷/۵۰
ترسیم نقشه مفهومی سلسله‌مراتبی	۸۵	۸۵	۹۰	۸۰	۸۵	۸۰	۹۰	۸۰	٪۸۴/۳۸
ترسیم نقشه مفهومی شبکه‌ای	۸۰	۹۰	۸۰	۸۰	۹۰	۸۰	۸۰	۸۵	٪۸۳/۱۳

همان‌گونه که در جدول ۱ مشخص است بر مبنای نظر داوران آزمون چهارگزینه‌ای بیشتر از دیگر آزمون‌ها محتوا و اهداف آموزشی مبحث را پوشش داده است. از بین آزمون‌های مبتنی بر نقشه مفهومی نیز آزمون تکمیل نقشه مفهومی سلسله‌مراتبی روایی محتوایی بالاتری دارد. کمترین میزان پوشش اهداف و محتوای آموزشی مبحث نیز مربوط به آزمون‌های ترسیم نقشه مفهومی شبکه‌ای است.

جدول ۲. نتایج آزمون کروسکال - والیس برای ارزشیابی رتبه‌های داوران در پنج نوع آزمون

نوع آزمون	تعداد داوران	میانگین رتبه	مجذور کای	درجه آزادی	معناداری
چهارگزینه‌ای	۸	۲۸/۱۳			
تکمیل نقشه مفهومی سلسله‌مراتبی	۸	۲۴			
تکمیل نقشه مفهومی شبکه‌ای	۸	۲۲/۳۸	۱۰/۴۱	۴	۰/۰۳۴
ترسیم نقشه مفهومی سلسله‌مراتبی	۸	۱۵/۳۱			
ترسیم نقشه مفهومی شبکه‌ای	۸	۱۲/۶۹			

جدول ۲ نتایج آزمون کروسکال - والیس برای بررسی تفاوت رتبه‌های داوران در مورد روایی محتوایی پنج آزمون مورد استفاده در پژوهش را نشان می‌دهد. همان‌گونه که مشخص است آزمون چهارگزینه‌ای از بالاترین میانگین رتبه و آزمون ترسیم نقشه مفهومی شبکه‌ای از پایین‌ترین میانگین رتبه برخوردار است. مقدار مجذور کای محاسبه شده برابر با ۱۰/۴۱ است که از لحاظ آماری در سطح ۰/۰۵ معنادار است؛ بنابراین می‌توان نتیجه گرفت که پنج نوع آزمون از روایی محتوایی متفاوتی برخوردارند.

با مشخص شدن وجود تفاوت بین رتبه‌های داوران در روایی محتوایی پنج آزمون می‌بایست بررسی شود که این تفاوت بین رتبه‌های داوران در کدام آزمون‌ها بوده است. برای این منظور از

آزمون تعقیبی بنفرونی برای مقایسه دو به دو آزمون‌ها استفاده شد. نتایج این آزمون در بخش زیر ارائه شده است:

جدول ۳. آزمون تعقیبی بنفرونی برای بررسی تفاوت رتبه‌های داوران در پنج آزمون

فاصله اطمینان %۹۵	معداری	خطای معیار	تفاوت میانگین	آزمون	حد	
					پایین	بالا
۱۷/۷۳	۱/۰۰۰	۵/۰۴	۲/۶۳	ترسیم سلسله‌مراتبی ترسیم شبکه‌ای	-۱۲/۴۸	۱۷/۷۳
۱۹/۲۳	۱/۰۰۰	۵/۰۴	۴/۱۳	چهارگزینه‌ای تکمیل سلسله مراتبی	-۱۰/۹۸	۱۹/۲۳
۲۰/۸۵	۱/۰۰۰	۵/۰۴	۵/۷۵	تکمیل شبکه‌ای	-۹/۳۵	۲۰/۸۵
۲۷/۹۱	۰/۱۵۶	۵/۰۴	۱۲/۸۱	ترسیم سلسله‌مراتبی	-۲/۲۹	۲۷/۹۱
۳۰/۵۴	۰/۰۴۲	۵/۰۴	۱۵/۴۴	ترسیم شبکه‌ای	۰/۳۴	۳۰/۵۴
۱۶/۷۳	۱/۰۰۰	۵/۰۴	۱/۶۳	تکمیل سلسله‌مراتبی تکمیل شبکه‌ای	-۱۳/۴۸	۱۶/۷۳
۲۳/۷۹	۰/۹۴	۵/۰۴	۸/۶۹	ترسیم سلسله‌مراتبی	-۶/۴۲	۲۳/۷۹
۲۶/۴۲	۰/۳۱۳	۵/۰۴	۱۱/۳۱	ترسیم شبکه‌ای	-۳/۷۹	۲۶/۴۲
۲۲/۱۷	۱/۰۰۰	۵/۰۴	۷/۰۶	تکمیل شبکه‌ای ترسیم سلسله‌مراتبی	-۸/۰۴	۲۲/۱۷
۲۴/۷۹	۰/۶۲۸	۵/۰۴	۹/۶۹	ترسیم شبکه‌ای	-۵/۴۲	۲۴/۷۹

نتایج اجرای آزمون تعقیبی بنفرونی نشان داد که از ده مقایسه انجام شده بین پنج آزمون مختلف تنها بین دو آزمون چهارگزینه‌ای و آزمون ترسیم نقشه مفهومی شبکه‌ای از لحاظ رتبه‌های داوران تفاوت معناداری وجود دارد ($P < ۰/۰۵$)؛ به عبارت دیگر میانگین رتبه‌های داوران در آزمون چهارگزینه‌ای (۲۸/۱۳) به طور معناداری بالاتر از میانگین رتبه‌های داوران در آزمون ترسیم نقشه مفهومی شبکه‌ای (۱۲/۶۹) است.

در این پژوهش برای آگاهی از میزان مطلوب بودن روایی ملاکی آزمون‌های مبتنی بر نقشه مفهومی ضریب همبستگی بین هر چهار آزمون مبتنی بر نقشه مفهومی با نمرات یک آزمون چهارگزینه‌ای محاسبه شد که نتایج آن‌ها در جدول زیر نشان داده شده است.

جدول ۴. ضریب همبستگی نمرات آزمون چهارگزینه‌ای با نمرات چهار آزمون مبتنی بر نقشه مفهومی

معداری	همبستگی با آزمون چهارگزینه‌ای		نوع آزمون
	n	r	
۰/۰۰۱	۲۷	۰/۶۱	تکمیل نقشه مفهومی سلسله‌مراتبی
۰/۰۰۳	۲۸	۰/۵۴	تکمیل نقشه مفهومی شبکه‌ای
۰/۰۰۱	۲۹	۰/۶۸	ترسیم نقشه مفهومی سلسله‌مراتبی
۰/۱۵	۲۵	۰/۳۰	ترسیم نقشه مفهومی شبکه‌ای

همان‌گونه که در جدول ۴ مشخص است آزمون ترسیم نقشه مفهومی سلسله‌مراتبی دارای بالاترین مقدار ضریب همبستگی پیرسون با آزمون چهارگزینه‌ای می‌باشد. کمترین مقدار ضریب همبستگی نیز متعلق به آزمون ترسیم نقشه مفهومی شبکه‌ای است که معنادار نیز نمی‌باشد. برای اطمینان از روایی افتراقی (تمییزی) آزمون‌های مختلف مبتنی بر نقشه مفهومی در پژوهش حاضر توان این آزمون‌ها در تمایز گروه‌های بالا و پایین پیشرفت تحصیلی (براساس میانگین نمرات دانش‌آموزان در آزمون چهارگزینه‌ای) مورد بررسی قرار گرفت. از یک آزمون t دونمونه‌ای مستقل نیز برای بررسی تفاوت میانگین دو گروه بالا و پایین در هر کدام از آزمون‌های مبتنی بر نقشه مفهومی استفاده شد.

جدول ۵. آزمون t دونمونه‌ای مستقل برای بررسی تفاوت میانگین گروه دارای پیشرفت تحصیلی بالا و پایین در آزمون‌های مختلف مبتنی بر نقشه مفهومی

تفاوت میانگین	معناداری	درجه آزادی	t	نوع آزمون
۴/۵۷	۰/۰۰۴	۱۴/۵۷	۳/۴۴	تکمیل نقشه مفهومی سلسله‌مراتبی
۳/۴۲	۰/۰۱۵	۱۵/۰۴	۲/۷۳	تکمیل نقشه مفهومی شبکه‌ای
۴/۶۳	۰/۰۰۲	۲۷	۳/۳۷	ترسیم نقشه مفهومی سلسله‌مراتبی
۲/۵	۰/۱۲	۲۳	۱/۶۱	ترسیم نقشه مفهومی شبکه‌ای

همان‌گونه که در جدول ۵ مشخص است بالاترین مقدار شاخص t متعلق به آزمون تکمیل نقشه مفهومی سلسله‌مراتبی و پایین‌ترین مقدار شاخص t نیز متعلق به آزمون ترسیم نقشه مفهومی شبکه‌ای می‌باشد که تنها آزمونی است که نتوانسته بین دو گروه دارای پیشرفت تحصیلی بالا و پایین تمییز ایجاد کند.

برای بررسی پایایی آزمون‌های تکمیل نقشه مفهومی از روش کودر - ریچاردسون ۲۱ استفاده شد؛ که ضریب به‌دست‌آمده برای آزمون تکمیل نقشه مفهومی سلسله‌مراتبی برابر با ۰/۸۵ و برای آزمون تکمیل نقشه مفهومی شبکه‌ای نیز برابر با ۰/۷۹ بود.

به‌منظور بررسی پایایی آزمون‌های ترسیم نقشه مفهومی نیز از روش پایایی بین نمره‌گذاران^۱ استفاده شد. برای این منظور ضریب همبستگی بین نمرات هشت نمره‌گذار محاسبه شد.

جدول ۶. ضرایب همبستگی نمرات داوران در دو آزمون ترسیم نقشه مفهومی سلسله‌مراتبی و شبکه‌ای

معناداری	تعداد	ضریب همبستگی پیرسون	آزمون
۰/۰۰۰۱	۲۹	۰/۹۵	ترسیم نقشه مفهومی سلسله‌مراتبی
۰/۰۰۰۱	۲۵	۰/۹۸	ترسیم نقشه مفهومی شبکه‌ای

1. Inter-rater reliability

همان‌گونه که در جدول ۶ مشخص است هر دو آزمون ترسیم نقشه مفهومی سلسله‌مراتبی و شبکه‌ای دارای ضرایب پایایی بین نمره‌گذاران بالا و معناداری می‌باشند. در جدول شماره ۷ نیز ضرایب پایایی هر چهار آزمون مبتنی بر نقشه مفهومی ارائه شده است:

جدول ۷. پایایی هر چهار آزمون مبتنی بر نقشه مفهومی

تکمیل نقشه مفهومی سلسله‌مراتبی	تکمیل نقشه مفهومی شبکه‌ای	ترسیم نقشه مفهومی سلسله‌مراتبی	ترسیم نقشه مفهومی شبکه‌ای
۰/۸۵	۰/۷۹	۰/۹۵	۰/۹۸

همان‌گونه که در جدول ۷ مشاهده می‌شود آزمون ترسیم نقشه مفهومی شبکه‌ای دارای بالاترین ضریب پایایی و آزمون تکمیل نقشه مفهومی شبکه‌ای دارای پایین‌ترین ضریب پایایی است. برای انتخاب مناسب‌ترین آزمون مبتنی بر نقشه مفهومی از بین چهار آزمون، آزمونی که از لحاظ شاخص‌های روایی محتوایی، روایی ملاکی، روایی افتراقی و پایایی ضرایب بالاتری داشته باشد به‌عنوان مناسب‌ترین آزمون انتخاب شد.

جدول ۸. مقادیر شاخص‌های روایی محتوایی، ملاکی، افتراقی و پایایی آزمون‌های مبتنی بر نقشه مفهومی

نوع آزمون	روایی محتوایی	روایی ملاکی	روایی افتراقی	پایایی
تکمیل نقشه مفهومی سلسله‌مراتبی	۸۸/۱۳	۰/۶۱	۳/۴۴	۰/۸۵
تکمیل نقشه مفهومی شبکه‌ای	۸۷/۵۰	۰/۵۴	۲/۷۳	۰/۷۹
ترسیم نقشه مفهومی سلسله‌مراتبی	۸۴/۳۸	۰/۶۸	۳/۳۷	۰/۹۵
ترسیم نقشه مفهومی شبکه‌ای	۸۳/۱۳	۰/۳۰	۱/۶۱	۰/۹۸

همان‌گونه که در جدول ۸ مشخص است آزمون‌های مختلف مبتنی بر نقشه مفهومی دارای مقادیر متفاوتی از لحاظ روایی محتوایی، روایی ملاکی، روایی افتراقی و پایایی می‌باشند. با مقایسه این چهار آزمون متوجه می‌شویم که آزمون تکمیل نقشه مفهومی سلسله‌مراتبی در دو شاخص روایی محتوایی و روایی افتراقی بالاتر از سایر آزمون‌هاست. ضعیف‌ترین شاخص‌ها نیز مربوط به آزمون ترسیم نقشه مفهومی شبکه‌ای می‌باشد. این آزمون در هر سه شاخص روایی محتوایی، روایی ملاکی و روایی افتراقی دارای کمترین ضرایب است.

بحث و نتیجه‌گیری

هدف پژوهش حاضر بررسی شاخص‌های روایی و پایایی آزمون‌های مبتنی بر نقشه مفهومی بود. نتایج پژوهش نشان داد که این آزمون‌ها دارای مقادیر متفاوتی از این شاخص‌ها هستند. یافته‌های پژوهش حاضر به‌طور کلی نشان داد که چهار آزمون مبتنی بر نقشه مفهومی از روایی محتوایی بالایی

برخوردارند. در زمینه روایی ملاکی، نتایج پژوهش نشان داد که چهار آزمون مبتنی بر نقشه مفهومی دارای مقادیر متفاوت همبستگی از ۰/۳۰ تا ۰/۶۸ با یک آزمون چهارگزینه‌ای پیشرفت تحصیلی می‌باشند. در زمینه روایی افتراقی سه آزمون تکمیل نقشه مفهومی سلسله‌مراتبی، ترسیم نقشه مفهومی سلسله‌مراتبی و تکمیل نقشه مفهومی شبکه‌ای به‌خوبی از عهده تشخیص تفاوت بین نمرات دو گروه پیشرفت تحصیلی بالا و پایین برآمدند اما آزمون ترسیم نقشه مفهومی شبکه‌ای از این توان تمیز برخوردار نبود. در زمینه پایایی نیز هر چهار آزمون مبتنی بر نقشه مفهومی از پایایی بالا و مناسبی برخوردار بودند.

نتایج پژوهش حاضر نشان داد که آزمون‌های مبتنی بر نقشه مفهومی از روایی محتوایی قابل قبولی برخوردار می‌باشند که با نتایج پژوهش مصرآبادی (۱۳۸۸) و هانگ و لین (۲۰۱۵) همسوست. در تبیین این یافته می‌توان اظهار داشت که در آزمون‌های ترسیم نقشه مفهومی (ترسیم نقشه‌های مفهومی شبکه‌ای و ترسیم نقشه‌های مفهومی سلسله‌مراتبی)، مفاهیم مهم حوزه موردنظر توسط متخصصان از قبل ارائه شده بود تا فراگیران آن‌ها را در نقشه ترسیمی خود بگنجانند. همین امر خود موجب به‌کارگیری این مفاهیم که محتوای مهم حوزه مورد آزمون را تشکیل می‌دهند، شد. در زمینه آزمون‌های تکمیل (تکمیل نقشه‌های مفهومی شبکه‌ای و تکمیل نقشه‌های مفهومی سلسله‌مراتبی) نیز مشخص است که خود این نقشه‌ها از قبل توسط محققان و متخصصان ساخته شده و مفاهیم و روابط مهم بین مفاهیم در آن‌ها لحاظ شده بود؛ بنابراین استفاده از نظرات متخصصان و کارشناسان در طراحی نقشه‌های مفهومی می‌تواند در افزایش روایی محتوایی تأثیرگذار بوده باشد. در همین رابطه، محققان اظهار داشته‌اند که تفاوت مهمی بین دانش محتوایی افراد متخصص و افراد تازه‌کار وجود دارد و دانش محتوایی متخصصان عمیق‌تر، منسجم‌تر و پیوسته‌تر است (کلاین، چانگ، اوسموندسون، هرل و اونیل^۱، ۲۰۰۲).

شواهد بیشتر برای تأیید روایی محتوایی نقشه‌های مفهومی می‌تواند توسط تحلیل‌های شناختی^۲ بعضی از محققان به‌دست آید. تحلیل‌های شناختی مانند «بسط و گسترش نوشتاری»^۳ و «تفکر با صدای بلند»^۴ درباره گزاره‌های یک نقشه مفهومی توانستند نه تنها صحت محتوا بلکه عمق و پیچیدگی درک و فهم یک فراگیر را نیز آشکار سازند (میساه^۵، ۲۰۱۰). برای مثال در روش بسط و گسترش نوشتاری از آزمودنی‌ها خواسته می‌شد تا یک یا چند تا از گزاره‌هایی که در نقشه‌های مفهومی ساخته بودند (و در محتوای متنی که مورد نقشه‌کشی قرار گرفته بود) را شرح داده و به‌صورت نوشتاری (کتبی) در یک یا چند بند توضیح دهند تا درک فراگیران را از محتوا مورد

1. Klein, Chung, Osmundson, Herl & O'Neil
2. Cognitive Analyses
3. written elaborations
4. think-aloud protocols
5. Mesa

وارسی قرار دهند. در روش تفکر با صدای بلند نیز از آزمودنی‌ها خواسته می‌شد تا هنگام انجام تکالیف نقشه مفهومی با صدای بلند فکر کنند. سپس این فعالیت به صورت صوتی یا تصویری ضبط شده و مورد تجزیه و تحلیل قرار می‌گرفت تا فرایندهای تفکری را که پشتیبان نقشه مفهومی ترسیم شده و گزاره‌های بکار رفته در آن بود کشف کنند؛ بنابراین، با این شیوه‌ها میزان درک شاگردان از محتوا مورد تحلیل قرار می‌گرفت و این تحلیل‌های شناختی توانستند میزان درک یا کج‌فهمی‌های عمیق یادگیرندگان را از محتوای مورد نظر آشکار سازند. علاوه بر این، تحلیل‌های شناختی این موضوع را که چگونه تکلیف نقشه مفهومی توانایی فراگیر را در بیان درک و فهمش محدود می‌سازد نیز به‌طور کامل نشان دادند.

در زمینه روایی ملاکی نیز نتایج پژوهش حاضر نشان از روایی ملاکی مناسب آزمون‌های مبتنی بر نقشه مفهومی داشت که با نتایج پژوهش مصرآبادی (۱۳۸۸) و اکین و همکاران (۲۰۱۶) همسو و با نتایج پژوهش ون و همکاران (۲۰۱۷) ناهمسو است. در تبیین روایی ملاکی مناسب سه آزمون مبتنی بر نقشه مفهومی (تکمیل نقشه مفهومی سلسله‌مراتبی، تکمیل نقشه مفهومی شبکه‌ای، ترسیم نقشه مفهومی سلسله‌مراتبی) می‌توان به نتایج پژوهش وست، پارک، پومروی و ساندوال^۱ (۲۰۰۲) استناد کرد. این پژوهشگران به منظور برآورد روایی ملاکی (از نوع پیش‌بینی) نمرات نقشه‌های مفهومی را پیش و پس از یک مداخله آموزشی مورد مقایسه قرار دادند. مقایسه کیفی نقشه‌های مفهومی پیش و پس از مداخله به‌خوبی تلفیق مفاهیم جدید به داخل نقشه‌ها، تمایز بیشتر مفاهیم موجود و روابط عرضی (همبرها) جدید را نشان داد. نتایج کمی پژوهش آن‌ها نیز نشان داد که نمرات نقشه‌های مفهومی پس از آموزش به‌صورت معناداری افزایش یافته‌اند هر چند که این معناداری در روش نمره‌دهی ساختاری مشاهده شد (که نشان‌دهنده تأثیر روش نمره‌دهی بر روایی ملاکی است). در همین رابطه، لیو و هینچی^۲ (۱۹۹۶) همبستگی بین نمرات نقشه‌های مفهومی و نمرات آزمون‌های سنتی را با استفاده از روش همسانی درونی بین مؤلفه‌های نقشه مفهومی (مانند سطوح سلسله‌مراتب، مثال‌ها و ...) تبیین کرده‌اند. آن‌ها اظهار داشتند که زمانی که همسانی درونی مؤلفه‌های نقشه مفهومی بالا باشد همبستگی بین نمرات نقشه مفهومی و نمرات آزمون‌های سنتی نیز بالاست و زمانی که همسانی درونی بین مؤلفه‌ها پایین باشد این همبستگی پایین خواهد بود. رایس، رایان و سامسون^۳ (۱۹۹۸) نیز اظهار داشتند زمانی که نمرات نقشه‌های مفهومی فراگیران فقط با روش گزاره‌های درست یا غلط به دست آید همبستگی بالایی (از ۰/۸۲ تا ۰/۸۷) با آزمون‌های استاندارد ملی و ایالتی نشان می‌دهند که نشان‌دهنده این است که نمرات نقشه‌های مفهومی روایی ملاکی (از نوع همزمان) بالایی دارند. آن‌ها همچنین اظهار داشتند که روش نمره‌دهی نقشه‌های مفهومی - به‌ویژه معیارهای مختص به نمره‌دهی - تعیین‌کننده مهم نیرومندی رابطه بین

1. West, Park, Pomeroy & Sandoval

2. Liu & Hinchey

3. Rice, Ryan & Samson

نمرات نقشه مفهومی و نمرات دیگر ابزارهای ارزشیابی است. برای مثال زمانی که سطوح سلسله مراتب از معیارهای نمره‌دهی نقشه‌های مفهومی حذف می‌شود همبستگی بین نمرات نقشه‌های مفهومی و نمرات آزمون‌های سنتی به شکل دراماتیکی افزایش می‌یابد. این موضوع می‌تواند در بالا بردن روایی ملاکی در پژوهش حاضر نیز تأثیرگذار بوده باشد چراکه در پژوهش حاضر مؤلفه همبرها از نقشه‌های مفهومی آزمودنی‌ها کنار گذاشته شد.

در رابطه با روایی ملاکی پایین آزمون‌های ترسیم نقشه مفهومی شبکه‌ای نیز باید عنوان کرد که ساختار شبکه‌ای مانند نقشه‌های مفهومی شبکه‌ای کاملاً متفاوت با دیگر انواع آزمون‌های مبتنی بر نقشه مفهومی می‌باشد. در واقع ساختار پیچیده‌تر این نقشه‌ها می‌طلبد تا هنگام ترسیم آن‌ها از توانایی ایجاد روابط مفهومی متقابل و سازمان‌دهی مفاهیم استفاده شود که متفاوت با سایر آزمون‌های مبتنی بر نقشه مفهومی است؛ همچنین در آزمون‌های ترسیم نقشه مفهومی شبکه‌ای فراگیران آزادی عمل بیشتری در هنگام آزمون دارند و محدودیت‌های کمتری برای آن‌ها وجود دارد که باعث می‌شود آن‌ها کمتر در راستای چارچوب فکری ارزیاب (آزمون‌گیرنده) و انتظارات او حرکت کنند؛ بنابراین از یک سو قابل‌انتظار است که نمره‌های چنین آزمون ساختار شکنی (آزمون ترسیم نقشه مفهومی شبکه‌ای) متفاوت از نمره‌های دیگر انواع آزمون‌ها باشد و از سوی دیگر پایین بودن همبستگی بین نمرات آزمون‌های ترسیم نقشه مفهومی شبکه‌ای با نمرات آزمون‌های سنتی نمی‌تواند امتیازی منفی برای آزمون‌های ترسیم نقشه مفهومی شبکه‌ای محسوب شود و روایی ملاکی آن‌ها را خدشه‌دار سازد چراکه اولاً برخلاف آزمون‌های سنتی هیچ‌گاه دو تکلیف ترسیم نقشه مفهومی کاملاً یکسان و مشابه نخواهند بود و ثانیاً نمرات ارزشیابی‌های مبتنی بر نقشه مفهومی هرگز به‌صورت کامل با ارزشیابی‌های سنتی همبسته نیستند زیرا آزمون‌های مبتنی بر نقشه مفهومی به‌گونه‌ای مستقیم‌تر درک ساختاری و سازمان‌یافته شاگردان را اندازه می‌گیرند (نواک و گوین^۱، ۱۹۸۴). نواک، گوین و یوهانسن^۲ (۱۹۸۳) نیز عنوان کردند که همبستگی پایین بین نمرات آزمون‌های ترسیم نقشه مفهومی و نمرات آزمون‌های سنتی گویای این موضوع است که نقشه‌های مفهومی و آزمون‌های سنتی جنبه‌های متفاوتی از توانایی‌های فراگیران را اندازه می‌گیرند. به‌خصوص آن‌ها اظهار داشتند که نقشه‌های مفهومی *توانایی‌های سطح بالاتر*^۳ (مانند تحلیل، ترکیب و ارزشیابی) را مورد ارزشیابی قرار می‌دهند در حالی که آزمون‌های سنتی *توانایی‌های سطح پایین‌تر*^۴ (مانند دانش، درک و فهم و کار بستن) را مورد ارزشیابی قرار می‌دهند.

نتایج پژوهش حاضر در زمینه روایی افتراقی نشان داد که سه آزمون تکمیل نقشه مفهومی سلسله‌مراتبی، تکمیل نقشه مفهومی شبکه‌ای و ترسیم نقشه مفهومی سلسله‌مراتبی می‌توانند بین

-
1. Novak & Gowin
 2. Johansen
 3. higher-order abilities
 4. lower-order abilities

دانش‌آموزان دارای پیشرفت تحصیلی بالا و دانش‌آموزان دارای پیشرفت تحصیلی پایین تمییز قائل شوند که با نتایج پژوهش مصرآبادی (۱۳۸۸) و کراوالیو (۲۰۱۰) همسوست. همچنین تنها آزمونی که در این پژوهش روایی افتراقی مناسبی را نشان نداد آزمون ترسیم نقشه مفهومی شبکه‌ای بود. در رابطه با روایی افتراقی مناسب آزمون‌های مبتنی بر نقشه مفهومی نتایج پژوهش داگوسای^۱ (۲۰۱۲) نشان داد که تفاوت آشکاری بین کیفیت نقشه‌های مفهومی افراد تازه‌کار و افراد خبره وجود دارد. نتایج وی نشان داد که تعداد مفاهیم، رابطه‌ها، همبرها و مثال‌ها در نقشه‌های افراد خبره بیشتر از افراد تازه‌کار است و در مجموع کیفیت نقشه‌های مفهومی افراد خبره مناسب‌تر و بهتر است. همچنین در زمینه غنای محتوایی نیز نقشه‌های مفهومی افراد خبره غنای بیشتری از نقشه‌های مفهومی افراد تازه‌کار دارند. پژوهش‌های دیگری نیز به شواهد کیفی درباره روایی افتراقی آزمون‌های مبتنی بر نقشه مفهومی دست یافته‌اند تا ساخت معانی اولیه و فرایند پیشرفت معناسازی را شرح دهند. برای نمونه، واکر و کینگ^۲ (۲۰۰۳) در پژوهش خود به مقایسه نقشه‌های مفهومی ساخت اساتید دانشگاه و نقشه‌های مفهومی ساخت دانشجویان پرداختند. آن‌ها به این نتیجه رسیدند که اساتید شبکه‌های متراکمی از مفاهیم و کاربردها را در نقشه‌های مفهومی‌شان به وجود آوردند درحالی‌که دانشجویان رابطه‌های اندکی را میان مفاهیم مربوط به محتوا برقرار ساختند. به‌علاوه، نقشه‌های مفهومی دانشجویان در طول زمان از لحاظ کیفی متغیر بودند، یعنی نقشه‌های مفهومی که بعداً ساخته می‌شدند واژه‌های دقیق‌تری داشتند، به‌صورت یکپارچه‌تر و منسجم‌تری سازمان‌یافته بودند و رابطه‌های بین مفاهیم بیشتری را در برمی‌گرفتند؛ بنابراین، آن‌ها نتیجه گرفتند که نقشه‌های مفهومی ابزارهای سودمندی برای به تصویر کشیدن فرایند تغییر شکل دانش از سطح مبتدی به سطح خبره هستند.

در رابطه با توان تفکیک پایین آزمون‌های ترسیم نقشه مفهومی شبکه‌ای نیز شاید بتوان گفت با توجه به اینکه در مبحث روایی ملاکی در پژوهش حاضر این آزمون با آزمون چهارگزینه‌ای (به‌عنوان یکی از روش‌های ارزشیابی سنتی) چندان همبستگی و سازگاری نشان نداد، استفاده از نقطه برش در نمرات آزمون چهارگزینه‌ای به‌عنوان معیار تفکیک دو گروه دارای پیشرفت تحصیلی بالا و پایین چندان مناسب آزمون‌های ترسیم نقشه مفهومی شبکه‌ای نباشد. لذا، شاید اگر از معیار تفکیک دیگری استفاده می‌شد نتیجه متفاوتی نیز در رابطه با روایی تفکیکی آزمون ترسیم نقشه مفهومی شبکه‌ای به دست می‌آمد.

بخش دیگری از نتایج نیز نشان داد که نقشه‌های مفهومی از پایایی مناسب و بالایی برخوردارند که با نتایج پژوهش مصرآبادی (۱۳۸۸)، اکین و همکاران (۲۰۱۶) و هانگ و لین (۲۰۱۵) همسوست. در رابطه با پایایی بالای نقشه‌های مفهومی، محققان اظهار داشته‌اند که زمانی که تمام

1. Dogusoy

2. Walker & King

فراگیران تکالیف یکسانی را انجام می‌دهند و فرایندهای نمره‌دهی به‌خوبی تعریف می‌شوند پایایی نیز افزایش می‌یابد؛ به‌علاوه زمانی که از طرح‌ها و روش‌های مشخص نمره‌دهی استفاده می‌شود، میزان پایایی بالاتر می‌رود چراکه این طرح‌ها می‌توانند محدودیت‌هایی را بر روی نمره‌دهی ارزشیابی‌ها اعمال نموده و در بیشتر موارد پایایی را افزایش دهند. طرح‌ها زمانی که قابل تجزیه و تحلیل بوده، موضوع مشخصی داشته و با نمونه‌ها و مثال‌ها همراه باشند، اثر بیشتری بر روی افزایش پایایی دارند (جانسون و سوینگبی^۱، ۲۰۰۷)؛ علاوه بر این، استفاده از نمره‌گذاران تعلیم‌دیده در زمینه استفاده از معیارهای نمره‌دهی نقشه‌های مفهومی می‌تواند در به دست آوردن ضرایب پایایی بالا تأثیرگذار باشد (اسنلسون، ۲۰۱۰). مسأله دیگر میزان آشنایی نمره‌گذاران نقشه‌های مفهومی با فرایند ارزشیابی نقشه‌های مفهومی است؛ چراکه پژوهشگران اظهار داشته‌اند که استفاده از ارزیابان و نمره‌گذارانی که به‌خوبی آموزش‌دیده باشند و کاملاً با فرایند ارزشیابی نقشه‌های مفهومی آشنا باشند در بالا بردن میزان پایایی تأثیرگذار است (کساب و حسین^۲، ۲۰۱۰؛ و اسنلسون، ۲۰۱۰) چراکه آشنایی با هدف طرح، محتوای دوره و سطوح توانایی آزمودنی‌ها می‌تواند در کاهش بار شناختی تحمیل‌شده روی نمره‌گذاران نقشه‌های مفهومی (در اثر توجه همزمان به انتخاب گزاره‌ها، صحت گزاره‌ها و دیگر عناصر نمره‌دهی در نقشه‌های مفهومی) کاملاً مؤثر باشد (پلومر، ۲۰۰۸). اسنلسون (۲۰۱۰) نیز به‌طور مستند نشان داد که در آزمون‌های مبتنی بر نقشه مفهومی پایایی میانگین نمره‌گذاری‌ها برای هر آزمودنی تابعی از تعداد نمره‌گذاران نقشه‌های مفهومی (و به نسبت کمتر تعداد موقعیت‌های نمره‌دهی^۳) است؛ به‌عبارت‌دیگر هر چه تعداد ارزیابان (نمره‌گذاران) نقشه‌های مفهومی افزایش یابد پایایی آزمون‌های مبتنی بر نقشه مفهومی نیز به همان نسبت افزایش خواهد یافت که با توجه به تعداد زیاد نمره‌گذاران در پژوهش حاضر این توجیه قابل قبول به نظر می‌رسد.

روی‌هم‌رفته، یافته‌های پژوهش حاضر نشان داد که چهار آزمون مبتنی بر نقشه مفهومی از لحاظ شاخص‌های روایی محتوایی، روایی ملاکی، روایی تمییزی و پایایی تفاوت‌های قابل‌ملاحظه‌ای با یکدیگر دارند. در زمینه روایی محتوایی و روایی افتراقی آزمون تکمیل نقشه مفهومی سلسله‌مراتبی شرایط بهتری نسبت به دیگر آزمون‌ها داشت. در زمینه روایی ملاکی آزمون ترسیم نقشه مفهومی سلسله‌مراتبی نسبت به دیگر آزمون‌ها مناسب‌تر به نظر می‌رسد. در زمینه پایایی نیز آزمون ترسیم نقشه مفهومی شبکه‌ای شرایط بهتری دارد. این یافته‌ها در راستای مبانی نظری و تجربی موجود در زمینه ارزشیابی‌های مبتنی بر نقشه مفهومی است. براساس نظرات روئیز-پریمو و شیولسون (۱۹۹۶) آزمون‌های مختلف مبتنی بر نقشه مفهومی می‌توانند جنبه‌های متفاوتی از دانش فراگیران را مورد اندازه‌گیری قرار دهند؛ لذا به جهت همین جنبه‌های متفاوت دانش فراگیران و نیز نوع تکلیفی که

1. Jonsson & Svingby

2. Kassab & Hussain

۳. منظور از موقعیت، نمره‌گذاری در زمان‌های مختلف است.

آزمون به فراگیر تحمیل می‌کند، شاخص‌های روایی و پایایی نیز می‌توانند متفاوت باشند؛ بنابراین می‌توان گفت که الزاماً بالا بودن شاخص‌های روایی و پایایی در تعدادی از این آزمون‌ها ناشی از ویژگی‌های خاص آن‌هاست (مصرآبادی، ۱۳۸۸).

در مجموع یافته‌های پژوهش حاضر از ایده استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی حمایت می‌کنند. این ایده فرایند ارزشیابی را به فرایند یادگیری پیوند می‌دهد (کو^۱، ۲۰۰۷). استفاده از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی جانشین ارزشیابی‌های سنتی، پتانسیل و جذابیت زیادی دارد. در همین زمینه، هیئت کنترل ارزشیابی ملی وزارت آموزش و پرورش آمریکا^۲ تأیید کرد که ارزشیابی مبتنی بر نقشه مفهومی می‌تواند از توانایی‌های علمی که سنجش آن‌ها توسط دیگر ابزارها مشکل است، بهره‌برداری کند؛ بنابراین این هیئت استفاده کلاسی و پژوهش‌های اکتشافی بیشتر در زمینه ارزشیابی مبتنی بر نقشه مفهومی را تشویق کرده است (سیپرین‌ساس، ۲۰۱۰). در همین راستا از سال ۲۰۰۹ میلادی نقشه‌های مفهومی به‌منظور ارزشیابی‌های استاندارد مقیاس وسیع^۳ در مرکز ملی ارزشیابی پیشرفت آموزشی ایالات متحده آمریکا^۴ (NAEP) مورد استفاده قرار گرفته‌اند تا تغییرات در درک مفهومی ایده‌های علوم طبیعی را مورد سنجش قرار دهند (شوندیمان، ۲۰۱۵). پژوهش حاضر نشان داد که امکان استفاده از نقشه‌های مفهومی برای اهداف ارزشیابی حداقل از لحاظ فنی وجود دارد. نتایج مثبت درباره روایی و پایایی روش‌های ارزشیابی مبتنی بر نقشه مفهومی به افزایش اطمینان برای استفاده بیشتر از ارزشیابی‌های مبتنی بر نقشه مفهومی کمک خواهد کرد؛ بنابراین ضروری است که مسئولان مربوطه در سیستم آموزش و ارزشیابی کشور نیز نگاه ویژه و متفاوتی به نقش کاربردی نقشه‌های مفهومی در زمینه استفاده شدن به‌عنوان یک ابزار ارزشیابی جدید و جایگزین داشته باشند. همچنین این ابزار ارزشیابی می‌تواند رابطه بین فرایند ارزشیابی آموزشی و علوم شناختی را که بعضی از محققان (میسلوی^۵، ۲۰۰۸) آن را به‌عنوان «شکاف در حال گسترشی^۶» که می‌بایست پل زده شود در نظر گرفته‌اند، برقرار سازد.

محدودیت‌ها:

۱. به آزمودنی‌ها بازخوردی درباره ارزشیابی مبتنی بر نقشه مفهومی داده نشد. چراکه ارزشیابی و نمره‌گذاری نقشه‌های مفهومی بعد از فرایند ارزشیابی آزمودنی‌ها صورت گرفت.

1. Ku

2. National Assessment Governing Board of the U.S. Department of Education

3. standardized large-scale assessments

4. U.S. National Assessment of Educational Progress

5. Mislevy

6. widening chasm

۲. آموزش نحوه تکمیل یا ترسیم نقشه‌های مفهومی در پژوهش حاضر نسبتاً کوتاه‌مدت بود که این موضوع می‌تواند در عملکرد ضعیف تعدادی از آزمودنی‌ها در آزمون‌های مختلف مبتنی بر نقشه مفهومی مؤثر بوده باشد.
۳. به دلیل اینکه آزمودنی‌های متفاوتی در هر یک از چهار گروه بکار گرفته شدند راهی وجود نداشت تا یادگیری مهارت‌های نقشه‌کشی یا موفقیت در آزمون‌های مبتنی بر نقشه مفهومی در طول زمان مورد اندازه‌گیری قرار گیرد.
۴. نمره‌گذاری آزمون‌های مبتنی بر نقشه مفهومی به‌ویژه در شیوه‌های ترسیم نقشه مفهومی ممکن است تا حدی تحت تأثیر ذهنیت نمره‌گذاران بوده باشد.
۵. در پژوهش حاضر فقط مؤلفه‌های نقشه و گزاره‌ها مورد ارزشیابی قرار گرفتند و دیگر نظام‌های نمره‌گذاری نقشه‌های مفهومی (مانند نظام نمره‌گذاری کل‌نگر، نظام‌های نمره‌گذاری کامپیوتری و ...) برای ارزشیابی درک کلی شاگردان از مفاهیم مورد استفاده قرار نگرفت.
۶. از جمله محدودیت‌های پژوهش حاضر عدم امکان استفاده از مقیاس‌های عینی (مانند روش ردیابی حرکات چشمی و ...) برای اندازه‌گیری بار شناختی بود.
۷. جامعه آماری این پژوهش محدود به دانش‌آموزانی از یک فرهنگ خاص (با زبانی خاص) بود، لذا تعمیم‌پذیری نتایج به پژوهش‌های بیشتری در بافت‌های تحصیلی دیگر و ملاحظه دامنه وسیعی از ویژگی‌های محیط یادگیری وابسته است.

پیشنهادات کاربردی:

۱. با توجه به روایی و پایایی مطلوب نقشه‌های مفهومی، معلمان، مربیان، اساتید و مسئولان امور ارزشیابی کشور می‌توانند از نقشه‌های مفهومی به‌عنوان یکی از روش‌های ارزشیابی جایگزین جهت ارزشیابی هر چه بهتر فرایند یاددهی-یادگیری استفاده کنند. این امر می‌تواند منجر به این شود که فراگیران با روش‌های متفاوت و جدید ارزشیابی آشنا شده و به آن‌ها عادت کنند. در واقع استفاده از نقشه‌های مفهومی به‌عنوان ابزار ارزشیابی تحصیلی خود مصداقی از توجه به علوم شناختی در فرایند ارزشیابی تحصیلی است.
۲. با توجه به اینکه در پژوهش حاضر نقشه‌های مفهومی توانستند تعدادی از کج‌فهمی‌های آزمودنی‌ها را به تصویر بکشند؛ لذا معلمان و مربیان می‌توانند از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی تکوینی (جهت پی بردن به کج‌فهمی‌ها و نقاط قوت و ضعف یادگیری شاگردان و اصلاح برنامه آموزشی) نیز استفاده کنند.
۳. پیشنهاد می‌شود که آزمون‌های تکمیل نقشه مفهومی به‌صورت گسترده‌تری در فرایند ارزشیابی تحصیلی مورد استفاده قرار گیرند چراکه براساس نتایج پژوهش حاضر این آزمون‌ها نسبت به دیگر آزمون‌های مبتنی بر نقشه مفهومی دارای روایی و پایایی مناسب‌تری بوده و بار شناختی کمتری بر روی آزمودنی‌ها تحمیل می‌کردند.

۴. پیشنهاد می‌شود مؤلفین و برنامه‌ریزان کتب درسی (به‌ویژه کتب زیست‌شناسی) در پایان هر درس، فصل یا بحث (به‌جای آزمون‌های چندگزینه‌ای یا تشریحی و یا همراه با آن‌ها) از یکی از آزمون‌های مبتنی بر نقشه مفهومی استفاده کنند. این امر می‌تواند بازخورد مناسبی از یادگیری شاگردان فراهم سازد که می‌تواند هم برای مربیان و هم برای خود شاگردان مفید باشد.
۵. پیشنهاد می‌شود دوره‌های آموزشی ضمن خدمتی با محوریت آموزش و آشنایی با آزمون‌های مبتنی بر نقشه مفهومی برای معلمان، مربیان و یا شاگردان اجرا شود.

تشکر و قدردانی

از اساتید محترم، معلمان و همراهانی که در به سرانجام رساندن این پژوهش نقش بسزایی داشته و با راهنمایی‌های ارزشمندشان در تمام مراحل انجام پژوهش همیار ما بوده‌اند کمال تشکر و قدردانی را داریم.

منابع

- Amadiou, F., Tricot, A., & Mariné, C. (2010). Interaction between prior knowledge and concept-map structure on hypertext comprehension, coherence of reading orders and disorientation. *Interacting with computers*, 22(2), 88-97.
- Arneson, B. T. (2005). *On the role of concept mapping assessment in today's constructivist classroom*. Doctoral dissertation, The University of Texas at Austin.
- Ciprian-Sas, I. (1999). *The multiple-choice concept map (MCCM): an interactive computer-based assessment method*. Doctoral dissertation, University of Nevada, Las Vegas.
- Dugosuy, B. (2012). *Cognitive analysis of experts and novices concept mapping processes*. Doctoral dissertation, Middle East Technical University.
- Ekin, B., Ulcinar Sagir, S., & Saltan, F. (2016). The Comparison on Evaluation of Concept Map and Structured Grid with Multiple-Choice Test. *Participatory Educational Research (PER)*, Special Issue 2016-II, 100-111.
- Hung, C. H., & Lin, C. Y. (2015). Using concept mapping to evaluate knowledge structure in problem-based learning. *BMC medical education*, 15(1), 1-9.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, 2(2), 130-144.
- Joseph, C., Conradsson, D., Wikmar, L. N., & Rowe, M. (2017). Structured feedback on students' concept maps: the proverbial path to learning?. *BMC medical education*, 17(1), 1-9.
- Kassab, S. E., & Hussain, S. (2010). Concept mapping assessment in a problem-based medical curriculum. *Medical teacher*, 32(11), 926-931.
- Klein, D. C., Chung, G. K., Osmundson, E., Herl, H. E., & O'Neil, H. F., Jr. (2002). *The validity of knowledge mapping as a measure of elementary students' scientific understanding* (CSE Tech. Rep. No. 557). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Ku, W. A. (2007). *Using concept maps to explore the conceptual knowledge of technology students: an exploratory study*. Doctoral dissertation, The Ohio State University.
- Liping, Z., Yan, W., Boqing, D. & Zengyao, Z. (2009). The comparison study of Chinese and American secondary school students' knowledge structure assessment technique - an experimental research based on concept map. *Frontiers of Education in China*, 4(2), 286-297.
- Liu, X. & Hinchey, M. (1996). The internal consistency of a concept mapping scoring scheme and its effect on prediction validity. *International Journal of Science Education*, 18(8), 921-937.
- Liu, T. C., Lin, Y. C., & Tsai, C. C. (2009). Identifying senior high school students' misconceptions about statistical correlation, and their possible causes: an exploratory study using concept mapping with interviews. *International Journal of Science and Mathematics Education*, 7(4), 791-820.

- Maker, C. J., & Zimmerman, R. H. (2020). Concept maps as assessments of expertise: Understanding of the complexity and interrelationships of concepts in science. *Journal of Advanced Academics*, 31(3), 254-297.
- Mesa, J. C. (2010). *The development and use of a concept mapping assessment tool with young children on family visits to a live butterfly exhibit*. Doctoral dissertation, University of Florida.
- Mesrabadi, J. (2008). *Feasibility study of using concept map-based evaluations in high school biology*. Ministry of education, Research and Curriculum Planning Organization, Curriculum Planning And Educational Innovation Research institute. (In Persian).
- Mislevy, R. J. (2008). How cognitive science challenges the educational measurement tradition. *Measurement*, 6(2), 124-141.
- Nasri, N., Roslan, S. N., Sekuan, M. I., Bakar, K. A., & Puteh, S. N. (2010). Teachers' perception on alternative assessment. *Procedia-Social and behavioral sciences*, 7, 37-42.
- Novak, J. D., & Gowin, B. (1984). *Learning how to learn*. New York: Cambridge University Press.
- Novak, J. D., Gowin, D. B. & Johansen, G. T. (1983). The use of concept mapping and Knowledge vee mapping with junior high school science students. *Science Education*, 67(5), 625-645.
- Plummer, K. J. (2008). *Analysis of the psychometric properties of two different concept-map assessment tasks*. Doctoral dissertation. Brigham Young University.
- Rice, D. C., Ryan, J. M., & Samson, S. M. (1998). Using concept maps to assess student learning in the sciences classroom: Must different methods compete?. *Journal of Research in Science Teaching*, 35(10), 1103-1127.
- Ruiz-Primo, M.A. & Shavelson, R.J. (1996). Problems and issues in use of concept maps in science assessment. *Journal of Research in Science Teaching*. 33, 569-600.
- Schmidt, H. J. (2006). Alternative approaches to concept mapping and implications for medical education: commentary on reliability, validity and future research directions. *Advances in Health Sciences Education*, 11(1), 69-76.
- Schwendimann, B. A. (2015). Concept maps as versatile tools to integrate complex ideas: From kindergarten to higher and professional education. *Knowledge Management & E-Learning: Special Issue on Novakian Concept Mapping in University and Professional Education*, 7(1), 73-99.
- Seif, A. (2008). *Educational measurement, assessment and evaluation*. Tehran: Doran publications, twenty seventh edition. (In Persian).
- Snelson, L. J. (2010). *Estimating the Reliability of Concept Map Ratings Using a Scoring Rubric Based on Three Attributes of Propositions*. Doctoral dissertation, Brigham Young University.
- Stoddart, T., Abrams, R., Gasper, E., & Canaday, D. (2000). Concept maps as assessment in science inquiry learning: A report of methodology. *International Journal of Science Education*, 22(12), 1221-1246.
- Walker, J. M., & King, P. H. (2003). Concept mapping as a form of student assessment and instruction in the domain of bioengineering. *Journal of Engineering Education*, 92(2), 167-178.

- Watson, M. K., Pelkey, J., Noyes, C. & Rodgers, M. (2016). Assessing impacts of a learning-cycle-based module on students' conceptual sustainability knowledge using concept maps and surveys. *Journal of Cleaner Production*, 133, 544-556.
- West, D. C., Park, J. K., Pomeroy, J. R., & Sandoval, J. (2002). Concept mapping assessment in medical education: a comparison of two scoring systems. *Medical education*, 36(9), 820-826.
- Woldeamanuel, Y. W., Abate, N. T., & Berhane, D. E. (2020). Effectiveness of Concept Mapping Based Teaching Methods on Grade Eight Students' Conceptual Understanding of Photosynthesis at Ewket Fana Primary School, Bahir Dar, Ethiopia. *EURASIA Journal of Mathematics, Science and Technology Education*, 16(12), 1-16.
- Won, M., Krabbe, H., Ley, S. L., Treagust, D. F., & Fischer H. E. (2017). Science Teachers' Use of a Concept Map Marking Guide as a Formative Assessment Tool for the Concept of Energy. *Educational Assessment*, 22(2), 95-110.
- Yin, Y., & Shavelson, R. J. (2008). Application of Generalizability Theory to Concept Map Assessment Research. *Applied Measurement in Education*, 21(3), 273-291.