

Original Paper

# Knowledge Graphs Based on Meta-Analysis Papers Improve the Quality of Case Formulation: Mixed Methods Design

Kenji Yokotani<sup>1,2</sup>, PhD; Yasumitsu Jikihara<sup>3</sup>, PhD; Kohei Koiwa<sup>4</sup>, PhD

<sup>1</sup>Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Tokushima, Japan

<sup>2</sup>Department of Data Science, PsychoBit Inc., Kobe, Japan

<sup>3</sup>Graduate School of Human Sciences, The University of Osaka, Suita, Japan

<sup>4</sup>Graduate School of Education, Hokkaido University of Education, Sapporo, Japan

## Corresponding Author:

Kenji Yokotani, PhD  
Graduate School of Technology, Industrial and Social Sciences  
Tokushima University  
1-1, Minamijosanjima-cho  
Tokushima 770-8502  
Japan  
Phone: 81 88-656-7000  
Email: [yokotanikenji@tokushima-u.ac.jp](mailto:yokotanikenji@tokushima-u.ac.jp)

## Abstract

**Background:** Case formulation (CF) is a core skill for therapists; however, creating high-quality CFs requires considerable time.

**Objective:** This study aims to demonstrate that providing a knowledge graph based on meta-analytic literature can enhance CF quality.

**Methods:** Five groups were established, including 4 large language model groups and 1 human expert group, each generating 25 CFs based on 25 vignettes. The control group with Claude (Sonnet 3.7; Anthropic) produced 25 CFs. The personalization group served as the control group with additional personalization prompts. The knowledge graph group used a large language model that generated 25 CFs, which was provided with a meta-analysis knowledge graph. Further incorporation of additional personalization prompts then comprised the knowledge graph with personalization group. Finally, the expert group consisted of 25 CFs generated by a human expert. These 125 CFs in total were evaluated for general quality (ie, correctness, completeness, feasibility, and consistency) using a 7-point scale and 18 essential elements with binary scores (0 or 1) by another human expert. The CFs were also qualitatively analyzed.

**Results:** The knowledge graph and knowledge graph with personalization groups scored significantly higher than the control group in terms of correctness, completeness, and feasibility. The expert group scored significantly higher on consistency than the machine-generated groups. Additionally, there was no significant difference in the feasibility scores among the knowledge graph, knowledge graph with personalization, and expert groups. The qualitative evaluation suggested that human CFs narrow the text to content that is easy for the client to read, whereas machine CFs are more likely to include expressions that are unnatural to the client.

**Conclusions:** These results indicate that providing knowledge graphs to novice therapists increases the correctness, completeness, and feasibility of CF. Providing experienced therapists with knowledge graphs is suggested to improve the quality of their CF and mental health services.

*JMIR Form Res* 2026;10:e76808; doi: [10.2196/76808](https://doi.org/10.2196/76808)

**Keywords:** case formulation; knowledge graph; family therapy; large language model; LLM; reframing

## Introduction

Case formulation (CF), or case conceptualization, is a hypothesis collaboratively developed by the therapist and the

client to address the client's psychological difficulties [1,2]. It influences both the choice of therapeutic approach [3] and treatment outcomes [4]. CF is a core skill for therapists [5]; however, extensive training is required to develop

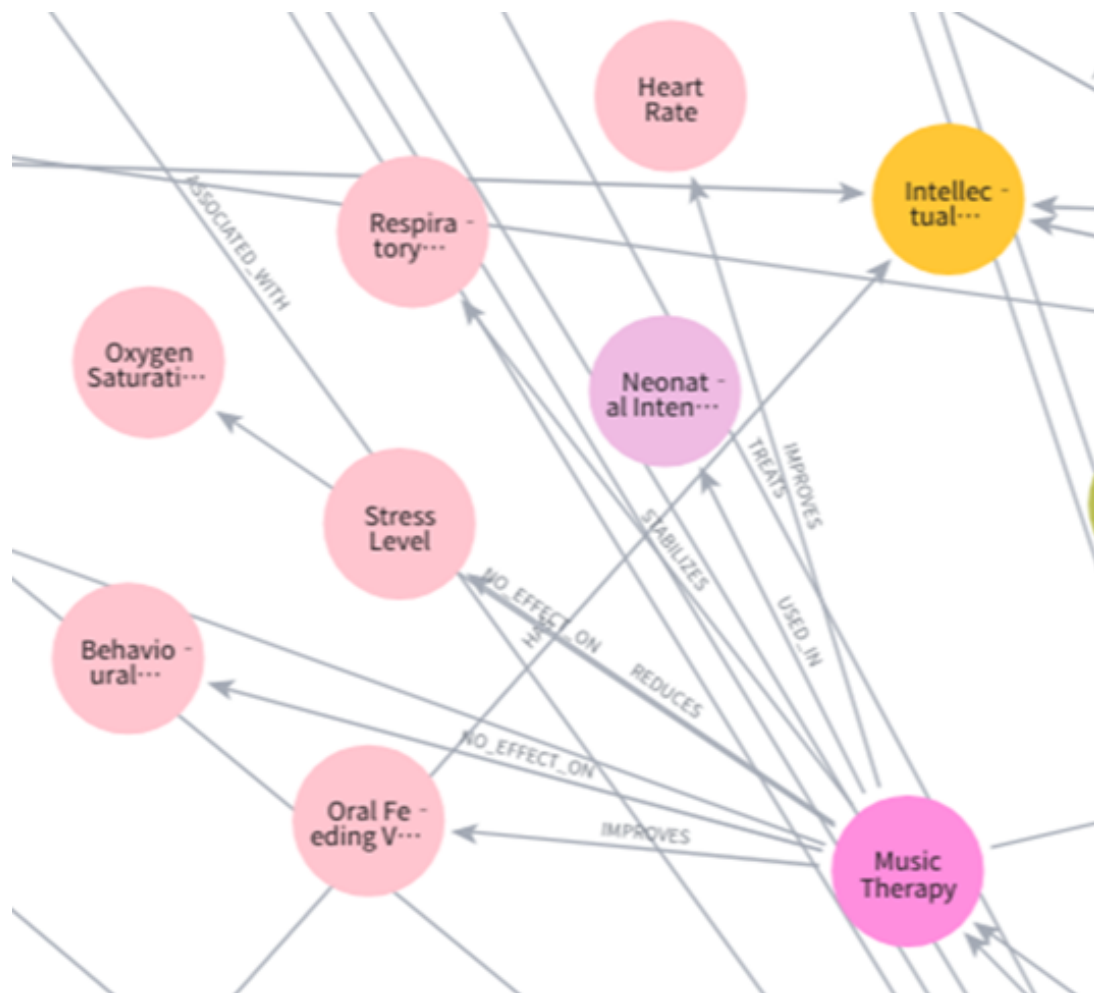
high-quality CF [6]. CF involves not only the diagnosis of mental disorders but also the comprehensive assessment of the client’s psychological difficulties and associated risks [7]. Consequently, it is a time-consuming process, with previous research indicating that the development of a single CF requires an average of approximately 52 minutes [8]. In recent years, the automated generation of CF through a large language model (LLM) has been used as an auxiliary tool for therapists to create high-quality CF. It has been suggested that by using this CF, therapists can produce high-quality CF within a relatively short training period [1]. This study demonstrates that by incorporating the collection of meta-analysis evidence as a knowledge graph into LLM [9,10], a high-quality CF can be generated. High-quality CF could serve as an auxiliary clinical tool for therapists to develop their CF skills [11].

The theoretical model of this study is an evidence-based CF model [12], which emphasizes the incorporation of

findings from previous research—that is, evidence—into CF. Evidence-based CF is currently considered the most reasonable approach [13]. Although therapists understand the importance of citing evidence, it has been pointed out that, in practice, CF is conducted without using evidence because of the time-consuming nature of literature searches [14]. However, theoretically, evidence-based assessments and CFs refine CFs and gradually improve their quality. [15].

Knowledge graphs are used to retrieve evidence collections. A knowledge graph is a network that systematically connects various pieces of knowledge [16] and represents them in a graph structure (Figure 1); knowledge graphs have been constructed from clinical trials [17], medical records [18], and research papers [16]. Decision-making for therapists can be supported by such knowledge graphs [19]. Furthermore, by incorporating a knowledge graph into the LLM, more accurate diagnoses [10] and higher-quality diagnostic report generation [9] become possible.

**Figure 1.** Examples of nodes and edges in a knowledge graph based on meta-analytic studies of parenting stress.



Based on previous findings [9,10], as well as the premise that an accurate diagnosis is an essential component of high-quality CF [20], it is posited that augmenting an LLM with a knowledge graph would result in the generation of high-quality CF. Furthermore, CF is personalized through collaboration with the client [21], and personalized CF has been suggested to be more effective in treatment than nonpersonalized CF

[4]. Based on these previous studies, this study also investigated the personalization of CF.

As indices of CF general quality, 4 indicators (ie, correctness, completeness, consistency, and feasibility) that were used in previous research [1] were used. In addition, as an index of the individual components of CF, 18 key

components derived from previous research [2] were used to examine their presence or absence. Moreover, a qualitative evaluation was conducted to examine CFs from multiple perspectives [22]. Standardized evaluation metrics for CF were not used in this study because their reliability and validity have not yet been sufficiently established [23].

The hypotheses of this study are as follows: the LLM augmented with both the knowledge graph and personalization (knowledge graph and personalization group) is expected to generate a higher-quality CF than an LLM with no augmentation (control group). For comparison, this study also created a group with an LLM augmented with a knowledge graph only (knowledge graph-only group), a group with an LLM augmented with personalization only (personalization-only group), and a group of human therapists (expert group).

#### Textbox 1. Example of a vignette.

I'm in my 40s, raising my first son born in 2018 and my second son born in 2020 who has autism and severe intellectual disability. I've been completely exhausted from caring for my second son and have been seeing a psychiatrist. Recently, when I consulted with Child Protection Services about short-stay respite care due to parental fatigue, unexpectedly my second son was taken into temporary protective custody the very next day.

Initially, I just wanted to communicate that I was at my limit with exhaustion and needed temporary childcare. However, as I explained my situation in detail, I ended up confessing that I had been so desperate that the terrifying thought, "Would I feel relief if I killed my child..." had crossed my mind. I also mentioned that I had run out of my prescribed medication. This became a serious issue, and protective custody was decided that same day.

When I explained the situation to my husband, I thought he would be angry, but he just quietly cried. Seeing him like that, I felt overwhelmingly guilty. I realized I had caused a terrible situation.... I felt so sorry for my second son as well.

About a month after my son was taken into protective custody, discussions about ending the protection finally began. The conditions include arranging visiting nurses and helpers for me, and most importantly, securing regular short-stay respite care. However, although the city office had told us that finding short-stay facilities was the parents' responsibility, it turned out this should have been handled by the consultation service office. The city office's incorrect information was one factor that led to this situation.

Currently, nearby facilities are full, and we're searching for short-stay options throughout the entire prefecture. I understand there may be criticism for writing this blog, but I hope my experience can be helpful to others in similar situations, so I plan to continue sharing my story.

### Construction of the Knowledge Graph

Semantic Scholar [26] was used to search for research studies related to the following 10 phrases: "maternal anxiety meta-analysis," "parental anxiety meta-analysis," "maternal depression meta-analysis," "parental depression meta-analysis," "autism spectrum disorder parenting meta-analysis," "attention deficit hyperactivity disorder parenting meta-analysis," "developmental disorder parenting meta-analysis," "intellectual disabilities parenting meta-analysis," "parenting child with chronic illness meta-analysis," and "parenting child with chronic medical problem meta-analysis." Up to 500 studies were retrieved per phrase, resulting in a total of 4251 studies. Subsequently, duplicate studies based on the article IDs were removed. Additionally, studies whose titles did not include both "meta" and "analysis" were excluded, resulting in 2233 selected studies. Finally, studies without available abstracts were excluded, leaving 1544 studies. The text used in the titles and abstracts of these studies (410,654 words) served as the corpus for constructing the knowledge graph.

## Methods

### Construction of 25 Vignettes

Clinical psychologist A, who is the sole author of a book on CF in family therapy [24], created 25 fictional parenting stress vignettes. Various episodes were extracted from parenting blogs on *Ameba Blog*, a popular blogging platform in Japan [25]. Each episode was summarized after removing personal information and commercial content (prior confirmation was obtained from the administrators of the *Ameba Blog*, indicating there were no issues with such usage). An example of a vignette is shown in [Textbox 1](#). Two additional fictional parenting stress vignettes were created using the same method as that used for the preliminary datasets in order to develop the evaluation criteria.

Next, each sentence within this corpus was translated into relational data for the knowledge graph using an LLM called Claude (Sonnet 3.5; "Claude-3-5-sonnet-20240620"; Anthropic) [27], which has been frequently used in clinical contexts [28]. The relational data in a knowledge graph explicitly indicates the relationships between entities. For example, the sentence "Parent Distress was positively associated with their Child Distress" translates into relational data as "[Parent Distress] -[:ASSOCIATED\_WITH]-> [Child Distress]," illustrating 2 entities and their connecting relationship. This resulted in the extraction of 7740 entities and 2348 relationships. The constructed knowledge graph is depicted in [Figure 1](#), where individual nodes represent entities and the edges connecting these nodes represent relationships. Knowledge graphs have also been used in clinical contexts [19].

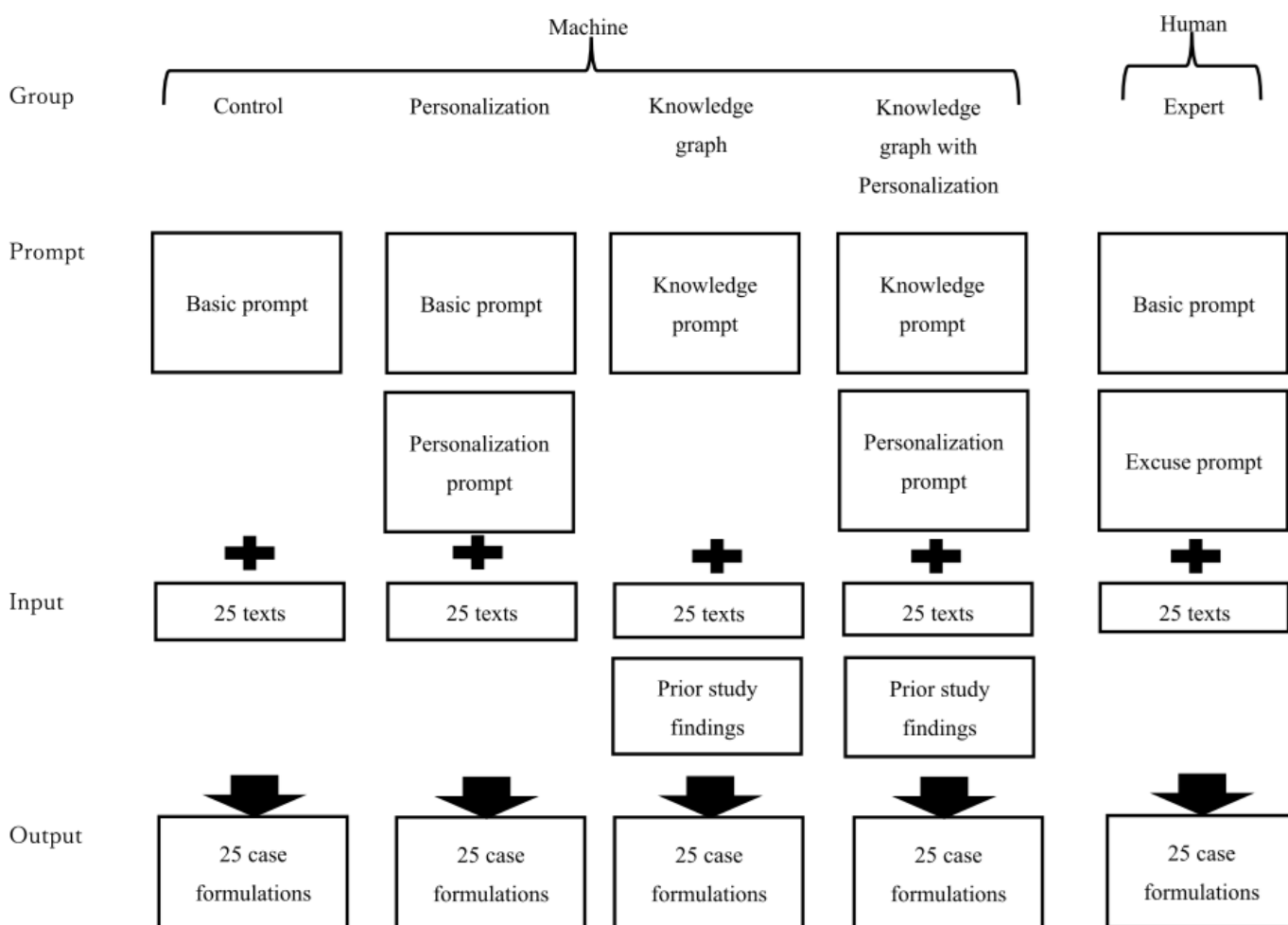
### Generation of CFs Across 5 Groups

First, for the control group, an LLM named Claude (Sonnet 3.7; "Claude-3-7-sonnet-20250219") [27], which has also been used in clinical contexts [29], was used with a basic prompt. It read the 25 vignettes ([Textbox 1](#)) and generated 25

CFs (Figure 2). Next, the personalization group incorporated a personalization prompt into the basic prompt; otherwise, the procedure was identical to that of the control group (Figure 2). For the knowledge graph group, the Japanese text was translated into English via Claude (Sonnet 3.7), and the previously constructed knowledge graph was queried using this translated English text. Based on prior studies [30,31], the similarity threshold was set at 0.9. Clinical psychologist A also judged the literature citations retrieved using this threshold to be appropriate. Relationships with similarity scores exceeding 0.9 between the English sentences and the relational data were extracted, compiled, and summarized as prior study findings (Table S1 in Multimedia

Appendix 1). These findings were used for the knowledge graph group. Additionally, the knowledge graph group used a knowledge graph prompt alongside the 25 vignettes to generate 25 CFs (Figure 2). In the knowledge graph and personalization groups, a personalization prompt was added to the knowledge graph group’s prompt. Otherwise, the procedures matched those of the knowledge graph group. Finally, clinical psychologist B, a human expert who has presented as a symposium speaker on CF in family therapy [32], read basic prompts, excuse prompts, and 25 vignettes, and then generated 25 CFs (Figure 2). Each of the 5 groups performed 25 CFs for the 25 vignettes, producing 125 CFs (5×25).

Figure 2. Comparison of prompts, input, and output across 5 groups.



### Evaluation of CFs

Clinical psychologist C, a Japanese expert who validated the assessment tools for stepfamilies in Japan [33], evaluated the previously created preliminary datasets (2 vignettes) across the 5 groups. These expert CFs were conducted only for the preliminary datasets by clinical psychologist A, rather than by clinical psychologist B, and were not used in the main evaluation of the 25 vignettes. The evaluation criteria consisted of 4 items assessing the general quality of CFs (on a 7-point scale) [1] and 18 items evaluating key components (presence or absence, coded as 0 or 1) [2]. Based on this evaluation, clinical psychologist A suggested

adjustments to ensure broader variability in general quality scores and recommended broadly defining the presence of the 18 key components. Clinical psychologist C agreed with these recommendations. Using these criteria, clinical psychologist C evaluated the 125 CFs, which were presented randomly to ensure that clinical psychologist C remained unaware of the group assignments, thus ensuring a single-blind condition.

### Statistics

A one-factor ANOVA was performed to determine the effects of the 5 groups on the quality of the CF. The Tukey multiple comparison test was used for comparisons between groups.

## Ethical Considerations

This study did not involve experimental participants, nor did it include personal data; therefore, an ethical review of personal data was not required.

the number of characters in CFs produced by human experts was significantly lower than that of the 4 LLM conditions. Although the evaluator was blinded to the source of each CF, it is possible that the evaluator inferred human authorship based on text length. Therefore, the results should be interpreted with caution.

## Results

### Comparison of Text Style Among the 5 Groups

Before testing our hypotheses, we compared the length of the CF texts across the 5 groups (Table 1). As shown in Table 1,

**Table 1.** Comparison of sentence structure, general quality, and individual key components of the 25 case formulations across the 5 groups. The significant differences in evaluation indicators were confirmed by multiple comparisons using the Tukey test.

	Control, mean (SD)	Personalization, mean (SD)	KG <sup>a</sup> , mean (SD)	KG_P <sup>b</sup> , mean (SD)	Expert, mean (SD)	Statistics F test (df)	P value
Sentence structure							
Number of characters <sup>c</sup>	885.4 (143.1)	1002.3 (142.6)	1040.2 (135.0)	1218.8 (155.1)	542.0 (182.2)	68.01 (4, 120)	<.001 <sup>d</sup>
General qualities							
Consistency <sup>e</sup>	5.36 (0.81)	5.72 (0.46)	5.64 (0.49)	5.80 (0.50)	6.20 (0.91)	5.28 (4, 120)	.001 <sup>d</sup>
Correctness <sup>f</sup>	4.56 (0.77)	5.16 (0.99)	5.96 (0.73)	6.00 (0.91)	5.12 (1.01)	11.88 (4, 120)	<.001 <sup>d</sup>
Completeness <sup>g</sup>	4.60 (0.76)	5.28 (0.74)	6.12 (0.60)	5.92 (0.91)	4.32 (0.90)	24.95 (4, 120)	<.001 <sup>d</sup>
Feasibility <sup>h</sup>	4.76 (1.01)	5.32 (0.56)	5.48 (0.51)	5.56 (0.92)	5.12 (0.88)	4.00 (4, 120)	.004 <sup>i</sup>
Individual key components							
List of problems	0.04 (0.20)	0.04 (0.20)	0.08 (0.28)	0.16 (0.37)	0.12 (0.33)	0.84 (4, 120)	.51
Sociocultural factors	0.04 (0.20)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	1.00 (4, 120)	.41
Tailored language and metaphors	0.08 (0.28)	0.36 (0.49)	0.20 (0.41)	0.08 (0.28)	0.12 (0.33)	2.60 (4, 120)	.04 <sup>j</sup>
Perpetuating factors	0.04 (0.20)	0.00 (0.00)	0.00 (0.00)	0.08 (0.28)	0.04 (0.20)	0.89 (4, 120)	.47
Protective factors	0.44 (0.51)	0.28 (0.46)	0.36 (0.49)	0.44 (0.51)	0.36 (0.49)	0.47 (4, 120)	.76
Personal meaning	0.16 (0.37)	0.08 (0.28)	0.08 (0.28)	0.08 (0.28)	0.04 (0.20)	0.59 (4, 120)	.67
Accessible language <sup>k</sup>	0.64 (0.49)	0.76 (0.44)	0.80 (0.41)	0.64 (0.49)	1.00 (0.00)	3.27 (4, 120)	.01 <sup>j</sup>
Physiological effects	0.00 (0.00)	0.00 (0.00)	0.04 (0.20)	0.08 (0.28)	0.00 (0.00)	1.37 (4, 120)	.25
Organic causes	0.08 (0.28)	0.04 (0.20)	0.08 (0.28)	0.16 (0.37)	0.04 (0.20)	0.80 (4, 120)	.53
Recent history	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	— <sup>l</sup>	—
Coping strategies <sup>m</sup>	0.44 (0.51)	0.36 (0.49)	0.52 (0.51)	0.32 (0.48)	0.04 (0.20)	4.07 (4, 120)	.004 <sup>i</sup>
Maintenance patterns	0.08 (0.28)	0.00 (0.00)	0.08 (0.28)	0.08 (0.28)	0.12 (0.33)	0.71 (4, 120)	.59
Relationship dynamics	0.28 (0.46)	0.20 (0.41)	0.32 (0.48)	0.28 (0.46)	0.32 (0.48)	0.29 (4, 120)	.89
Strengths and achievements	0.72 (0.46)	0.72 (0.46)	0.80 (0.41)	0.68 (0.48)	0.68 (0.48)	0.29 (4, 120)	.89
Childhood history	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	—	—
Cognitive schemas	0.04 (0.20)	0.00 (0.00)	0.00 (0.00)	0.04 (0.20)	0.08 (0.28)	0.89 (4, 120)	.47
Meaning making <sup>n</sup>	0.20 (0.41)	0.48 (0.51)	0.24 (0.44)	0.36 (0.49)	0.12 (0.33)	2.59 (4, 120)	.04 <sup>j</sup>
Perceived support	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.04 (0.20)	0.04 (0.20)	0.75 (4, 120)	0.56

<sup>a</sup>KG: knowledge graph.

<sup>b</sup>KG\_P: knowledge graph with personalization.

<sup>c</sup>Control < expert:  $P<.001$ ; control < KG\_P:  $P=.004$ ; expert < KG:  $P<.001$ ; expert < KG\_P:  $P<.001$ ; expert < personalization:  $P=.006$ ; and KG\_P < personalization:  $P<.001$ .

<sup>d</sup>\*\*\* $P<.001$ .

<sup>e</sup>Control < expert:  $P=.002$  and KG < expert:  $P=.03$ .

<sup>f</sup>Control < KG:  $P<.001$ ; control < KG\_P:  $P<.001$ ; expert < KG:  $P=.01$ ; expert < KG\_P:  $P=.006$ ; personalization < KG:  $P=.02$ ; and personalization < KG\_P:  $P=.01$ .

<sup>g</sup>Control < KG:  $P<.001$ ; control < KG\_P:  $P<.001$ ; control < personalization:  $P=.02$ ; expert < KG:  $P<.001$ ; expert < KG\_P:  $P<.001$ ; expert < personalization:  $P=.003$ ; personalization < KG:  $P=.002$ ; and personalization < KG\_P:  $P=.04$ .

<sup>h</sup>Control < KG:  $P=.02$  and control < KG\_P:  $P=.005$ .

<sup>i</sup>\*\* $P<.01$ .

<sup>j</sup>\* $P < .05$ .

<sup>k</sup>Control < expert:  $P = .02$  and KG\_P < expert:  $P = .02$ .

<sup>l</sup>Not applicable.

<sup>m</sup>Expert < control:  $P = .02$ <sup>n</sup> and “expert < KG:  $P = .003$ .

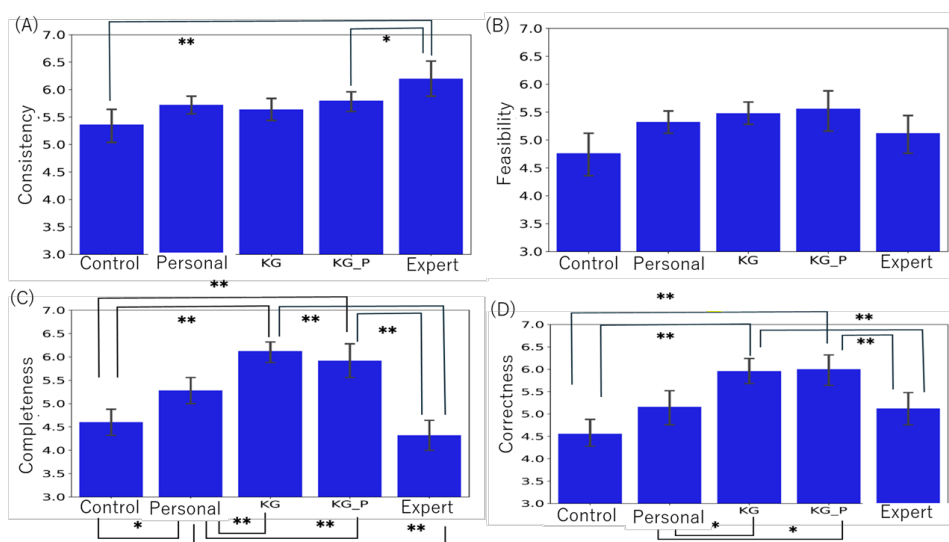
<sup>n</sup>Expert < personalization:  $P = .04$ .

## Evaluation of the General Quality of CF

The general quality of CF was evaluated (Table 1). Regarding consistency, the expert CFs showed significantly higher scores than the machine-generated CFs (Figure 3A). Regarding correctness, both the knowledge graph group and the knowledge graph with personalization group demonstrated significantly higher scores than the control group (Figure 3D). Surprisingly, the correctness scores of the knowledge graph and knowledge graph with personalization groups were significantly higher than those of the expert group (Table 1). Similar results were confirmed for completeness, indicating that both the knowledge graph group and the knowledge graph with personalization group had significantly higher

completeness scores than the control group (Figure 3C). The completeness scores of these 2 groups were also significantly higher than those of the expert group (Table 1). Regarding feasibility, the knowledge graph group and the knowledge graph with personalization group showed significantly higher scores than the control group (Figure 3B). However, there was no significant difference in feasibility between the 2 groups. These results suggest that the correctness and completeness of CF can be improved by incorporating a meta-analysis knowledge graph. Although the consistency of the machine-generated CF did not reach human expert levels, providing machines with knowledge graphs allowed the feasibility to reach levels comparable to those of experts.

**Figure 3.** Impact of the knowledge graph (KG) on the general quality of the case formulation. (A) Consistency; (B) Feasibility; (C) Completeness; (D) Correctness. KG: knowledge gap; KG\_P: Knowledge graph with personalization.

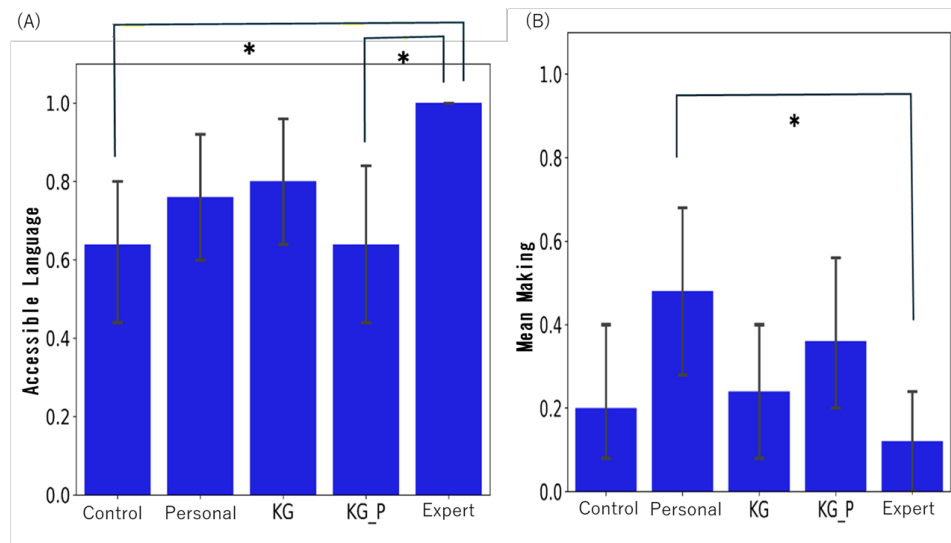


## Evaluation of Key Components in CF

The key components of CF were compared (Table 1). For the accessible language component, the expert group scored significantly higher than the control group and the knowledge graph with personalization group (Figure 4A). Regarding the mean-making component, the personalization group scored

significantly higher than the expert group (Figure 4B). These findings suggest that human experts are more likely to use accessible language than machines and that personalization prompts might increase the mean-making components within the CF.

**Figure 4.** Group comparisons for accessible language (A) and mean-making (B) components in the case formulation. KG: knowledge graph; KG\_P: knowledge graph with personalization. \* $P < .05$ .



### Qualitative Analysis of CF

A qualitative analysis of the CFs from the 5 groups based on the vignette in [Textbox 1](#) was conducted ([Table 2](#)). Concerning the use of prior study findings, citations were present in the knowledge graph and the knowledge graph with personalization groups, but absent in the control, personalization, and expert groups ([Table 2](#)). Considering that human experts typically require more time to conduct literature searches,

leading to fewer citations during CF [14], it is possible that the control, personalization, and expert groups did not perform adequate literature searches during CF creation. In contrast, the knowledge graph and knowledge graph with personalization groups were provided with a list of relevant prior study findings in advance ([Figure 2](#); [Table S1 in Multimedia Appendix 1](#)), facilitating easier use of prior study findings.

**Table 2.** Comparison of evidence use and reframing practices among the 5 groups.

	Control	Personalization	KG <sup>a</sup>	KG_P <sup>b</sup>	Expert
Use of prior study findings	None	None	“Insights from research indicates that parental stress, anxiety, and depression can also impact a child’s condition. Thus, caring for yourself is crucial for your child’s well-being as well. Parents of children with developmental disabilities are particularly prone to experiencing high stress levels, and early support has been shown to positively affect the health of both parents and children.”	<ul style="list-style-type: none"> <li>• “However, research clearly shows that such thoughts arise from extreme fatigue and stress, and parental stress closely relates to a child’s condition.”</li> <li>• “Research indicates that paternal depression also affects children’s behavior.”</li> <li>• “Research shows autism spectrum disorder (ASD) is associated with restricted and repetitive behaviors.”</li> </ul>	None

	Control	Personalization	KG <sup>a</sup>	KG_P <sup>b</sup>	Expert
Content reframing the client's self-blame	"First, reaching out to the Child Guidance Center for help was a very courageous and appropriate decision. Recognizing your limits and seeking support is a sign of being a good parent. The resulting temporary protective custody may have been unexpected, but it was likely a necessary step to ensure the safety of yourself, your younger son, and your entire family."	Raising a child with autism and severe intellectual disabilities involves unimaginable difficulties. You seem to be blaming yourself for having had the thought, 'Wouldn't it be easier if I killed my child?' However, please remember that this was a temporary thought arising from extreme exhaustion and overwhelming stress—it does not make you a bad parent. Rather, seeking help when these thoughts occurred was an incredibly brave action on your part."	"You might be experiencing guilt and regret because your consultation with the Child Guidance Center unexpectedly led to temporary protective custody. However, this situation is not a 'failure' on your part. Rather, seeking help when you felt overwhelmed demonstrates responsible parenting."	<ul style="list-style-type: none"> <li>• "Let go of the self-blame ('It's my fault this happened'). Instead, recognize that your decision to seek help was an essential first step toward receiving future support."</li> </ul>	"I think it is very meaningful that you have put your experiences into words and shared them up to this point. Particularly, the thought, 'Wouldn't it be easier if I killed my child...?' is something that must have been incredibly difficult for anyone to say out loud. Nevertheless, honestly expressing this was an essential SOS—a significant signal that you were hurting and needed help. The fact that you had the strength to send this SOS is extremely important."
Client expressions used during reframing	None	"Wouldn't it be easier if I killed my child?"	"Failure"	<ul style="list-style-type: none"> <li>• "It's my fault this happened"</li> </ul>	"Wouldn't it be easier if I killed my child...?"
Comments after response and evaluation					"When creating the CF, I was particularly mindful of the client's perspective, anticipating that they would read it. Due to the significant amount of information, I was concerned whether clients could fully absorb it, so I selectively included content that I deemed manageable for the current client and omitted content that might be less accessible. Particularly in severe cases, while a diagnostic impression existed, there were instances where I was uncertain whether directly confronting the client with such information was appropriate." (Respondent, clinical psychologist B)

<sup>a</sup>KG: Knowledge graph.

<sup>b</sup>KG\_P: Knowledge graph with personalization.

The CFs were also examined from a reframing perspective. Reframing involves therapists actively attempting to transform their clients' thoughts, which is a common technique in family therapy [34]. The vignette in [Textbox 1](#) depicts a mother expressing self-blame, making reframing particularly applicable [35]. All 5 groups applied reframing to this aspect ([Table 2](#)). While reframing recommends directly using client phrases [36], only the expert, personalization, and knowledge graph with personalization groups adhered to this recommendation ([Table 2](#)). The control group did not use client expressions during reframing and the knowledge graph group used only single words, suggesting that they did not

fully adhere to this recommendation ([Table 2](#)). This analysis indicates that personalization prompts facilitate the extraction of client-specific expressions, enhancing meaning within CF.

[Table 2](#) also includes comments from the experts following their responses and evaluations. The responding expert (clinical psychologist B) indicated a deliberate practice of writing only content considered easily acceptable to the recipient rather than exhaustively documenting the entirety of the CF. This practice is thought to contribute to the higher consistency observed in human-generated CFs compared to those generated by machines. Furthermore, an evaluation

expert (clinical psychologist C) pointed out the unnatural expressions present in machine-generated CFs when directly returned to clients. This observation aligns with the lower scores for accessible language found in machine-generated CFs compared to human-generated CFs.

## Discussion

### Main Findings

This study demonstrates that the correctness and completeness of CF increase when using a knowledge graph based on meta-analyses. In addition, the use of a knowledge graph resulted in the feasibility of CF reaching a level comparable to that of human experts. These findings support the validity of the evidence-based CF model [12]. Incorporating the findings of meta-analytic studies into CF has 3 advantages. First, integrating academic findings enhances the accuracy of CF content [12]. Second, the inclusion of diverse academic findings increases CF completeness [15]. Third, because meta-analyses often include specific therapeutic methods, their incorporation into CF enhances feasibility [13]. This study provides concrete data to support the validity of the evidence-based CF model [12]. This study is valuable in that it enhances CF by incorporating a knowledge graph derived from meta-analytic findings into an LLM. Although previous studies have applied knowledge graphs to diagnostic assessment [30,31], research using knowledge graphs specifically for CF remains scarce. This study is novel in demonstrating that the use of a knowledge graph can improve the quality of CF. Furthermore, previous CF research has predominantly focused on cognitive-behavioral therapy in Western countries [2,23]; incorporating Asian family therapy has expanded the cultural diversity of CF research [1].

### Strengths of Human Experts

Regarding consistency, the influence of the knowledge graph was minimal, and human expert CF generally scored higher than machine-generated CF. Human experts summarized the content in paragraphs rather than bullet points, potentially contributing to their higher consistency scores. In addition, human experts included more elements of accessible language than machines. Although the machines were prompted to use simple language and adhere to these instructions, the human experts still exhibited more explicit emotional expressions. The experts effectively combined the expressions presented in the vignettes to convey nuanced meanings. Nuanced expression is a difficult area for LLMs. Although LLMs may have an advantage over the general population [37], human experts currently show an advantage, and a similar trend arises in a variety of specialties [38].

### Effects of Personalization

Personalization had a limited impact on the general quality or specific key components of CF, which is inconsistent with previous findings [4,21]. However, qualitative analysis indicated that personalization facilitates the extraction of client-specific expressions, promoting meaning-making within CF. During vignette creation, writer-specific expressions were

likely diminished because of the removal of proper nouns, potentially obscuring the personalization effect. Individual-specific expressions may also have been lost during the summarization process. One previous study has demonstrated that a 2-stage approach—first summarizing individual-specific characteristics and subsequently performing diagnostic reasoning—yields higher diagnostic accuracy in LLMs than direct diagnosis alone [28]. Accordingly, adopting a 2-stage approach in which individual-specific expressions are first extracted and then used to generate CFs may facilitate the development of CFs that better capture individual-specific characteristics. Future research should use longer vignettes with more writer-specific expressions and a 2-stage approach to better examine personalization effects [4].

### Practical Implications

This study has 2 practical implications. First, knowledge of graph-generated CF can be used to train novice therapists [11]. Novice therapists can compare their CFs with machine-generated CFs based on knowledge graphs, which helps them identify missing perspectives or components in their formulations [39]. Second, knowledge of graph-generated CFs can support expert therapists. Experts often lack time for literature searches, resulting in limited citations of previous research [14]. Machines can automatically extract meta-analytic literature data and present experts with current evidence, thereby improving CF quality through the timely integration of such evidence [15]. Since the knowledge extracted from meta-analyses may remain at a relatively superficial level, and even when the cited knowledge is accurate, its application in practice requires careful consideration to avoid placing undue burden on clients [30,31]. Therefore, when such knowledge is used in clinical settings, professional experience and ethical judgment are essential [28]. Indeed, the clinician involved in this study reported intentionally reducing the length of the CF to minimize the client's burden. Because CF generation by machines is less time-consuming and more cost-effective [8], appropriately reviewing machine-generated CFs by human experts could contribute to improving the quality of CFs produced by humans.

### Limitations

This study has 4 limitations. First, there was a potential bias in the constructed data, as only 1 individual was involved in vignette creation, 1 in response, and 1 in evaluation. Because this study used only a single evaluator, interrater reliability could not be calculated. The scores used as the evaluation metric could not be examined in relation to other external measures, and thus the validity of the metric was not established. In addition, because test-retest assessments were not conducted for this measure, test-retest reliability could not be calculated. To ensure the reliability and validity of these measures, future research should incorporate test-retest evaluations and CF measures with established validity [8]. Securing multiple vignette creators, respondents, and evaluators can prevent such bias [2]. Second, human experts' CFs differed significantly from machine-generated CFs in style and length, potentially enabling evaluators to infer human authorship

despite the blinded condition. Evaluators may have identified and preferentially rated the machine responses negatively [40]. Because the CFs produced by human experts were shorter, the evaluator may have recognized them as being authored by human experts. This recognition may have contributed to the higher consistency ratings observed for the human expert condition. Future research should standardize the length and style of responses to ensure more robust blinding. Third, vignette-style presentation may not be the optimal way to present clinical cases. Presenting clinical cases in a vignette style reportedly resulted in lower diagnostic accuracy among human experts compared with presentations that list symptoms alone [41]. Similarly, when LLMs evaluate cases presented in a vignette style, they tend to generate an excessive number of diagnoses [42]. Therefore, future studies should examine the quality of CF using a format in which symptoms are presented solely as structured lists. Fourth, qualitative analysis in this study did not use thematic analysis [43]. In future research, more

comprehensive qualitative analyses could be achieved by conducting thematic analysis and ensuring transparency in the coding process.

## Conclusions

Despite these limitations, our study demonstrated the effectiveness of knowledge graphs in enhancing the general quality of CF, which is consistent with studies showing the usefulness of knowledge graphs in the clinical domain [9, 10,16-19]. A comparison of human and machine-generated CFs also highlighted human strengths in terms of clarity and machine strengths in terms of completeness and correctness. Future studies should demonstrate improvements in the quality of therapist-generated CFs by presenting machine-generated CFs with meta-analysis knowledge graphs to novice and expert therapists [5]. Using machines as supplementary tools to enhance CF quality will enable therapists to rapidly produce high-quality CFs [3], ultimately improving the mental health services provided to many individuals [20].

---

## Acknowledgments

This study was partially presented in Japanese as a poster at the 89th Annual Convention of the Japanese Psychological Association in 2025. In this study, ChatGPT (GPT-5) was used to assist with computer-program code generation and the translation of Japanese manuscripts into English. All artificial intelligence (AI)-assisted outputs were subsequently reviewed and substantially revised by the authors. The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2026) [44], the following tasks were delegated to generative AI (GAI) tools under full human supervision: code generation, code optimization, data collection, and translation. The GAI tools used were: ChatGPT (GPT-4.1, GPT-5, GPT-5.1, GPT-5.2, GPT-5.3, and GPT-5.4) and Claude (Sonnet 3.5 and Sonnet 3.7). The responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

---

## Funding

This study was financially supported by the Japan Society for the Promotion of Science (grant 23H00080).

---

## Data Availability

Data supporting the findings of this study are available from the corresponding author (KY) upon reasonable request.

---

## Authors' Contributions

Conceptualization: KY  
Data curation: KY  
Formal analysis: KY  
Funding acquisition: KY  
Investigation: KY, YJ, KK  
Methodology: KY  
Project administration: KY  
Resources: YJ, KK  
Software: KY  
Validation: KY  
Visualization: KY  
Writing – original draft: KY  
Writing – review & editing: KY

---

## Conflicts of Interest

None declared.

---

## Multimedia Appendix 1

Knowledge graphs.

[\[DOCX File \(Microsoft Word File\), 31 KB-Multimedia Appendix 1\]](#)

## References

1. Hsieh LH, Liao WC, Liu EY. Feasibility assessment of using ChatGPT for training case conceptualization skills in psychological counseling. *Comput Hum Behav Artif Hum*. Aug 2024;2(2):100083. [doi: [10.1016/j.chbah.2024.100083](https://doi.org/10.1016/j.chbah.2024.100083)]
2. Throrer NE, Bucci S, Morris L, Berry K. The key components of a clinical psychology formulation: a consensus study. *Br J Clin Psychol*. Jun 2024;63(2):213-226. [doi: [10.1111/bjc.12455](https://doi.org/10.1111/bjc.12455)] [Medline: [38235902](https://pubmed.ncbi.nlm.nih.gov/38235902/)]
3. Easden MH, Kazantzis N. Case conceptualization research in cognitive behavior therapy: a state of the science review. *J Clin Psychol*. Mar 2018;74(3):356-384. [doi: [10.1002/jclp.22516](https://doi.org/10.1002/jclp.22516)] [Medline: [28776663](https://pubmed.ncbi.nlm.nih.gov/28776663/)]
4. Lundkvist-Houndoumadi I, Thastum M, Hougaard E. Effectiveness of an individualized case formulation-based CBT for non-responding youths with anxiety disorders. *J Child Fam Stud*. Feb 2016;25(2):503-517. [doi: [10.1007/s10826-015-0225-4](https://doi.org/10.1007/s10826-015-0225-4)]
5. Kendjelic EM, Eells TD. Generic psychotherapy case formulation training improves formulation quality. *Psychotherapy (Chic)*. 2007;44(1):66-77. [doi: [10.1037/0033-3204.44.1.66](https://doi.org/10.1037/0033-3204.44.1.66)] [Medline: [22122169](https://pubmed.ncbi.nlm.nih.gov/22122169/)]
6. Dudley R, Ingham B, Sowerby K, Freeston M. The utility of case formulation in treatment decision making; the effect of experience and expertise. *J Behav Ther Exp Psychiatry*. Sep 2015;48:66-74. [doi: [10.1016/j.jbtep.2015.01.009](https://doi.org/10.1016/j.jbtep.2015.01.009)] [Medline: [25748793](https://pubmed.ncbi.nlm.nih.gov/25748793/)]
7. Mohtashemi R, Stevens J, Jackson PG, Weatherhead S. Psychiatrists' understanding and use of psychological formulation: a qualitative exploration. *BJPsych Bull*. Aug 2016;40(4):212-216. [doi: [10.1192/pb.bp.115.051342](https://doi.org/10.1192/pb.bp.115.051342)] [Medline: [27512592](https://pubmed.ncbi.nlm.nih.gov/27512592/)]
8. Abbas MJ, Fosker H, Dudson H, Ramewal S. A comparison of the quality of integrated case formulations produced by UK psychiatric trainees and an artificial intelligence-assisted application. *BJPsych Bull*. Oct 13, 2025:1-6. [doi: [10.1192/bjb.2025.10160](https://doi.org/10.1192/bjb.2025.10160)] [Medline: [41077965](https://pubmed.ncbi.nlm.nih.gov/41077965/)]
9. Gao Y, Li R, Croxford E, et al. Leveraging medical knowledge graphs into large language models for diagnosis prediction: design and application study. *JMIR AI*. Feb 24, 2025;4:e58670. [doi: [10.2196/58670](https://doi.org/10.2196/58670)] [Medline: [39993309](https://pubmed.ncbi.nlm.nih.gov/39993309/)]
10. Jia M, Duan J, Song Y, Wang J. MedIKAL: integrating knowledge graphs as assistants of llms for enhanced clinical diagnosis on EMRS. *Proceedings of the 31st International Conference on Computational Linguistics*. 2025:9278-9298. URL: <https://aclanthology.org/2025.coling-main.624.pdf> [Accessed 2026-03-17]
11. Abbas M, Walton R, Johnston A, Chikoore M. Evaluation of teaching an integrated case formulation approach on the quality of case formulations: randomised controlled trial. *Psychiatrist*. Apr 2012;36(4):140-145. [doi: [10.1192/pb.bp.110.033746](https://doi.org/10.1192/pb.bp.110.033746)]
12. Eells TD, Lombart KG. Theoretical and evidence-based approaches to case formulation. In: *Forensic Case Formulation*. John Wiley & Sons, Ltd; 2011:1-32. ISBN: 9781119977018
13. Pachankis JE, Soulliard ZA, Morris F, Seager van Dyk I. A model for adapting evidence-based interventions to be LGBTQ-affirmative: putting minority stress principles and case conceptualization into clinical research and practice. *Cogn Behav Pract*. Feb 2023;30(1):1-17. [doi: [10.1016/j.cbpra.2021.11.005](https://doi.org/10.1016/j.cbpra.2021.11.005)]
14. Huisman P, Kangas M. Evidence-based practices in cognitive behaviour therapy (CBT) case formulation: what do practitioners believe is important, and what do they do? *Behav change*. Apr 2018;35(1):1-21. [doi: [10.1017/bec.2018.5](https://doi.org/10.1017/bec.2018.5)]
15. Christon LM, McLeod BD, Jensen-Doss A. Evidence-based assessment meets evidence-based treatment: an approach to science-informed case conceptualization. *Cogn Behav Pract*. Feb 2015;22(1):36-48. [doi: [10.1016/j.cbpra.2013.12.004](https://doi.org/10.1016/j.cbpra.2013.12.004)]
16. Alam F, Giglou HB, Malik KM. Automated clinical knowledge graph generation framework for evidence based medicine. *Expert Syst Appl*. Dec 2023;233:120964. [doi: [10.1016/j.eswa.2023.120964](https://doi.org/10.1016/j.eswa.2023.120964)]
17. Chen Z, Peng B, Ioannidis VN, Li M, Karypis G, Ning X. A knowledge graph of clinical trials ([formula: see text]). *Sci Rep*. Mar 18, 2022;12(1):4724. [doi: [10.1038/s41598-022-08454-z](https://doi.org/10.1038/s41598-022-08454-z)] [Medline: [35304504](https://pubmed.ncbi.nlm.nih.gov/35304504/)]
18. Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a health knowledge graph from electronic medical records. *Sci Rep*. Jul 20, 2017;7(1):5994. [doi: [10.1038/s41598-017-05778-z](https://doi.org/10.1038/s41598-017-05778-z)] [Medline: [28729710](https://pubmed.ncbi.nlm.nih.gov/28729710/)]
19. Lyu K, Tian Y, Shang Y, et al. Causal knowledge graph construction and evaluation for clinical decision support of diabetic nephropathy. *J Biomed Inform*. Mar 2023;139:104298. [doi: [10.1016/j.jbi.2023.104298](https://doi.org/10.1016/j.jbi.2023.104298)] [Medline: [36731730](https://pubmed.ncbi.nlm.nih.gov/36731730/)]
20. Macneil CA, Hasty MK, Conus P, Berk M. Is diagnosis enough to guide interventions in mental health? Using case formulation in clinical practice. *BMC Med*. Sep 27, 2012;10:111. [doi: [10.1186/1741-7015-10-111](https://doi.org/10.1186/1741-7015-10-111)] [Medline: [23016556](https://pubmed.ncbi.nlm.nih.gov/23016556/)]
21. van den Bergh R, Olthof M, Goldbeck F, et al. The content of personalised network-based case formulations. *J Contemp Psychother*. Sep 2024;54(3):181-192. [doi: [10.1007/s10879-023-09613-7](https://doi.org/10.1007/s10879-023-09613-7)]
22. Spencer HM, Dudley R, Johnston L, Freeston MH, Turkington D, Tully S. Case formulation-a vehicle for change? Exploring the impact of cognitive behavioural therapy formulation in first episode psychosis: a reflexive thematic analysis. *Psychol Psychother*. Jun 2023;96(2):328-346. [doi: [10.1111/papt.12442](https://doi.org/10.1111/papt.12442)] [Medline: [36480353](https://pubmed.ncbi.nlm.nih.gov/36480353/)]

23. Flinn L, Braham L, das Nair R. How reliable are case formulations? A systematic literature review. *Br J Clin Psychol*. Sep 2015;54(3):266-290. [doi: [10.1111/bjc.12073](https://doi.org/10.1111/bjc.12073)] [Medline: [25515566](https://pubmed.ncbi.nlm.nih.gov/25515566/)]
24. Yokotani K. *Illustrated Case-Based Family Therapy: Systems and Narrativist Case Formulation and Intervention* [Book in Japanese]. Tomishobo; 2022. ISBN: 978-4-86616-155-6
25. Takano M, Taka F, Ogiue C, Nagata N. Online harassment of Japanese celebrities and influencers. *Front Psychol*. 2024;15:1386146. [doi: [10.3389/fpsyg.2024.1386146](https://doi.org/10.3389/fpsyg.2024.1386146)] [Medline: [38686089](https://pubmed.ncbi.nlm.nih.gov/38686089/)]
26. Fricke S. Semantic Scholar. *J Med Libr Assoc*. Jan 2018;106(1):145-147. [doi: [10.5195/jmla.2018.280](https://doi.org/10.5195/jmla.2018.280)]
27. Models overview. Claude APIDocs. 2021. URL: <https://docs.anthropic.com/en/docs/about-claude/models/all-models> [Accessed 2025-04-17]
28. Kurokawa R, Ohizumi Y, Kanzawa J, et al. Diagnostic performances of Claude 3 Opus and Claude 3.5 Sonnet from patient history and key images in Radiology's "Diagnosis Please" cases. *Jpn J Radiol*. Dec 2024;42(12):1399-1402. [doi: [10.1007/s11604-024-01634-z](https://doi.org/10.1007/s11604-024-01634-z)] [Medline: [39096483](https://pubmed.ncbi.nlm.nih.gov/39096483/)]
29. Feng J, Zheng Q, Wu C, et al. M<sup>3</sup>Builder: a multi-agent system for automated machine learning in medical imaging. Presented at: AI for Clinical Applications Agentic AI CMLLMs CREATE 2025 2025 2025; Sep 23-27, 2025; Daejeon, South Korea. URL: [https://link.springer.com/chapter/10.1007/978-3-032-06004-4\\_12#citeas](https://link.springer.com/chapter/10.1007/978-3-032-06004-4_12#citeas) [Accessed 2026-03-17] [doi: [10.1007/978-3-032-06004-4\\_12](https://doi.org/10.1007/978-3-032-06004-4_12)]
30. Gao S, Yu K, Yang Y, et al. Large language model powered knowledge graph construction for mental health exploration. *Nat Commun*. Aug 13, 2025;16(1):7526. [doi: [10.1038/s41467-025-62781-z](https://doi.org/10.1038/s41467-025-62781-z)]
31. Lui SJ, Xiang C, Krishnaswamy S. KAMEL: knowledge aware medical entity linkage to automate health insurance claims processing. Presented at: AAAI'24/IAAI'24/EAAI'24: Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence; Feb 20-27, 2024; British Columbia, Canada. [doi: [10.1609/aaai.v38i21.30314](https://doi.org/10.1609/aaai.v38i21.30314)]
32. Uchiyama A, Wakashima K, Koiwa K, Nagano K, Kakei K. Learning brief therapy and family therapy—case formulation in brief therapy and family therapy [article in Japanese]. Presented at: 43rd Annual Conference of the Japanese Association of Clinical Psychology; Oct 6, 2024.
33. Jikihara Y, Nozawa S, Ando S, Higginbotham BJ, Schramm DG, Adler-Baeder F. Exploring the factor structure and concurrent validity of two stepfamily measures for Japanese couples. *J Divorce Remarriage*. May 19, 2023;64(2-4):156-172. [doi: [10.1080/10502556.2023.2242751](https://doi.org/10.1080/10502556.2023.2242751)]
34. Moran G, Diamond GM, Diamond GS. The relational reframe and parents' problem constructions in attachment-based family therapy. *Psychother Res*. Jul 2005;15(3):226-235. [doi: [10.1080/10503300512331387780](https://doi.org/10.1080/10503300512331387780)] [Medline: [22011152](https://pubmed.ncbi.nlm.nih.gov/22011152/)]
35. Munroe M, Al-Refae M, Chan HW, Ferrari M. Using self-compassion to grow in the face of trauma: the role of positive reframing and problem-focused coping strategies. *Psychol Trauma*. Apr 2022;14(S1):S157-S164. [doi: [10.1037/tra0001164](https://doi.org/10.1037/tra0001164)] [Medline: [34735189](https://pubmed.ncbi.nlm.nih.gov/34735189/)]
36. Lyddon WJ, Clay AL, Sparks CL. Metaphor and change in counseling. *Jour of Counseling & Develop*. Jul 2001;79(3):269-274. [doi: [10.1002/j.1556-6676.2001.tb01971.x](https://doi.org/10.1002/j.1556-6676.2001.tb01971.x)]
37. Sorin V, Brin D, Barash Y, et al. Large Language Models and empathy: systematic review. *J Med Internet Res*. Dec 11, 2024;26:e52597. [doi: [10.2196/52597](https://doi.org/10.2196/52597)] [Medline: [39661968](https://pubmed.ncbi.nlm.nih.gov/39661968/)]
38. Yokotani K, Takagi G, Wakashima K. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. *Comput Human Behav*. Aug 2018;85:135-145. [doi: [10.1016/j.chb.2018.03.045](https://doi.org/10.1016/j.chb.2018.03.045)]
39. Liese BS, Esterline KM. Concept mapping: a supervision strategy for introducing case conceptualization skills to novice therapists. *Psychotherapy (Chic)*. Jun 2015;52(2):190-194. [doi: [10.1037/a0038618](https://doi.org/10.1037/a0038618)] [Medline: [25730627](https://pubmed.ncbi.nlm.nih.gov/25730627/)]
40. Thuillard S, Adams M, Jelmini G, Schmutz S, Sonderegger A, Sauer J. When humans and computers induce social stress through negative feedback: effects on performance and subjective state. *Comput Human Behav*. Aug 2022;133:107270. [doi: [10.1016/j.chb.2022.107270](https://doi.org/10.1016/j.chb.2022.107270)]
41. Payton EM, Graber ML, Bachiasvili V, Mehta T, Dissanayake PI, Berner ES. Impact of clinical note format on diagnostic accuracy and efficiency. *Health Inf Manag J*. Sep 2024;53(3):183-188. [doi: [10.1177/18333583231151979](https://doi.org/10.1177/18333583231151979)]
42. Sarma KV, Hanss KE, Halls AJM, et al. Integrating expert knowledge into large language models improves performance for psychiatric reasoning and diagnosis. *Psychiatry Res*. Jan 2026;355:116844. [doi: [10.1016/j.psychres.2025.116844](https://doi.org/10.1016/j.psychres.2025.116844)] [Medline: [41270691](https://pubmed.ncbi.nlm.nih.gov/41270691/)]
43. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. Jan 2006;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]

44. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Account Res. Apr 2026;33(3):2544331*. [doi: [10.1080/08989621.2025.2544331](https://doi.org/10.1080/08989621.2025.2544331)] [Medline: [40781729](https://pubmed.ncbi.nlm.nih.gov/40781729/)]

---

## Abbreviations

**CF:** case formulation

**LLM:** large language model

---

*Edited by Amaryllis Mavragani, Ivan Steenstra; peer-reviewed by Balu Bhasuran; submitted 01.May.2025; final revised version received 05.Mar.2026; accepted 06.Mar.2026; published 30.Jun.2026*

*Please cite as:*

*Yokotani K, Jikihara Y, Koiwa K*

*Knowledge Graphs Based on Meta-Analysis Papers Improve the Quality of Case Formulation: Mixed Methods Design*

*JMIR Form Res 2026;10:e76808*

*URL: <https://formative.jmir.org/2026/1/e76808>*

*doi: [10.2196/76808](https://doi.org/10.2196/76808)*

© Kenji Yokotani, Yasumitsu Jikihara, Kohei Koiwa. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 30.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.