

A dialogue system for identifying need deficiencies in moral education

Journal of Pacific Rim Psychology

Volume 15: 1–14

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1834490921998589

journals.sagepub.com/home/pac



Penghe Chen , Yu Lu, Shengquan Yu, Qi Xu and Jiefei Liu

Abstract

Moral education refers to the cultivation of ideals, moral quality, culture, and discipline. One of its main tasks is to analyze students' problem behaviors and identify their underlying need deficiencies. Previous psychological research has focused on studying how distinct factors affect psychological needs and problem behaviors. However, these findings have provided only scattered guidelines for identifying students' need deficiencies, which are difficult for inexperienced teachers and parents to apply systematically. To address these issues, we attempt to answer two key research questions in this work. First, how do we define a theoretical framework so that the psychological research findings can be systematically applied to identify students' need deficiencies? Second, can the latest AI technologies be employed to identify such need deficiencies automatically? To answer these research questions, we first build a theoretical framework to summarize all the factors relevant to the students' problem behaviors and need deficiencies. After that, we propose and develop a task-oriented dialogue system that can properly inquire about different aspects of students' information and automatically infer their need deficiencies. We conduct comprehensive experiments to evaluate the system's performance with real-life cases. The results show that the built dialogue system could effectively serve as a diagnostic tool to identify the students' need deficiencies.

Keywords

need deficiency, problem behavior, moral education, dialogue system

Received 29 January 2020; accepted 30 January 2021

Introduction

Moral education is defined as the cultivation of ideals, moral quality, culture, and disciplines, and the core objective is to promote personal well-being and character development (Lee & Ho, 2005). Essentially, moral education aims to reduce violence, incivility, and misconduct by modifying students' problem behaviors while assisting in healthy moral development (Koh, 2012; Schuitema et al., 2008). Past literature has demonstrated that effective moral education can promote behavioral advancement and grade improvement among students (Jeynes, 2019). One main task of moral education is to manage students' problem behaviors, such as truancy and fighting in school, which are undesirable since they are not consistent with social norms (Jessor, 2010). Psychology researchers have argued that, according to Maslow's hierarchy of needs, need deficiency drives students' behaviors. Therefore, problem behaviors could be reduced by mitigating the deficiencies (Harper et al., 2003).

Traditionally, teachers and parents have mainly relied on their own experiences to manage students' problem behaviors, but inexperienced persons usually lack such capabilities and expertise. In addition, it is difficult to master and apply the obscure psychological theories, which further increases the difficulties of teachers and parents in identifying students' need deficiencies.

Artificial intelligence (AI) aims to understand the essence of intelligence and enable machines to perform like human beings. The current AI research covers various key fields, including natural language processing, computer vision, reasoning, and decision-making, and

Advanced Innovation Center for Future Education, School of Educational Technology, Beijing Normal University

Corresponding author:

Yu Lu, No.19, Xijiekouwai St, Haidian District, Beijing, 100875, P.R. China.

Email: luyu@bnu.edu.cn



it has achieved great progress and even outperforms humans in specific tasks (e.g., image classification and GO game). AI technologies have been applied in different domains, such as finance, healthcare, transportation, and education. One typical application is building dialogue systems, which usually converse with people and help them complete a specific task like restaurant reservation and flight booking. During the interaction process, the dialogue system could keep querying the users to acquire the necessary information for the task, and the user could reply using natural language either in text or voice. For example, a dialogue system might query the information from the user on the favorite food type, location, and price range to accomplish a restaurant reservation task. Such advancements in the AI domain inspire us to employ the corresponding technologies for moral education, especially help to manage students' problem behaviors and identify their need deficiencies.

Specifically, we target two key research questions in this work: 1) how do we define a theoretical framework so that the psychological research findings can be systematically applied to analyze students' need deficiencies, and 2) how to employ the AI technologies to identify students' need deficiencies automatically. To answer these two questions, we first build a theoretical framework of problem behaviors and need deficiencies by investigating the existing psychological research findings. After that, guided by the built theoretical framework, we propose and design a task-oriented dialogue system to automatically identify students' need deficiencies underlying problem behaviors. To evaluate the effectiveness of the proposed dialogue system, comprehensive experiments are conducted with real-life cases. The experimental results demonstrate that the proposed dialogue system could correctly identify 44% need deficiencies of the participants, and the number of dialogue turns (i.e., the rounds of ask-response in one dialogue) is 12 on average.

Problem behavior and need deficiency

Problem behaviors refer to undesirable behaviors that are not consistent with social norms and usually raise concerns or control responses from others (Jessor, 2010). Due to immature physical and psychological development, elementary and middle school students are easily affected by environmental factors such as family, school, and society during their growth. With the negative influence from certain external factors, students may present problem behaviors that deviate from family and social standards (Maggs & Galambos, 1993). Researchers have worked to categorize problem behaviors into different types for better management (Achenbach & Ruffle, 2000). Peterson

proposed a personality versus conduct dichotomy, in which students who tend to show impulses toward society are labeled as having conduct problems (i.e., aggression), whereas students whose issues are more covert are labeled as having personality problems (i.e., social withdrawal) (Peterson & Donald, 1961). On the other hand, Miller argued that the comparison should be between aggression and inhibition, which focused on emotional problems (Miller, 1968).

Different frameworks have been proposed to identify students' problem behaviors. The Achenbach Child Behavior Checklist (CBCL) was the first tool to use empirical, multi-axis, and cross-assessor measurement methods to identify students' problem behaviors. Specifically, three types of forms were designed: the Teacher Report Form, the Youth Self-Reports, and the Direct Forms. These forms are considered to have complete reliability and validity based on a series of cross-cultural studies (Achenbach, 1991). To assist teachers in resolving the classroom behavior management problems, Scarpaci (2007) proposed a five-step approach: identifying the problem, the objectives to be achieved, the solution, the implementation, and the evaluation (IOSIE). These classifications and frameworks provide guidelines on building the theoretical framework.

Research work has also been conducted to analyze the reasons underlying students' problem behaviors. Maslow argued that human behaviors are driven by the lack of satisfaction with psychological needs (Maslow, 2013). Harper claimed that natural disasters, violence, abuse, poverty, lack of school and community resources, and emotional deprivation contribute to students' unmet basic needs (Harper et al., 2003). Dennis et al. indicated that individual development was affected by the interactions between individual characteristics and environmental factors (Dennis et al., 2005). Jessor (2010) pointed out that the personality system and the perceived environmental system influenced students' behaviors. Therefore, we think that students' problem behaviors are driven by the lack of satisfaction with their basic needs, which are influenced by external environmental and internal individual factors.

Psychology researchers have also explored how various distinct factors affect students' problem behaviors and need deficiencies. Previous studies have shown that the Big Five traits were closely related to students' problem behaviors (Van et al., 2013). For instance, Ehrler et al. (1999) found that the students with low scores in agreeableness and conscientiousness might exhibit conduct problems and attention deficits, and the neuroticism trait was associated with anxiety and depression. In addition, family conditions also affect student behavioral problems. Hoffmann's (2006) study showed that changes in parents' marital status

could increase students' probability of engaging in problem behaviors. Fomby and Christie (2013) found that adolescents from unstable families (e.g., blended families) were more likely to engage in aggressive and antisocial behaviors than adolescents in stable families. Pinquart and Martin (2017) observed that authoritarian, permissive, and neglectful parenting styles caused externalizing problems with high probabilities.

Moreover, Purwati and Japar (2017) found that parents' education levels affected how parents educated their children, which subsequently affected students' problem behaviors. Rosario et al. (2017) showed that students with less-educated parents presented more aggressive behaviors. Furthermore, the family's socio-economic status can also influence students' behaviors. McGrath and Elgar (2015) found that students from higher-status backgrounds presented fewer internalizing and externalizing problems. Sieh, Meijer et al. (2010) and Sieh, Visser-Meily et al. (2012) suggested that parents' health conditions influenced students' family education, which subsequently influenced students' problem behaviors.

In addition to family factors, school factors also heavily affect students' behaviors. For instance, Maryam et al. (2019) showed that students who were rejected by peer groups tended to present more internalizing problems. Farrell et al. (2017) found that students might be influenced by friends' delinquent behaviors, including aggression, drug use, and misconduct. In addition, mass media also plays an important role in students' school and social lives. Wahab et al. (2017) showed that extreme addictions to social media might result in depression. In short, all such distinct factors should be considered building the construction of a theoretical framework to identify the student need deficiencies underlying problem behaviors.

Besides, research has also been conducted on behavioral intervention and management. From the family perspective, Leijten et al. (2018) pointed out that strengthening students' relationships with their parents can effectively reduce their problem behaviors. Kazdin et al. (2018) proved that proper training and managing parents' behaviors effectively reduced their children's aggression. From the school perspective, Spiller et al. (2019) pointed out that a positive school atmosphere could help students form good behaviors. Specifically, clear expectations, communication, and consistent feedback are all effective for student behavioral management in the classroom environment. The existing work does not provide a systematic theoretical framework for need deficiency identification, indicating the necessity of building a theoretical framework. The research findings summarized by the previous work provide a basis for building the theoretical framework.

Task-oriented dialogue system

A typical task-oriented dialogue system usually consists of four main modules (Gao et al., 2019). The first is the natural language understanding (NLU) module that interprets users' utterances to extract users' intents and task-related semantic slot information. The second is the dialogue state tracking (DST) module used for tracking the dialogue state. The third is the policy learning (PL) module that takes charge of making decisions on the next system action based on the current dialogue state. The fourth is the natural language generation (NLG) module that transforms specific system actions into a natural language response. Through multiturn dialogue, a task-oriented dialogue system can acquire the necessary information and complete the task automatically.

With the advancement of AI technology, various dialogue systems have been designed for different tasks. Some systems are designed for booking tasks. For example, Li et al. (2017) developed a dialogue system for movie-ticket booking. Wen et al. (2016) built a dialogue system to help users search for restaurants and make reservations. Yan et al. (2017) proposed a dialogue system to assist customers in different purchase-related tasks in online shopping. Dialogue systems have also been developed to solve information-searching tasks. For instance, Papangelis et al. (2018) designed a spoken dialogue system to help users make decisions through information navigation. Dhingra et al. (2016) proposed a dialogue agent for accessing information from knowledge bases.

In addition, the dialogue system has also been implemented for the automatic diagnosis of medical diseases. Tang et al. (2016) designed a group of anatomical models that emulated different hospital experts to diagnose diseases. In subsequent work, Kao et al. (2018) designed a hierarchical model to implement a master model for controlling the selection between different anatomical models. Peng et al. (2018) employed the reward-shaping and feature-rebuilding techniques to improve the model for quicker disease diagnosis. Wei et al. (2018) built another model to automatically diagnose patients' diseases based on symptom information acquired through dialogues. To further improve the performance, medical knowledge graph was incorporated into the dialogue model (Xu et al., 2019). AI technologies have also been explored in the diagnosis of psychological issues. For example, Washington et al. (2020) summarized and categorized how different data-driven methods could be applied in autism for digital therapeutic phenotyping in computational psychiatry.

By reviewing existing systems, we can see that task-oriented dialogue systems have been utilized in various domains. However, they have seldom been adopted for

problem behavior management in moral education. Identifying need deficiencies is similar to diagnosing medical diseases, which inspires us to employ this technology to solve the problem of need deficiency identification. In this work, we propose building a dialogue system for need deficiency identification and explore the feasibility of applying AI technology to solve psychological problems.

Theoretical framework

This section presents a theoretical framework that summarizes relevant factors for identifying the need deficiencies underlying problem behaviors (RQ1). Extensive research has been conducted to analyze problem behaviors and need deficiencies. These findings are informative for uncovering the reasons for problem behavior but are also too scattered to be systematically employed by amateurs. Therefore, an integrated framework summarizing the relevant factors is essential for need deficiency identification. Theoretically, the number of factors relevant to students' problem behaviors and need deficiencies is large, and such factors might often influence each other. In practice, it is still challenging to cover all the relevant factors and meanwhile fully capture the interplay between them. Therefore, the proposed framework mainly focuses on the key factors and simplifies their joint effects on the students' problem behaviors.

Maslow (2013) suggested that psychological needs mainly drive behaviors, and students' problem behaviors are caused by the lack of satisfaction concerning their psychological needs. Therefore, in attempts to intervene in students' problem behaviors, the related influencing factors should be considered. Jessor (2010) proposed a problem behavior theory with three systems: the personality system, the perceived environment system, and the behavior system. Additionally, he pointed out that the personality system and the perceived environment system interacted and affected students' problem behaviors. Based on these research findings, we define a framework for need deficiency identification, which provides a consolidated classification of the types of need deficiencies as well as a systematic classification of the relevant factors from three categories: problem behaviors, external environmental factors, and internal individual factors.

Classification of need deficiencies

Need deficiencies are categorized according to Maslow's hierarchy of needs. In the "The Theory of Human Motivation," Maslow systematically explained

Table 1 Classification of need deficiencies

	Category
Need deficiencies	Physiological need deficiency
	Safety need deficiency
	Belongingness and love need deficiency
	Esteem need deficiency
	Cognition need deficiency

the hierarchy of needs theory by discussing physiological needs, safety needs, belongingness and love needs, esteem needs, and self-realization needs (Maslow, 2013). In this framework, we replace the self-realization needs with the cognitive needs, considering that the demand for self-realization refers to a stage fusing goodness and beauty, which usually appears in later life stages. In contrast, cognitive needs represent the need for understanding the surrounding world, which is more common for primary and secondary school students. A detailed classification of need deficiencies is shown in Table 1.

Classification of problem behaviors

Problem behaviors are classified mainly according to the CBCL Teacher Report Form (Achenbach, 2015). However, some modifications are applied based on real-life case analysis. Specifically, problem behaviors are classified into three categories: externalization problems, internalization problems, and other problems. Externalization problems are the "externalization syndrome" of behaviors, which refer to social adaptation problems, including attack, bullying, sabotage, and others. Externalization problems are further divided into aggressive behaviors and rule-breaking behaviors. Internalization problems are the "internalization syndrome" of behaviors, which refer to emotional distress problems or nonsocial behavioral problems, including anxiety and depression. Internalization problems are further divided into social withdrawal, depression, and anxiety. Problems that do not belong to the above two categories are called "other problems," including learning problems, egocentricity, and special problems. A detailed classification of problem behaviors is shown in Table 2.

Classification of external environmental factors

External environmental factors mainly denote environmental factors affecting students' behaviors. Based on existing research findings, we categorize external environmental factors into factors related to the family environment, school environment, and social environment. According to previous theoretical and empirical research, we summarize the specific factors of the

Table 2 Classification of problem behaviors

	Category	Specific factor
Problem behaviors	Externalizing problems	Aggressive behavior, rule-breaking behavior
	Internalizing problems	Social withdrawal, depression, anxiety
	Other problems	Learning problems, egocentricity, special problems

family environment as family structure, parenting style, parents' education background, parents' health condition, parents' delinquent behavior, and parents' socioeconomic status. The school environment also has a significant influence on the formation of students' behavior. Based on previous research, we summarize the specific factors of the school environment as teacher leadership style, peer acceptance, and peer influence. The social environment concerns the influence of society. In this framework, we emphasize the influence of social media and cultural customs. A detailed classification of external environmental factors is shown in Table 3.

Classification of internal individual factors

Problem behaviors are influenced not only by external environmental factors but also by students' internal individual factors. Specifically, information about internal individual factors can be classified into two categories: demographic information and personality information. Demographic information denotes students' characteristics that are closely related to problem behaviors. In this work, we define the specific factors of demographic information as grade, gender, health condition, and social group. In addition, the personality characteristics of individuals are significantly correlated with their problem behaviors (Ehrler et al., 1999). In this framework, we employ the five-factor model of personality to represent students' personalities: neuroticism, extroversion, openness, agreeableness, and conscientiousness. A detailed classification of internal individual factors is shown in Table 4.

Method

Participants

The participants included 223 primary school students (7- to 13-year-old pupils, $M_{age} = 9.98$, $SD_{age} = 1.80$), 207 middle school students (13- to 16-year-old middle school students, $M_{age} = 14.23$, $SD_{age} = 0.85$), and 198 high school students (16- to 19-year-old high school students, $M_{age} = 16.13$, $SD_{age} = 1.08$). The participants included 450 boys and 178 girls. The data were mainly obtained from real-life cases collected by the "Chinese Good Teachers" public platform of Beijing Normal University (<http://haolaoshi.bnu.edu.cn/>).

Ultimately, 628 effective real-life cases were collected, which constitute the data for this study. Subsequently, two domain experts were invited to manually and independently annotate the cases according to the proposed theoretical framework. Specifically, aspects of students' problem behaviors, external environmental factors, internal individual factors, and need deficiencies were labeled. The kappa value between these two annotations is 0.83.

Considering the sensitivity of moral education and possible ethical concerns, we employed the following strategies to protect privacy. Before the data collection process, all the participants, including the students, their parents, and the teachers, were explicitly informed that the collected data would be used for the research purpose. The participants were given the option to opt-out of the data sharing before signing the data usage agreement. After the data were collected, we also carefully notified each participant and removed those who felt uncomfortable being involved. To further protect privacy, we irreversibly anonymized the identities and randomly sampled only a subset of the participants to build our system. In addition, the collected real-life student cases were not directly exposed to system users but were utilized only to train the dialogue model which infers the students' need deficiencies based on the learned rules. Furthermore, our dialogue system questions users only about students' behaviors and some relevant factors in the practical application. Personal characteristics such as names and schools are not required; hence, the system cannot identify any specific students. In addition, the system infers students' need deficiencies in real-time without storing any of the students' information. The data will not be provided to any schools, governments, or other power holders.

Material

To measure the effectiveness of the task-oriented dialogue system on the task of automatic need deficiency identification, we develop a task-oriented dialogue system based on the summarized theoretical framework and test it with real-life cases (RQ2). Through multi-turn dialogue, the proposed task-oriented dialogue system can acquire various aspect information of students and infer their need deficiencies accordingly.

Table 3 Classification of external environmental factors

	Category	Specific Factor
External environmental factors	Family environment	Family structure, parenting style, parents' education background, parents' health condition, parents' delinquent behaviors, parents' socioeconomic status
	School environment	Teacher leadership style, peer acceptance, peer influence
	Social environment	Mass media, cultural customs

Table 4 Classification of internal individual factors

	Category	Specific factor
Internal individual factors	Demographic information	Grade, gender, health condition, health condition, social group
	Personality information	Neuroticism, extraversion, openness, agreeableness, conscientiousness

To build a task-oriented dialogue system, one fundamental job is to define the required information to complete the targeted task. That information is the basis for the dialogue state and system actions. The required information will be defined in this work according to the factors summarized in the proposed theoretical framework. Specifically, 29 different factors, including problem behaviors, external environmental factors, and internal individual factors, are summarized for need deficiency identification. These factors also define the dialogue state representation. In addition, system actions mainly include two categories: *request* and *inform*. The *request* action aims to request specific information about students, such as checking whether they have exhibited aggressive behaviors. In total, 29 specific request actions are defined according to the 29 factors. The *inform* action aims to inform users about the specific need deficiencies of students, such as telling that they have a deficiency in esteem needs. Specifically, five different inform actions are defined according to the five types of need deficiencies.

Based on the information defined according to the theoretical framework, we design and implement the dialogue system. As shown in Figure 1, the proposed dialogue system consists of four main functional modules: NLU, DST, PL, and NLG (Chen et al., 2017). The NLU module processes a user's reply to extract the student's information, such as whether he has aggressive behaviors. In the developed dialogue system, the long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997) network is adopted to generate the semantic vector for user input. An LSTM network is a typical recurrent neural network that has been widely used in natural language processing recently. Relying on a gating mechanism, the LSTM can solve the long-term dependency issue in sequential data processing. Hence, we utilize the

LSTM to interpret the user's utterances in this dialogue system. The DST module updates the dialogue state with another LSTM network based on the semantic vector of user input. This dialogue state represents students' information acquired to that point and is utilized by the PL module to compute the next system action.

Based on the current dialogue state, we adopt a reinforcement learning model, specifically a deep Q-learning network (DQN) model (Mnih et al., 2015), to learn dialogue policy that decides whether to request more information from the user or to present the derived need deficiency to the user. As one of the three main paradigms of machine learning, reinforcement learning targets solving sequential decision-making problems. Recently, deep learning techniques have been integrated into reinforcement learning models to improve model performance. The DQN is a typical deep reinforcement learning model that utilizes a deep neural network to calculate the Q-value in the Q-learning model (Mnih et al., 2015). With the generated system actions, the NLG module utilizes a template-based model to generate natural language responses for the user. This system can acquire different information about a student through multiturn dialogue and automatically infer his need deficiencies underlying the problem behaviors.

A task-oriented dialogue system is often trained with user simulators. In this work, we also utilize a user simulator to train the dialogue model. The system architecture of system training is elaborated in Figure 1, which consists of four main components. The *schema* is defined according to the theoretical framework summarizing all the factors relating to identifying need deficiency, which is also the basis for defining the case structure and system actions. The *case database* manages real-life cases collected from the online platform and annotated according to the

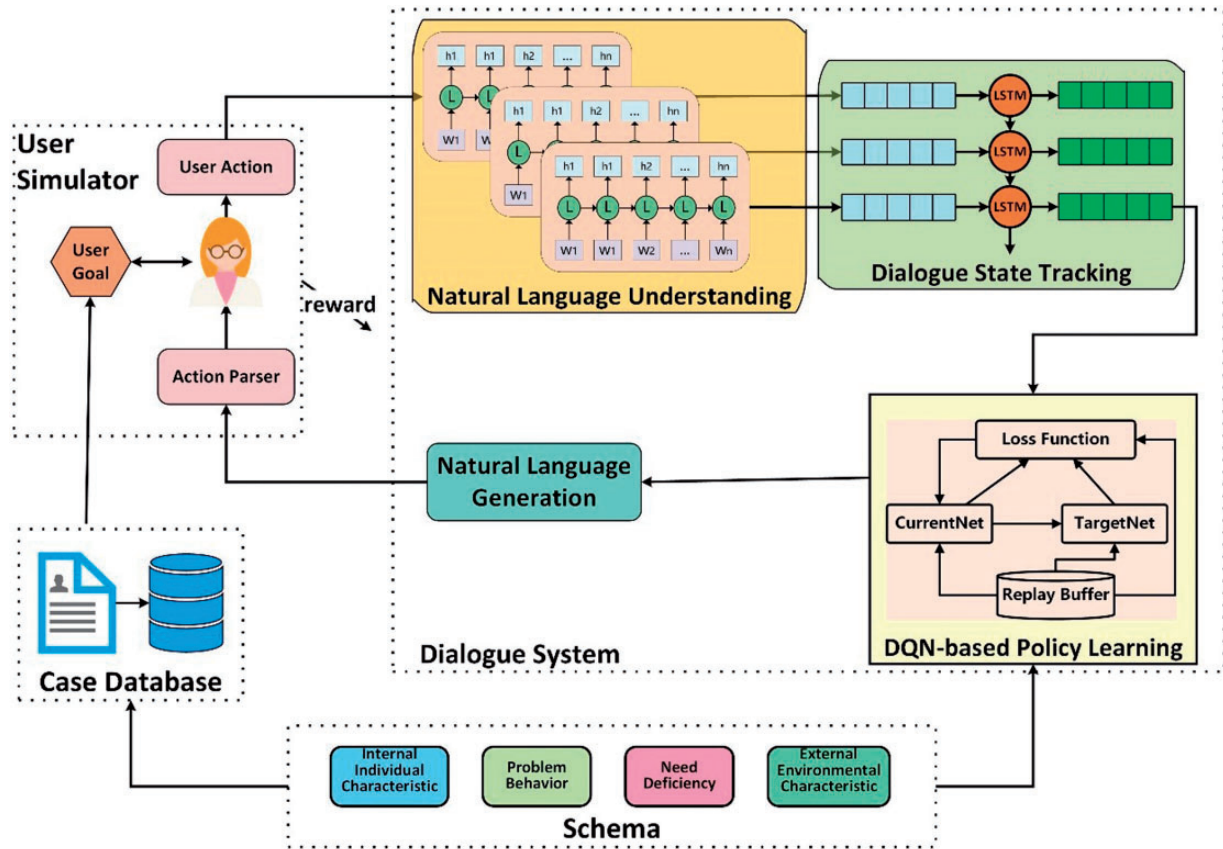


Figure 1 Architecture of proposed task-oriented dialogue system.

schema. These cases summarize real-life experiences for identifying need deficiencies. The *user simulator* emulates a user and interacts with the dialogue system based on information provided by the case randomly selected from the database. The *dialogue system* acquires students' information and identifies the need deficiencies underlying their problem behaviors. Utilizing such an architecture, we train our dialogue system and evaluate whether a task-oriented dialogue system can automatically identify students' need deficiencies.

Procedure

We measure whether a task-oriented dialogue system is feasible for automatically identifying students' need deficiencies by checking the system's performance on unseen real-life cases (RQ2). Specifically, before a new dialogue session starts, a real-life case is randomly selected to test the system. In the first dialogue turn, the dialogue system randomly chooses one factor to request from the user. For example, the system may issue a question as "Does the student have any aggressive behavior?" After checking the student's behaviors

in the chosen case, the user replies upon the student's problem behaviors as "The student often fights with his classmates." Interpreting this reply, the system knows that the student has aggressive behaviors and updates the dialogue state. Subsequently, according to the dialogue state, the dialogue system chooses another factor to request from the user. For example, the system requests the student's gender by issuing a question as "Is the student a boy or a girl?" Accordingly, based on information recorded in the case, the user replies that "The student is a boy." The dialogue continues by requesting different information about the student. When enough information is obtained, the dialogue system may infer and inform that the student has deficiencies in belongingness and love needs. By comparing this response with the real need deficiencies outlined in the case, we know whether the system's identification result is right or wrong. Repeating this procedure using different real-life cases, we compute system performance according to the recorded results.

To further evaluate the performance of the dialogue system, we compare the dialogue system with different models. Specifically, three different models are selected as baselines. First, we compare the system performance

with *multiclass classification models of machine learning*. These models treat the identification of need deficiency as a multiclass classification problem. Utilizing all the factors summarized by the theoretical framework as model features, those models employ different algorithms to calculate the probabilities of different need deficiencies and choose the one with maximum probability as the student's need deficiency. In this work, the decision tree (DT) (Safavian & Landgrebe, 1991) and gradient boosting (GB) (Friedman, 2002) models are selected as baselines. The DT model tries to learn different rules to classify each datum into different categories, which inherently supports multiclassification. The GB model is an ensemble machine-learning model that ensembles several weak classifiers to improve classification accuracy. Both DT and GB are typical classification models in machine learning. Second, we compare the dialogue system to a *dialogue system built with a rule-based policy*. This policy utilizes handcrafted rules to make decisions on the next system action. Whenever new factor information is obtained, it checks whether the information captured is enough to infer the need deficiency. If more information is required, it requests the factor that differentiates different need deficiencies the most as the next action. Third, we compare the dialogue system to a *dialogue system built with random policy*. This policy selects the next system action randomly from the action space to respond to the user action at each dialogue turn.

Three different experiments are conducted to measure system performance. First, experiments are conducted to check the system performance of different-sized training data. Specifically, the experiments are conducted with 50%, 60%, 70%, 80%, 90% of the data for training, with the rest for testing. Second, we conduct experiments to examine the effects of different parameter settings on system performance. Specifically, we conduct experiments with epoch numbers and rewards. Epoch numbers represent the number of iterations to train the dialogue system; we check how the performance varies from 1 epoch to 500 epochs. Reward setting defines the different costs and benefits of different system actions, and we measure the average number of turns with different reward settings. Third, one exemplary dialogue is analyzed to demonstrate the diagnosing process.

Indices of system performance

To evaluate the performance of the dialogue system, we compare the performance with baselines according to three commonly used metrics. *Success rate* denotes the percentage of identifying the need deficiencies with the right result in the total number of tested cases. This metric is identical to the accuracy metric used by the

classification models. *Average reward* denotes the average reward value for all testing cases. Each system action has a reward value, in which a negative reward denotes a penalty. The correct identification of a need deficiency results in a positive reward, and incorrect identification results in a negative reward. A large reward value denotes better performance. *Average turns* denote the average number of dialogue turns for all the testing cases. A small number of turns means the system can accurately select essential information to request and infer the need deficiency accordingly.

Results

System performance

The results are presented in Table 5, where *success* denotes the success rate metric, *reward* denotes the average reward obtained for different policies, and *turns* denotes the average turns in each dialogue session.

Impact of system parameters

The result regarding how system performance changes with the number of epochs is shown in Figure 2. The success rate is low at the beginning epochs, but increases as the number of epochs increases and then turns to flat starting at 300. The result of the average number of turns with different reward settings is shown in Figure 3. When the positive reward has an equal or larger absolute value compared to the negative reward, the average number of turns is small, which means the dialogue does not continue very well. On the other hand, if the negative reward has a larger absolute value, the average number of turns is higher, which implies the dialogue continues well.

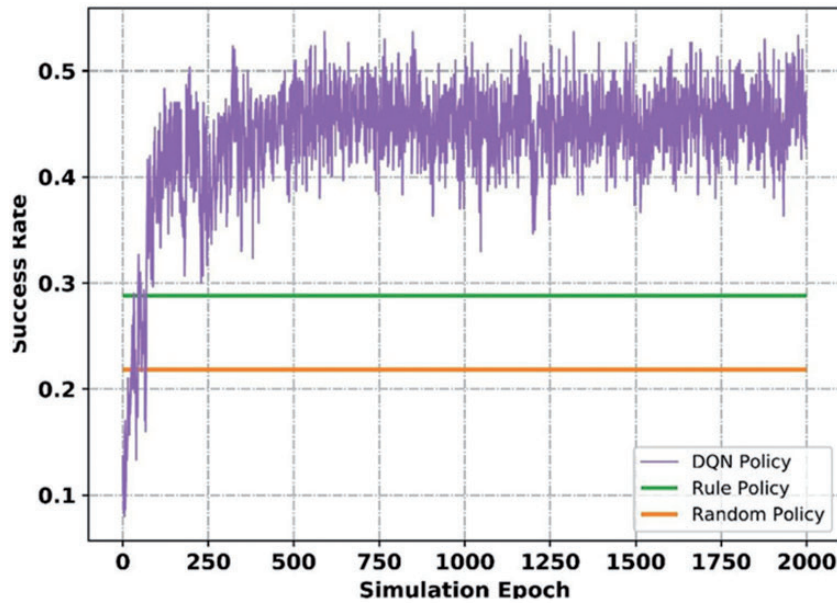
Exemplary dialogue

An example illustrating the interactive dialogue between the system and the user is shown in Table 6. In this example, the system initiates the dialogue by requesting the gender of the student. The reply "boy" determines the next essential factor in identifying a need deficiency is whether this student engages in aggressive behavior. With the answer "yes," the system infers that information about his parents' parenting style is needed. Following the same process, the system requests information about other factors (e.g., family structure and peer acceptance). It gives back a final result based on the information requested: the student has deficiencies in belongingness and love needs. Checking details of the dialogue, we can learn that the student's parents have adopted an *uninvolved*

Table 5. Experimental results of different baselines.

Model	Training data size				
	50%	60%	70%	80%	90%
DT					
Success	0.48 ± 0.03	0.50 ± 0.04	0.53 ± 0.06	0.54 ± 0.08	0.55 ± 0.06
GB					
Success	0.61 ± 0.05	0.63 ± 0.03	0.66 ± 0.05	0.67 ± 0.07	0.67 ± 0.03
Random policy					
Success	0.22 ± 0.05	0.21 ± 0.11	0.22 ± 0.12	0.20 ± 0.16	0.22 ± 0.12
Reward	-67.98 ± 2.08	-70.64 ± 1.02	-67.00 ± 2.01	-71.72 ± 5.02	-67.28 ± 5.32
Turns	26.19 ± 1.22	26.34 ± 1.08	25.67 ± 2.09	26.75 ± 3.02	25.82 ± 5.02
Rule policy					
Success	0.29 ± 0.02	0.28 ± 0.05	0.29 ± 0.06	0.30 ± 0.03	0.28 ± 0.05
Reward	-231.34 ± 3.05	-236.60 ± 3.02	-230.51 ± 5.09	-227.36 ± 4.02	-236.67 ± 3.02
Turns	124.88 ± 1.03	126.39 ± 2.02	125.01 ± 3.03	125.02 ± 3.02	124.75 ± 2.02
DQN policy					
Success	0.41 ± 0.08	0.43 ± 0.05	0.43 ± 0.06	0.44 ± 0.05	0.44 ± 0.08
Reward	-17.90 ± 2.08	-15.33 ± 3.12	-15.53 ± 2.02	-15.70 ± 1.82	-14.66 ± 2.02
Turns	5.26 ± 1.03	6.09 ± 2.03	7.44 ± 2.02	9.61 ± 2.02	11.11 ± 3.02

Note: DT denotes the decision tree classification model. GB denotes the gradient boosting ensemble learning model.

**Figure 2** Impact of simulation epoch number.

parenting style, and his family structure is a *one-parent family*. Additionally, his peer acceptance level at school is classified as *neglected*. With all the available information and research findings on emotional abandonment by parents (Harper & Ibrahim, 1999; Richards, 1999), it is reasonable to conclude that the student's love and belongingness needs might be unsatisfied because he tries to obtain attention and care from others by fighting with peers.

Discussion

Based on the experimental results, we report several significant findings. First, compared with the baseline models, the dialogue system can achieve similar performance with a success rate between 0.4 and 0.44 for different-sized training data. This finding suggests that our system is effective in identifying need deficiencies. Second, the dialogue system returns a result within

just 12 turns on average, which means it has successfully recognized the essential factors to request and infer the type of need deficiencies. Third, the developed dialogue system significantly outperforms the random policy and rule-based policy, proving the effectiveness of utilizing the DQN model in learning dialogue policy.

The experimental results demonstrate how the dialogue system identifies students' need deficiencies is consistent with psychological research findings. For example, Harper and Stone found that the method of satisfying one's basic needs is often influenced by race, social class, economic, political, as well as environmental resources (Harper & Stone, 2003). The needs for belongingness and love denote the needs to belong to and feel loved by a group, such as one's family,

religious group, workgroup, professional group, social club or fraternity, or even one's youth gang. Therefore, emotional abandonment by parents or emotional rejection by parents and peers will lead to unsatisfied belonging and love needs. Emotional abandonment by parents further leads to severe psychological pain and injury, which may result in distorted impulses to hurt others or hurt themselves (Harper & Ibrahim, 1999; Richards, 1999). In addition, other basic needs, such as self-esteem, refers to the self-esteem of one's achievements, as well as behaviors worthy of respect from others based on one's accomplishments, status, or appearance (Corr et al., 2015). The need for self-esteem will not be satisfied if teachers disrespect or psychologically reject students (Thompson & Rudolph, 1992). The underlying inferring logic of the dialogue system is consistent with those findings.

The experimental results also show that the dialogue system technology can be effectively utilized as a diagnostic tool to solve psychological problems such as need deficiency identification through integration with psychology theories. First, through multiturn dialogue interaction, a task-oriented dialogue system can automatically acquire information of student's characteristics and systematically diagnose the need deficiencies, which demonstrates that the dialogue system is effective in accomplishing the need deficiency identification task.

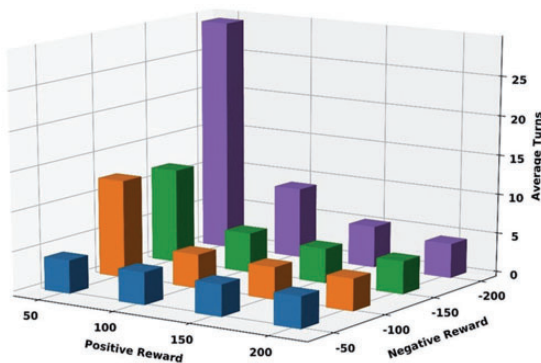


Figure 3 Impact of reward setting.

Table 6. Example of multiturn dialogue.

Turn	Role	Utterance	Action
1	Sys.	Welcome, may I know the gender of your student?	request(gender)
2	User	He is a boy.	inform(gender=boy)
3	Sys.	Got it! Has he engaged in aggressive activities like fighting with peers?	request(aggressive behavior)
4	User	Yes. He usually fights with others or even bullies youngsters.	inform(aggressive behavior = True)
5	Sys.	Okay. Do you know which kind of parenting style his family has adopted?	request(parenting style)
6	User	Emmmm, I am not sure, but I heard that his parents seldom take care of him. He has to live on his own most of the time.	inform(parenting style = uninvolved)
7	Sys.	It seems that his parents are uninvolved. And what is the structure of his family?	request(family structure)
8	User	As far as I know, his parents are divorced. He lives with his father, but his father is not often at home.	inform(family structure = One-parent family)
9	Sys.	It seems that he needs more care from us. What is the level of the student's peer acceptance in school?	request(peer acceptance)
10	User	The class committee reports that he is introverted and easily overlooked by peers.	inform(peer acceptance = Neglected)
11	Sys.	Alright, based on the information you provided, this student might be unsatisfied with belongingness and love needs.	inform(deficiency = Belongingness and Love needs)

Second, the proposed theoretical framework is important for identifying need deficiencies utilizing AI technologies. Both the dialogue system and classification models are built according to the theoretical framework, which ensures that the processes of analyzing students' information and deriving the need deficiencies are consistent with psychological findings. This guarantees the theoretical correctness of the system. The accuracy of the testing with the unseen cases also proves the effectiveness of the dialogue system.

Third, compared to other methods, a dialogue system requires less information to derive the type of need deficiencies, which significantly reduces service costs and improves service applicability. For a typical method such as a multiclassifier, in order to derive the need deficiency for a specific student, it should acquire the student's full information on all aspects, even though some aspects may not be necessary for developing an inference for this student. In contrast, the dialogue system infers a student's need deficiencies based on essential information only. Based on the user's reply in each dialogue turn, the system can adaptively decide the next important piece of information it needs to acquire and then determines the need deficiency when enough information is captured.

Fourth, the natural language interaction significantly improves service usability. With the traditional manner, teachers and parents must master and apply the psychology theories based on a limited understanding. In contrast, using this dialogue system, users just need to provide and describe the students' information according to their observations and obtain suggestions regarding the possible reasons underlying problem behaviors from the system without necessarily knowing all of the theories. In addition, such a system can be easily deployed to an online platform or integrated into a mobile application, which can significantly improve its scalability and service accessibility. Therefore, AI technologies can be appropriately applied to solve psychological problems such as need deficiency diagnosis. This motivates us to explore and adopt appropriate AI technologies to solve psychological problems that are difficult to diagnose using traditional approaches.

Comparing with the existing and successful dialogue systems for restaurant reservation or flight booking, the proposed system still has its limitations in terms of system performance and model design. It is mainly because the task of student's need deficiency identification is much more complex, and concerns many psychological factors that are not easy to measure and capture. For the tasks like restaurant reservation or flight booking, the required information and the logic model behind are usually easy to obtain and clear, and accordingly the system performance can be guaranteed. Hence, more research should be carried out to design

better models from both psychological and technical sides, especially on capturing the joint effects of psychological factors on students' need deficiencies, which is one of the main tasks in our future work.

Last but not least, it is necessary to alert the potential ethical and policy issues on utilizing the proposed dialogue system. Similar to other AI-driven systems, building and operating such intelligent agents would continuously collect a large amount of private user data and obtain sensitive analytical results. The corresponding information and the analytical results have the potential risks to be abused for surveillance and other purposes. Hence, privacy-preserving techniques and policies are crucial to protect students' privacy and avoid unnecessary surveillance in practice. For example, anonymization and encryption techniques should be applied to the entire process of data collection and management. Meanwhile, the global laws or regulations should be made regarding disclosure, consent, access, and reasonable usage of such AI-driven systems, similar to the general data protection regulation (GDPR) in Europe. If necessary, such AI-driven systems can be operated and managed by non-profit organizations rather than local schools or governments, where the practitioners should ensure no improper behavior norms and controls are imposed.

Conclusion and future work

In this work, we answered two research questions. One question concerned how do we define a theoretical framework for analyzing students' problem behavior and need deficiencies systematically. By summarizing relevant factors explored in previous psychological research, we classify need deficiencies into five categories: physiological needs, safety needs, belongingness and love needs, esteem needs, and cognitive needs. Additionally, we categorize the factors for identifying these need deficiencies into three groups: problem behaviors, external environmental factors, and internal individual factors. The proposed framework further classifies these groups into specific factors. The second question concerned whether AI technology can be employed to diagnose students' need deficiencies automatically. Specifically, based on factors defined in the theoretical framework, we developed a task-oriented dialogue system to identify need deficiencies. The experimental results demonstrated that a developed dialogue system could effectively acquire different information about students and diagnose the students' need deficiencies through multiturn dialogue interaction. An analysis of the diagnosis dialogue shows that the dialogue system's inferences regarding need deficiencies are consistent with psychological theories. In the future, the developed dialogue system will be

deployed on a nationwide online platform for moral education to verify its effectiveness on a large scale.

Acknowledgements

This research is partially supported by the National Natural Science Foundation of China (No. 61807003 and No. 62077006), and the Fundamental Research Funds for the Central Universities.

ORCID iD

Penghe Chen  <https://orcid.org/0000-0003-1894-3552>

References

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: Department of Psychiatry, University of Vermont.
- Achenbach, T. M. (2015). *Achenbach system of empirically based assessment (ASEBA)*. The encyclopedia of clinical psychology.
- Achenbach, T. M., Ruffle, T. M. (2000). The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in Review*, 21(8), 265–271. <https://doi.org/10.1542/pir.21-8-265>
- Chen, H., Liu, X., Yin, D., Tang, J. (2017). A survey on dialogue systems: Recent advances and new frontiers. *Acm Sigkdd Explorations Newsletter*, 19(2), 25–35. <https://doi.org/10.1145/3166054>. 3166058
- Corr, P. J., Deyoung, C. G., McNaughton, N. (2015). Motivation and personality. *Social & Personality Psychology Compass*, 7(3), 158–175. <https://doi.org/10.1111/spc3.12016>
- Dennis, J. M., Phinney, J. S., Chuateco, L. I. (2005). The role of motivation, parental support, and peer support in the academic success of ethnic minority first-generation college students. *Journal of college student development*, 46(3), 223–236. <https://doi.org/10.1353/csd.2005.0023>
- Dhingra, B., Li, L., Li, X., Gao, J., Chen, Y., Ahmed, F., et al. (2016). Towards end-to-end reinforcement learning of dialogue agents for information access. *arXiv preprint arXiv:1609.00777*.
- Ehrler, D. J., Evans, J. G., McGhee, R. L. (1999). Extending Big-Five theory into childhood: A preliminary investigation into the relationship between Big-Five personality traits and behavior problems in children. *Psychology in the Schools*, 36(6), 451–458. [https://doi.org/10.1002/\(SICI\)1520-6807\(199911\)36:6<451::AID-PITS1>3.0.CO;2-E](https://doi.org/10.1002/(SICI)1520-6807(199911)36:6<451::AID-PITS1>3.0.CO;2-E)
- Farrell, A. D., Thompson, E. L., Mehari, K. R. (2017). Dimensions of peer influences and their relationship to adolescents' aggression, other problem behaviors and prosocial behavior. *Journal of Youth & Adolescence*, 46(6), 1351–1369. <https://doi.org/10.1007/s10964-016-0601-4>
- Fomby, P., Christie, A. S. (2013). Family structure instability and mobility: The consequences for adolescents' problem behavior. *Social Science Research*, 42(1), 186–201. <https://doi.org/10.1016/j.ssresearch.2012.08.016>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Stats & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Gao, J., Galley, M., Li, L. (2019). *Neural approaches to conversational AI: Question answering, task-oriented dialogues and social chatbots*. Paper presented at the Now Foundations and Trends. <http://dx.doi.org/10.1561/15000000074>
- Harper, F. D., Harper, J. A., Stills, A. B. (2003a). Counseling children in crisis based on Maslow's Hierarchy of Basic Needs. *International Journal for the Advancement of Counselling*, 25(1), 11–25. <https://doi.org/10.1023/A:1024972027124>
- Harper, F. D., Ibrahim, F. A. (1999). Violence and schools in the USA: Implications for counseling. *International Journal for the Advancement of Counselling*, 21(4), 349–366. <https://doi.org/10.1023/A:1005651113740>
- Harper, F. D., Stone, W. O. (2003). Transcendent counseling: An existential, cognitive-behavioral theory. *Culture and counseling: New approaches*, 9, 233–251. <https://doi.org/10.1007/BF00120245>
- Hoffmann, J. P. (2006). Family structure, community context, and adolescent problem behaviors. *Journal of Youth and Adolescence*, 35(6), 867–880. <https://doi.org/10.1007/s10964-006-9078-x>
- Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Jessor, R. (2010). Problem-behavior theory, psychosocial development, and adolescent problem drinking. *Addiction*, 82(4), 331–342. <https://doi.org/10.1111/j.1360-0443.1987.tb01490.x>
- Jeynes, W. H. (2019). A meta-analysis on the relationship between character education and student achievement and behavioral outcomes. *Education and Urban Society*, 51(1), 33–71. <https://doi.org/10.1177/0013124517747681>
- Kao, H., Tang, K., Chang, E. Y. (2018, 2018-01-01). *Context-aware symptom checking for disease diagnosis using hierarchical reinforcement learning*. Paper presented at the In Thirty-Second AAAI Conference on Artificial Intelligence. <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17130/15944>
- Kazdin, A. E., Glick, A., Pope, J., Kaptchuk, T. J., Lecza, B., Carrubba, E., et al. (2018). Parent management training for conduct problems in children: Enhancing treatment to improve therapeutic change. *International Journal of Clinical and Health Psychology*, 18(2), 91–101. <https://doi.org/10.1016/j.ijchp.2017.12.002>
- Koh, C. (2012). Moral development and student motivation in moral education: A Singapore study. *Australian Journal of Education*, 56(1), 83–101. <https://doi.org/10.1177/000494411205600107>
- Lee, W. O., Ho, C. H. (2005). Ideopolitical shifts and changes in moral education policy in China. *Journal of Moral Education*, 34(4), 413–431. <https://doi.org/10.1080/03057240500410160>
- Leijten, P., Melendez Torres, G. J., Gardner, F., Van Aar, J., Schulz, S., Overbeek, G. (2018). Are relationship enhancement and behavior management “The Golden Couple” for

- disruptive child behavior? Two meta-analyses. *Child development*, 89(6), 1970–1982. <https://doi.org/10.1111/cdev.13051>
- Li, X., Chen, Y., Li, L., Gao, J., Celikyilmaz, A. (2017). End-to-end task-completion neural dialogue systems. *arXiv preprint arXiv:1703.01008*.
- Maggs, J. L., Galambos, N. L. (1993). Alternative structural models for understanding adolescent problem behavior in two-earner families. *The Journal of Early Adolescence*, 13(1), 79–101. <https://doi.org/10.1177/0272431693013001005>
- Maryam, Z., Ali, M., Kaisa, A., Jari-Erik, N., Noona, K., Marja-Kristiina, L. (2019). Supportive parenting buffers the effects of low peer acceptance on children's internalizing problem behaviors. *Child & Youth Care Forum*, 48(6), 865–887. <https://doi.org/10.1007/s10566-019-09510-y>
- Maslow, A. H. (2013). *A theory of human motivation*. Simon and Schuster.
- McGrath, P. J., Elgar, F. J. (2015). Effects of socio-economic status on behavioral problems. *International Encyclopedia of the Social & Behavioral Sciences*, 2(2), 477–482. <https://doi.org/10.1016/B978-0-08-097086-8.23052-3>
- Miller, L. C. (1968). Louisville behavior check list for males, 6–12 years of age. *Psychological Reports*, 21(3), 885–896. <https://doi.org/10.2466/pr0.1967.21.3.885>
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533. <https://doi.org/10.1038/nature14236>
- Papangelis, A., Papadakos, P., Stylianou, Y., Tzitzikas, Y. (2018). *Spoken dialogue for information navigation*. Paper presented at the In Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue (pp. 229–234). <https://www.aclweb.org/anthology/W18-5025>
- Peng, Y., Tang, K., Lin, H., Chang, E. (2018). *Refuel: Exploring sparse features in deep reinforcement learning for fast disease diagnosis*. Paper presented at the In Advances in Neural Information Processing Systems (pp. 7322–7331). <https://dl.acm.org/doi/10.5555/3327757.3327834>
- Peterson, Donald, R. (1961). Behavior problems of middle childhood. *Journal of Consulting Psychology*, 25(3), 205–209. <http://dx.doi.org/10.1037/h0038994>
- Pinquart, Martin. (2017). Associations of parenting dimensions and styles with internalizing symptoms in children and adolescents: A meta-analysis. *Marriage & Family Review*, 53(7), 613–640. <https://doi.org/10.1080/01494929.2016.1247761>
- Purwati, P., Japar, M. (2017). Parents' education, personality, and their children's disruptive behaviour. *International Journal of Instruction*, 10(3), 227–240. <https://doi.org/10.12973/iji.2017.10315a>
- Richards, B. M. (1999). Suicide and internalised relationships: A study from the perspective of psychotherapists working with suicidal patients. *British Journal of Guidance & Counselling*, 27(1), 85–98. <https://doi.org/10.1080/03069889908259717>
- Rosario, C., Gutiérrez-Cobo, M. J., Fernández-Berrocal, P. (2017). Parental education and aggressive behavior in children: A moderated-mediation model for inhibitory control and gender. *Frontiers in Psychology*, 8, 1181. <https://doi.org/10.3389/fpsyg.2017.01181>
- Safavian, S. R., Landgrebe, D. (1991). A survey of Decision Tree Classifier methodology. *IEEE Transactions on Systems Man and Cybernetics*, 21(3), 660–674. <https://doi.org/10.1109/21.97458>
- Scarpaci, R. T. (2007). IOSIE: A method for analyzing student behavioral problems. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 80(3), 111–116. <https://doi.org/10.3200/TCHS.80.3.111-116>
- Schuitema, J., Dam, G. T., Veugelers, W. (2008). Teaching strategies for moral education: A review. *Journal of curriculum studies*, 40(1), 69–89. <https://doi.org/10.1080/00220270701294210>
- Sieh, D. S., Meijer, A. M., Oort, F. J., Visser-Meily, J., Van der Leij, D. (2010). Problem behavior in children of chronically ill parents: a meta-analysis. *Clinical child and family psychology review*, 13(4), 384–397. <https://doi.org/10.1007/s10567-010-0074-z>
- Sieh, D. S., Visser-Meily, J. M. A., Oort, F. J., Meijer, A. M. (2012). Risk factors for problem behavior in adolescents of parents with a chronic medical condition. *European Child & Adolescent Psychiatry*, 21(8), 459–471. <https://doi.org/10.1007/s00787-012-0279-4>
- Spiller, A. N., Caro, K., Garay, J. A., Marcu, G. (2019, 2019–01–01). *Supporting behavior management with a classroom display providing immediate feedback to students*. Paper presented at the In Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare (pp. 159–168). <https://doi.org/10.1145/3329189.3329201>
- Tang, K., Kao, H., Chou, C., Chang, E. Y. (2016). *Inquire and diagnose: Neural symptom checking ensemble using deep reinforcement learning*. Paper presented at the In Proceedings of NIPS Workshop on Deep Reinforcement Learning. http://infolab.stanford.edu/~echang/NIPS_DeepRL_2016_Symptom_Checker.pdf
- Thompson, C. L., Rudolph, L. B. (1992). *Counseling children (3rd ed.)*. Thomson Brooks/Cole Publishing Co.
- Van, D. A. A. L., Prinzie, P., Dekovi, M., De Haan, A. D., Asscher, J. J., Widiger, T. (2013). The development of personality extremity from childhood to adolescence: Relations to internalizing and externalizing problems. *Journal of Personality & Social Psychology*, 105(6), 1038–1048. <https://doi.org/10.1037/a0034441>
- Wahab, N. A., Othman, M. S., Muhammad, N. (2017). The influence of the mass media in the behavior of students: A literature study. *International Journal of Academic Research in Business and Social Sciences*, 7(8), 166–174. <http://dx.doi.org/10.6007/IJARBS/v7-i8/3218>
- Washington, P., Park, N., Srivastava, P., Voss, C., Kline, A., Varma, M., et al. (2020). Data-driven diagnostics and the potential of mobile artificial intelligence for digital therapeutic phenotyping in computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(8), 759–769. <https://doi.org/10.1016/j.bpsc.2019.11.015>

- Wei, Z., Liu, Q., Peng, B., Tou, H., Chen, T., Huang, X., et al. (2018, 2018-01-01). Task-oriented dialogue system for automatic diagnosis. Paper presented at the In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 201-207). <https://www.aclweb.org/anthology/P18-2033>
- Wen, T., Vandyke, D., Mrksic, N., Gasic, M., Rojas-Barahona, L. M., Su, P., et al. (2016). A network-based end-to-end trainable task-oriented dialogue system. *arXiv preprint arXiv:1604.04562*.
- Xu, L., Zhou, Q., Gong, K., Liang, X., Tang, J., Lin, L. (2019). *End-to-end knowledge-routed relational dialogue system for automatic diagnosis*. Paper presented at the In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 7346-7353). <https://doi.org/10.1609/aaai.v33i01.33017346>
- Yan, Z., Duan, N., Chen, P., Zhou, M., Zhou, J., Li, Z. (2017). *Building task-oriented dialogue systems for online shopping*. Paper presented at the In Thirty-First AAAI Conference on Artificial Intelligence. <https://ojs.aaai.org/index.php/AAAI/article/view/11182>