Review on Neural Question Generation for Education Purposes: Supplementary Materials

Said Al Faraby[1], Adiwijaya[1*] and Ade Romadhony[1]

[1*]School of Computing, Telkom University, Bandung, Indonesia.

*Corresponding author(s). E-mail(s):
adiwijaya@telkomuniversity.ac.id;
Contributing authors: saidalfaraby@telkomuniversity.ac.id;
aderomadhony@telkomuniversity.ac.id;

Types of Questions

Table 1: Taxonomy Inference Questions according to (Chikalanga, 1992)

| Basic Category	Specific Types	Question-Answer Relationship
Lexical	Pronominal inferences Ambiguous / unfamiliar word meanings	Textually / scriptally Implicit
Propositional	Logical Informational - Referential - Spatio-temporal Logical Explanatory - Motivational - Causative Enablement	Textually Implicit
Pragmatic	Elaborative Informational - Referential - Spatiotemporal Elaborative Explanatory - Motivational - Causative - Enablement Evaluative	Scriptally Implicit

The taxonomy above also groups the types of inference based on whether or not external knowledge is required to answer the questions. *Textually implicit* means that the answer itself can be found in the text, although readers are required to make an inference first. *Scriptally implicit* means that the answer is not in the text but in the mind of the readers, and is based on their knowledge and experience.

Included literature

Table 2: Excluded literature statistics

No	Exclusion Criteria	Number of Excluded Literature.
1	PDF not found	44
2	Not question generation	13
3	Not automatic	7
4	Not neural	79
5	Secondary study	4
6	Low quality	1

Rate of publication
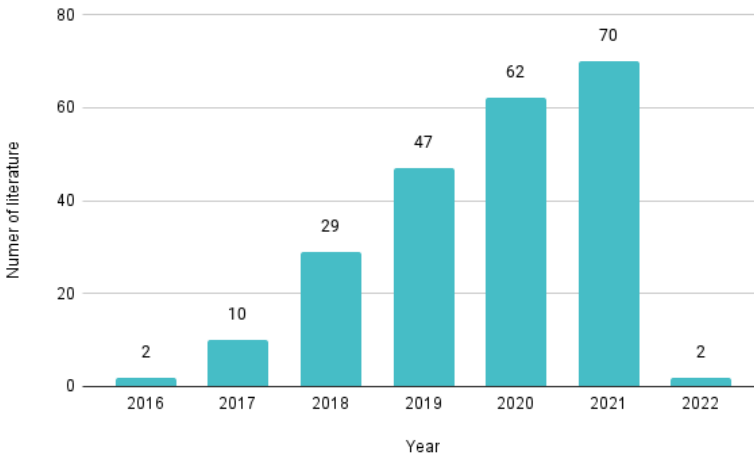
Fig. 1: Number of NQG studies published by year.

Publication venue

Table 3: Top publication venues.

No	Publication Venue	Number of studies
1	Conference on Empirical Methods in Natural Language Processing (EMNLP)	29
2	Annual Meeting of the Association for Computational Linguistics (ACL)	24
3	Arxiv	23
4	Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (ACL-IJCNLP)	11
5	Conference on Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)	10
6	AAAI Conferenceon Artificial Intelligence (AAAI)	9
7	International World Wide Web Conference (WWW)	8
8	Conference on Computational Linguistics (COLING)	7
9	Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)	7
10	The European Chapter of the ACL (EACL)	6
11	International Natural Language Generation Conference (INLG)	6
12	Natural Language Processing and Chinese Computing (NLPCC)	6
13	Workshop on Machine Reading for Question Answering (MRQA)	5
14	International Joint Conference on Artificial Intelligence (IJCAI)	4
15	Thesis	4
16	ACM International Conference on Information and Knowledge Management (ACM-CIKM)	3
17	ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR)	3
18	International Journal of Artificial Intelligence in Education (IJAIED) - Springer	2
19	Chinese National Conference on Computational Linguistics (CCL)	2
20	Conference on Computational Natural Language Learning (CoNLL)	2
21	International Conference on Learning Representations (ICLR)	2
22	ACM Conference on Learning at Scale (ACM-L@S)	2
23	Transactions of the Association for Computational Linguistics (TACL)	2
24	Workshop on Neural Generation and Translation (WNGT)	2

Existing Work on NQG for Education Purposes

Table 4: Comparison of literature that are closely related to educational purposes. D=Dataset, C=Context, A=Answer, QG=Method for question generation, AE=Method for Answer Extraction/Generation, IN=Input Representation, Ev=Evaluator

Lit	D, C, A	Method	QT	Eval	Edu Rel	Objectives
(Z. Wang et al., 2018)	D: SQuAD (tr), Openstax (ts) C: Text, 1-5 sentences A: Short span (word, phrase)	QG: LSTM-based Encoder-Decoder	Factoid	Ev: Crowdsourced Criteria: Fluency, relevance, human-like	Evaluation with data from education domain	Naturalness
(Willis et al., 2019)	D: SQuAD C: Text, paragraph A: Abstractive	AE: LSTM-based Encoder-Decoder QG : QG-Net	Factoid	Ev: Domain Expert Criteria: Matching extracted keywords	Using domain expert (classroom teacher) for evaluation. Answer Extraction method that is more correlated with experts' answers.	Usefulness
(S. Cao & Wang, 2021)	D: Developed from Reddit and Yahoo A : Long, Abstractive	IN: Semantic graph, Exemplar, Question Template, QG: BART	Verification, Disjunctive, Concept, Extent, Example, Comparison, Cause, Consequence, Procedural, Judgmental	Ev: - Criteria: Diversity: - Type - Syntax - Answer content Content quality: - Appropriateness, - Answerability, - Scope	The type of question comes from cognitive science	Naturalness Diversity Controllability (Question type)

Lit	D, C, A	Method	QT	Eval	Edu Rel	Objectives
(Steuer, Filighera, Meuser, & Rensing, 2021)	D: OpenStax A: Long, Extractive	AE: Linguistics rule-based QG:	Definitional	Ev: Expert Criteria: Horbach scheme	Data from the education domain Question types come from cognitive science Evaluation related to education	Naturalness Usefulness
(Jia, Zhou, Sun, & Wu, 2021)	D: RACE C: Passage A: Abstractive	IN: Dependency graph QG: LSTM-based Encoder Decoder	Factoid and non-factoid	Ev: - Criteria: Fluency, Relevancy, Answerability	Data from education domain	Naturalness
(Steuer, Filighera, & Rensing, 2020)	D: SQuAD (tr) RACE (ts), LearningQ (ts) C: Passage A: Extractive	IN: Dependency graph AE: Linguistics rule-based QG: fine-tuned UniLM	Factoid	Ev : 2 Annotators Criteria: Grammar Answerability Usefulness	Data from education domain	Naturalness
(Cheng, Ding, et al., 2021)	D: Wikipedia article C: Passage A: Short, Extractive	Summarization AE: fine-tuned T5 QG: fine-tuned T5 Filtering	Factoid	Ev: Crowdsourced 50 people Criteria: - Usefulness - Comprehensibility - Correctness	Application related to education, Evaluation usefulness Subject : history, geography	Naturalness, Controllability (level of context detail)

Lit	D, C, A	Method	QT	Eval	Edu Rel	Objectives
(Murakhovs'ka, Wu, Niu, Liu, & Xiong, 2021)	D: SQuAD, NewsQA, TriviaQA, SearchQA, HotpotQA, NQ, NarQA, MCTest, BoolQ, Quoref, DROP, QAConv C : Passage A: Short, Long, Extractive, Abstractive	QG: fine-tuned T5, fine-tuned BART	Factoid and non-factoid (including Yes-No)	Ev: Author Criteria: - Fluency - Relevancy	Can generate varied cognitive levels questions	Naturalness, Controllability (difficulty)
(Krishna & Iyyer, 2019)	D: SQuAD, QuAC, CoQA C:Passage A: Extractive	QG: LSTM-based Encoder Decoder	Specific (Factoid), General and Yes-No (Non-Factoid)	Ev: Crowd sourced Criteria: - Fluency - Relevancy - Answerability	Can generate varied cognitive levels questions	Naturalness, Controllability (specificity)
(Stasaski, Rathod, Tu, Xiao, & Hearst, 2021)	D: SQuAD, TQA C: Passage A: Extractive	AE: Linguistics rule-based QG: ProphetNet	Cause, Consequence	Ev: Crowd sourced Criteria: - Correct question type - Matching Answer	Data from education domain Subject: Life Science, Earth Science, Physical Science	Naturalness
(Qu, Jia, & Wu, 2021)	D: SQuAD, RACE C: Passage A: Generative	AE: ProphetNet QG: ProphetNet	Non-Factoid	Ev: 3 evaluators Criteria: - Fluency - Relevance - Answerability	Data from education domain	Naturalness

Tasks

Table 5: NQG tasks related to education purposes

Task	Sub-category	Citation
Question Generation from Reading Material	Context domain	General Q. Zhou et al. (2017), Du, Shao, and Cardie (2017), Pan, Xie, Feng, Chua, and Kan (2020), Gupta, Chauhan, Akella, Ekbal, and Bhattacharyya (2020), etc. Privacy policy (Lamba & Hsu, 2021) Children storybook (Yao et al., 2022) University Subjects Z. Wang et al. (2018), Steuer et al. (2021) Financial Jayakumar, Krishnakumar, Peddagopu, and Sridhar (2020), Car manual M. Delpisheh (2020), Mahdavi, An, Davoudi, Delpisheh, and Gohari (2020), School Science Stasaski et al. (2021) English Exam (Jia et al., 2021)
	Output or application	Flashcard (Cheng, Ding, et al., 2021) Interactive Reading(Syed et al., 2020) Question generator website (Fung, Kwok, Lee, Chui, & U, Leong Hou, 2020) Quiz for News (Lelkes, Tran, & Yu, 2021)
Word Problem Generation	Context domain	Mathematical (Q. Zhou & Huang, 2019), (T. Cao, Zeng, Zhao, Mansur, & Chang, 2021), (Z. Wang, Lan, & Baraniuk, 2021), (T. Liu et al., 2021) Mathematical statistics (Keller, 2021) SQL Guo et al. (2018), Yu and Jiang (2021)
Conversation	Output or application	Forum/chat interaction (Y. Wang, Liu, Huang, & Nie, 2018), (W. Wang, Feng, Wang, & Zhang, 2019), (Ling, Cai, Chen, & de Rijke, 2020), (J. Lee, Liang, & Fong, 2021), (Shen, Meng, Zhang, Feng, & Zhou, 2021) Sequential questions (Nakanishi, Kobayashi, & Hayashi, 2019), (Gao, Li, King, & Lyu, 2019), (Y. Wang, Rong, Zhang, Zhou, & Xiong, 2020), (Chai & Wan, 2020), (Gu, Mirshekari, Yu, & Sisto, 2021) Information seeking (Qi, Zhang, & Manning, 2020), (Scialom & Staiano, 2020) Interview (Su, Wu, Huang, Hong, & Huang, 2018), (B, Agnihotri, & Jayagopi, 2020), (Rao S B, Agnihotri, & Babu Jayagopi, 2021) Clarification (Y.T. Cao, Rao, & Daumé, 2019), (Zamani, Dumais, Craswell, Bennett, & Lueck, 2020), (Sekulić, Aliannejadi, & Crestani, 2021), (Majumder, Rao, Galley, & McAuley, 2021)

Datasets

Table 6: List of top datasets and education-related NQG literatures that used them

Dataset	Freq	Edu QG Studies
SQuAD (Rajpurkar, Zhang, Lopyrev, & Liang, 2016)	105	(Fung et al., 2020; Krishna & Iyyer, 2019; Syed et al., 2020; Willis et al., 2019)
MS MARCO (Nguyen et al., 2016)	19	
HotpotQA (Z. Yang et al., 2018)	17	
NewsQA (Trischler et al., 2016)	12	
NaturalQuestions (Kwiatkowski et al., 2019)	11	
SimpleQuestions (Bordes, Usunier, Chopra, & Weston, 2015)	5	
TriviaQA (Joshi, Choi, Weld, & Zettlemoyer, 2017)	5	(G. Chen, Yang, & Gasevic, 2019)
BioASQ (Tsatsaronis et al., 2015)	5	
WikiQA (Y. Yang, Yih, & Meek, 2015)	3	
LearningQ (G. Chen, Yang, Hauff, & Houben, 2018)	3	(G. Chen et al., 2019, 2018; Steuer et al., 2020)
PathQuestions (M. Zhou, Huang, & Zhu, 2018)	3	
QuAC (Choi et al., 2018)	3	(Krishna & Iyyer, 2019)
KorQuAD (Lim, Kim, & Lee, 2019)	3	
RACE (Lai, Xie, Liu, Yang, & Hovy, 2017)	3	(G. Chen et al., 2019; Jia et al., 2021; Steuer et al., 2020)
WebQuestionsSP (Yih, Richardson, Meek, Chang, & Suh, 2016)	3	
WikiSQL (Zhong, Xiong, & Socher, 2017)	3	
BoolQ (Clark et al., 2019)	2	
CoQA (Reddy, Chen, & Manning, 2019)	2	(Krishna & Iyyer, 2019)
OpenStax (Rice University, 1999)	2	(Steuer et al., 2021; Z. Wang et al., 2018)
DROP (Dua et al., 2019)	2	
Hi-QuAD (Kumar, Joshi, Mukherjee, Ramakrishnan, & Jyothi, 2019)	2	
Car Manual (E. Delpisheh et al., 2019)	2	
Dolphin18K (D. Huang, Shi, Lin, Yin, & Ma, 2016)	2	(T. Cao et al., 2021)
NewsQuizQA	1	(Lelkes et al., 2021)
FairyTaleQA	1	(Yao et al., 2022)
MCTest	1	(G. Chen et al., 2019)
Inquisitive (Ko, Chen, Huang, Durrett, & Li, 2020)	1	(Ko et al., 2020)
SQUASH (Krishna & Iyyer, 2019)	1	(Krishna & Iyyer, 2019)
Arithmetic (Hosseini, Hajishirzi, Etzioni, & Kushman, n.d.)	1	(Z. Wang et al., 2021)

Dataset	Freq	Edu QG Studies
MAWPS (Koncel-Kedziorski, Roy, Amini, Kushman, & Hajishirzi, 2016)	1	(Z. Wang et al., 2021)
Math32K (Y. Wang, Liu, & Shi, 2017)	1	(Z. Wang et al., 2021)
Reddit Custom (S. Cao & Wang, 2021)	1	(S. Cao & Wang, 2021)

Answers

Table 7: Various combinations of answer characteristics used in the literature

Gen/Giv	AA/AU	In/Out	Ex/Ab	Sh/Lg	Form	Citation
Giv	AA	In	Ex	Sh	Free	(Q. Zhou et al., 2017), (Song, Wang, Hamza, Zhang, & Gildea, 2018), (Sun et al., 2018), (Kim, Lee, Shin, & Jung, 2019), (L. Dong et al., 2019), (Y. Chen, Wu, & Zaki, 2019), (Ma, Zhu, Zhou, & Li, 2020), (Y. Chen, Wu, & Zaki, 2020), (S. Wang et al., 2020), (Yuan, He, & Dai, 2021), etc
			Ab	Lg	Free	(Jia et al., 2021)
			Ex, Ab	Sh, Lg	MC, YN, Free	(Murakhovs'ka et al., 2021), (Yuan et al., 2022)
		Out	Ab	Lg	Free	(S. Cao & Wang, 2021)*, (Mishra et al., 2020)*
Gen	AA	In	Ex	Sh	Free	(Z. Wang et al., 2018), (Z. Yang, Hu, Salakhutdinov, & Cohen, 2017), (Reddy, Raghu, Khapra, & Joshi, 2017), (Subramanian et al., 2018), (Alberti, Andor, Pitler, Devlin, & Collins, 2019), (Jayakumar et al., 2020), (Fung et al., 2020), (Steuer et al., 2021)
			Ab	Sh	Free	(Willis et al., 2019)
	AU	In	Ex	Sh	Free	(Cui et al., 2021)

Gen/Giv	AA/AU	In/Out	Ex/Ab	Sh/Lg	Form	Citation
					MC	(Lelkes et al., 2021)
-	AU	In	Ex	Sh	Free	(Du et al., 2017), (Scialom, Piwowarski, & Staiano, 2019), (Lopez, Cruz, Cruz, & Cheng, 2021), (X. Wu, Jiang, & Wu, 2020)
			Ab	Sh, Lg	Free	(Krishna & Iyyer, 2019)
		Out	-	-	Free	(Y. Wang et al., 2018), (Zhu et al., 2019), (Nakanishi et al., 2019), (Rao S B et al., 2021)

System Objectives

Table 8: Objectives of NQG systems and their representative approaches

Goal	Feature	Method
Naturalness	Feature engineering	Linguistics features (eg, PoS, Named-Entity, SRL, Dependency, Syntactic) (Q. Zhou et al., 2017), (Du & Cardie, 2018),(Ji, Lyu, Cao, & Cheng, 2021) ,(Pan et al., 2020)
		Answer position (Sun et al., 2018), (B. Liu et al., 2019), (B. Liu, Wei, Niu, Chen, & He, 2020), (Q. Huang et al., 2021), (Yin, Zhou, Small, & May, 2021)
		Placeholder strategy (Scialom et al., 2019)
	Use more specific information	Predict additional keywords from context before generating (Hu, Liu, Ma, Zhao, & Yan, 2018), (B. Liu et al., 2019), (B. Liu et al., 2020)
		Predict question word before generating (X. Dong et al., 2018), (Kang, Puerto San Roman, & Myaeng, 2019; W. Zhou, Zhang, & Wu, 2019)
		Extract answers-relevant relation before generating (Li, Gao, Bing, King, & Lyu, 2019)
		Use path in knowledge graph as input to generator (S. Wang et al., 2020)
		Continued on next page

---

[0] Codes: Giv=answers are given, Gen=answers are generated, AA=answer-aware, AU=answer-unaware, In=answers are inside context, Out=answers are outside context, Ex=extractive, Ab=abstractive, Sh=short, Lg=long, Fill=fill-in-the-blank, MC=multiple-choice, Y/N=yes/no, Free=free-text, *=answer-only without context.

Table 8 – continued from previous page

Goal	Feature	Method
	Add context from external knowledge	Add more context from knowledge bases (C. Liu, Liu, He, Nie, & Zhao, 2019) Incorporate external knowledge (Xin, Hao, Dawei, & Yunfang, 2021), (M. Delpisheh, 2020)
	Improve encoder	Separating and masking answer words to avoid copying answer words(Kim et al., 2019) Improve passage-answer interaction via encoder (Y. Chen et al., 2019), (Z. Liu, Huang, Huang, & Zhao, 2020)
	Improve decoder	Partial copy allowing morphological changes (Qiu & Xiong, 2019) 2nd decoder for refinement (Nema, Mohankumar, Khapra, Srinivasan, & Ravindran, 2019) Improve passage-answer interaction via decoder (L. Wang, Xu, Lin, Zheng, & Shen, 2020) Remember what has been copied or generated (Benmalek, Khabsa, Desu, Cardie, & Banko, 2019), (Fei, Zhang, & Zhou, 2021)
	Additional learning method	Adding RL rewards (paraphrase, answerability, fluency, or relevance) (S. Zhang & Bansal, 2019), (Xie, Pan, Wang, Kan, & Feng, 2020) Multi-task to overcome wrong question copied (Tuan, Shah, & Barzilay, 2020) Contrastive learning (W.S. Cho et al., 2021) Combine different QG models with teacher-student learning (Kang, Hong, Puerto San Roman, & Myaeng, 2020) Coreference resolution to previous question (Gao, Li, et al., 2019) Question ranker based on answer (Qiu & Xiong, 2019), (W. Wang et al., 2019) Question ranker based on well-formed (Faruqui & Das, 2018)
Usefulness	Without external information	Learning from sentence contains answer (Du & Cardie, 2017) Graph-based sentence importance (LexRank) (G. Chen et al., 2019) Use linguistics rule to select pedagogically valuable target answer (Yao et al., 2022)
	Require external information	Use learning objective as additional input (Shimmei & Matsuda, 2021) Use linguistics rule to select pedagogically valuable target answer (Yao et al., 2022)
	~~Global~~	Question templates (Yu & Jiang, 2021)

Diversity

Table 8 – continued from previous page

Goal	Feature	Method
		Classify input words into several types and use typed decoders (Y. Wang et al., 2018)
	Local	Sample/predict several target answers (Harrison & Walker, 2018; K. Wu, Hong, Zhu, Tang, & Zhang, 2019)
		Sample/predict several contents from context (J. Cho, Seo, & Hajishirzi, 2019; Z. Zhang & Zhu, 2021), (Z. Wang et al., 2020), (Z. Zhang & Zhu, 2021)
		Variational decoder (Bahuleyan, Mou, Vechtomova, & Poupart, 2018; Guo et al., 2018; D.B. Lee, Lee, Jeong, Kim, & Hwang, 2020), (Shinoda, Sugawara, & Aizawa, 2021)
		Use several entailed texts from context (Matsumoto, Hasegawa, Yamakawa, & Mitamura, 2018)
		Use paraphrasing (Jia, Zhou, Sun, & Wu, 2020), (D. Liu et al., 2020)
		Sampling at inference (Sultan, Chandel, Astudillo, & Castelli, 2020)
		Use different answer-relevant relations as input (Li et al., 2019)
		Predict and use several question words (X. Wu et al., 2020), (Z. Wang et al., 2020),
Controllability	Specific-type	13 question words, by initializing decoder with question word (Z. Zhang, 2020)
		10 question types, by templates (S. Cao & Wang, 2021)
	Difficulty-level	2 levels, determined by QA system, controlled by a value to initialize decoder (Gao, Bing, Chen, Lyu, & King, 2019)
		2 levels, determined by linguistic rules, controlled by one-hot vector to decoder (Kumar, Hua, et al., 2019)
		n levels , determined by number of hops, controlled by number of iteration of the 2nd decoder (Cheng, Li, et al., 2021)
		2 levels, determined by length of answer, controlled by length of answer (Murakhovs'ka et al., 2021)
Personalization		Difficulty based on student knowledge tracing (Srivastava & Goodman, 2021)
		Based on readers' background (Stewart & Mihalcea, 2021)

References

Alberti, C., Andor, D., Pitler, E., Devlin, J., Collins, M. (2019, July). Synthetic QA corpora generation with roundtrip consistency. *Proceedings of the*

57th annual meeting of the association for computational linguistics (pp. 6168–6173). Florence, Italy: Association for Computational Linguistics.

B, P.R.S., Agnihotri, M., Jayagopi, D.B. (2020, September). Automatic follow-up question generation for asynchronous interviews. , 10–20.

Bahuleyan, H., Mou, L., Vechtomova, O., Poupart, P. (2018, August). Variational attention for Sequence-to-Sequence models. *Proceedings of the 27th international conference on computational linguistics* (pp. 1672–1682). Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Benmalek, R., Khabsa, M., Desu, S., Cardie, C., Banko, M. (2019, July). Keeping notes: Conditional natural language generation with a scratchpad encoder. *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 4157–4167). Florence, Italy: Association for Computational Linguistics.

Bordes, A., Usunier, N., Chopra, S., Weston, J. (2015, June). Large-scale simple question answering with memory networks. https://arxiv.org/abs/1506.02075 [cs.LG]

Cao, S., & Wang, L. (2021, August). Controllable open-ended question generation with a new question type ontology. *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 6424–6439). Online: Association for Computational Linguistics.

Cao, T., Zeng, S., Zhao, S., Mansur, M., Chang, B. (2021). Generating math word problems from equations with topic consistency maintaining and commonsense enforcement. *Artificial neural networks and machine learning – ICANN 2021* (pp. 66–79). Springer International Publishing.

Cao, Y.T., Rao, S., Daumé, H., III. (2019). Controlling the specificity of clarification question generation. *WNLP@ ACL* (pp. 53–56). winlp.org.

Chai, Z., & Wan, X. (2020). Learning to ask more: Semi-autoregressive sequential question generation under dual-graph interaction. *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 225–237).

Chen, G., Yang, J., Gasevic, D. (2019). A comparative study on Question-Worthy sentence selection strategies for educational question generation. *Artificial intelligence in education* (pp. 59–70). Springer International Publishing.

Chen, G., Yang, J., Hauff, C., Houben, G.-J. (2018, June). LearningQ: A Large-Scale dataset for educational question generation. *Twelfth international AAAI conference on web and social media.*

Chen, Y., Wu, L., Zaki, M.J. (2019, September). Reinforcement learning based Graph-to-Sequence model for natural question generation. *International conference on learning representations.*

Chen, Y., Wu, L., Zaki, M.J. (2020, April). Toward subgraph guided knowledge graph question generation with graph neural networks. https://arxiv.org/abs/2004.06015 [cs.CL]

Cheng, Y., Ding, Y., Pascual, D., Richter, O., Volk, M., Wattenhofer, R. (2021). WikiFlash: Generating flashcards from wikipedia articles. *AAAI 2021 workshop on AI education-35th AAAI conference on artificial intelligence (AAAI).* tik-old.ee.ethz.ch.

Cheng, Y., Li, S., Liu, B., Zhao, R., Li, S., Lin, C., Zheng, Y. (2021, August). Guiding the growth: Difficulty-Controllable question generation through Step-by-Step rewriting. *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 5968–5978).

Chikalanga, I. (1992). A suggested taxonomy of inferences for the reading teacher. *Reading in Foreign Language*, *8*(2), 697–709.

Cho, J., Seo, M., Hajishirzi, H. (2019, November). Mixture content selection for diverse sequence generation. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 3121–3131). Hong Kong, China: Association for Computational Linguistics.

Cho, W.S., Zhang, Y., Rao, S., Celikyilmaz, A., Xiong, C., Gao, J., . . . Dolan, B. (2021, April). Contrastive multi-document question generation. *Proceedings of the 16th conference of the european chapter of the association for computational linguistics: Main volume* (pp. 12–30). Online: Association for Computational Linguistics.

Choi, E., He, H., Iyyer, M., Yatskar, M., Yih, W.-T., Choi, Y., . . . Zettlemoyer, L. (2018). QuAC: Question answering in context. *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 2174–2184). Brussels, Belgium: Association for Computational Linguistics.

Clark, C., Lee, K., Chang, M.-W., Kwiatkowski, T., Collins, M., Toutanova, K. (2019, May). BoolQ: Exploring the surprising difficulty of natural Yes/No questions.
  https://arxiv.org/abs/1905.10044 [cs.CL]

Cui, S., Bao, X., Zu, X., Guo, Y., Zhao, Z., Zhang, J., Chen, H. (2021, February). OneStop QAMaker: Extract Question-Answer pairs from text in a One-Stop approach.
  https://arxiv.org/abs/2102.12128 [cs.CL]

Delpisheh, E., Alzghool, M., An, A., Davoudi, H., Delpisheh, M., Gohari, E., Mahdavi, S. (2019, August). *Context-aware question and answer generation from car manuals* (Tech. Rep.). York University.

Delpisheh, M. (2020). *Neural question generation with transfer learning and utilization of external knowledge* (Unpublished doctoral dissertation). York University, Toronto.

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., . . . Hon, H.-W. (2019). Unified language model pre-training for natural language understanding and generation. *Advances in neural information processing systems* (Vol. 32). Curran Associates, Inc.

Dong, X., Hong, Y., Chen, X., Li, W., Zhang, M., Zhu, Q. (2018). Neural question generation with semantics of question type. *Natural language processing and chinese computing* (pp. 213–223). Springer International Publishing.

Du, X., & Cardie, C. (2017). Identifying where to focus in reading comprehension for neural question generation. *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2067–2073).

Du, X., & Cardie, C. (2018, July). Harvesting paragraph-level Question-Answer pairs from Wikipedia. *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1907–1917). Melbourne, Australia: Association for Computational Linguistics.

Du, X., Shao, J., Cardie, C. (2017). Learning to ask: Neural question generation for reading comprehension. *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1342–1352). Stroudsburg, PA, USA: Association for Computational Linguistics.

Dua, D., Wang, Y., Dasigi, P., Stanovsky, G., Singh, S., Gardner, M. (2019, June). DROP: A reading comprehension benchmark requiring discrete reasoning over paragraphs. *Proceedings of the 2019 conference of the*

north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers) (pp. 2368–2378). Association for Computational Linguistics.

Faruqui, M., & Das, D. (2018). Identifying well-formed natural language questions. *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 798–803). Brussels, Belgium: Association for Computational Linguistics.

Fei, Z., Zhang, Q., Zhou, Y. (2021, November). Iterative GNN-based decoder for question generation. *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 2573–2582). Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Fung, Y.-C., Kwok, J.C.-W., Lee, L.-K., Chui, K.T., U, Leong Hou. (2020). Automatic question generation system for english reading comprehension. *Technology in education. innovations for online teaching and learning* (pp. 136–146). Springer Singapore.

Gao, Y., Bing, L., Chen, W., Lyu, M., King, I. (2019, August). Difficulty controllable generation of reading comprehension questions. *Proceedings of the Twenty-Eighth international joint conference on artificial intelligence* (pp. 4968–4974). California: International Joint Conferences on Artificial Intelligence Organization.

Gao, Y., Li, P., King, I., Lyu, M.R. (2019, July). Interconnected question generation with coreference alignment and conversation flow modeling. *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 4853–4862). Florence, Italy: Association for Computational Linguistics.

Gu, J., Mirshekari, M., Yu, Z., Sisto, A. (2021, April). ChainCQG: Flow-Aware conversational question generation. *Proceedings of the 16th conference of the european chapter of the association for computational linguistics: Main volume* (pp. 2061–2070). Online: Association for Computational Linguistics.

Guo, D., Sun, Y., Tang, D., Duan, N., Yin, J., Chi, H., . . . Zhou, M. (2018). Question generation from SQL queries improves neural semantic parsing. *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 1597–1607). Brussels, Belgium: Association for Computational Linguistics.

Gupta, D., Chauhan, H., Akella, R.T., Ekbal, A., Bhattacharyya, P. (2020, December). Reinforced multi-task approach for multi-hop question

generation. *Proceedings of the 28th international conference on computational linguistics* (pp. 2760–2775). Barcelona, Spain (Online): International Committee on Computational Linguistics.

Harrison, V., & Walker, M. (2018, November). Neural generation of diverse questions using answer focus, contextual and linguistic features. *Proceedings of the 11th international conference on natural language generation* (pp. 296–306). Tilburg University, The Netherlands: Association for Computational Linguistics.

Hosseini, Hajishirzi, Etzioni, Kushman. (n.d.). Learning to solve arithmetic word problems with verb categorization. *EMNLP*.

Hu, W., Liu, B., Ma, J., Zhao, D., Yan, R. (2018, February). Aspect-based question generation.

Huang, D., Shi, S., Lin, C.-Y., Yin, J., Ma, W.-Y. (2016, August). How well do computers solve math word problems? Large-Scale dataset construction and evaluation. *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 887–896). Berlin, Germany: Association for Computational Linguistics.

Huang, Q., Fu, M., Mo, L., Cai, Y., Xu, J., Li, P., ... others (2021). Entity guided question generation with contextual structure and sequence information capturing. *Proc. Conf. AAAI Artif. Intell.*.

Jayakumar, H., Krishnakumar, M.S., Peddagopu, V.V.V., Sridhar, R. (2020, October). RNN based question answer generation and ranking for financial documents using financial NER. *Sādhanā*, *45*(1), 269.

Ji, T., Lyu, C., Cao, Z., Cheng, P. (2021, October). Multi-Hop question generation using hierarchical Encoding-Decoding and context switch mechanism. *Entropy*, *23*(11), 1449.

Jia, X., Zhou, W., Sun, X., Wu, Y. (2020). How to ask good questions? try to leverage paraphrases. *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 6130–6140).

Jia, X., Zhou, W., Sun, X., Wu, Y. (2021). EQG-RACE: Examination-Type question generation. *AAAI*. aaai.org.

Joshi, M., Choi, E., Weld, D., Zettlemoyer, L. (2017, July). TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension. *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1601–1611). Vancouver, Canada: Association for Computational Linguistics.

Kang, J., Hong, G., Puerto San Roman, H., Myaeng, S.-H. (2020, November). Regularization of distinct strategies for unsupervised question generation. *Findings of the association for computational linguistics: EMNLP 2020* (pp. 3266–3277). Online: Association for Computational Linguistics.

Kang, J., Puerto San Roman, H., Myaeng, S.-H. (2019, November). Let me know what to ask: Interrogative-Word-Aware question generation. *Proceedings of the 2nd workshop on machine reading for question answering* (pp. 163–171). Hong Kong, China: Association for Computational Linguistics.

Keller, S.U. (2021, September). Automatic generation of word problems for academic education via natural language processing (NLP). https://arxiv.org/abs/2109.13123 [cs.CL]

Kim, Y., Lee, H., Shin, J., Jung, K. (2019, July). Improving neural question generation using answer separation. *AAAI*, *33*(01), 6602–6609.

Ko, W.-J., Chen, T.-Y., Huang, Y., Durrett, G., Li, J.J. (2020, November). Inquisitive question generation for high level text comprehension. *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 6544–6555). Online: Association for Computational Linguistics.

Koncel-Kedziorski, R., Roy, S., Amini, A., Kushman, N., Hajishirzi, H. (2016, June). MAWPS: A math word problem repository. *Proceedings of the 2016 conference of the north American chapter of the association for computational linguistics: Human language technologies* (pp. 1152–1157). San Diego, California: Association for Computational Linguistics.

Krishna, K., & Iyyer, M. (2019, July). Generating Question-Answer hierarchies. *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 2321–2334). Florence, Italy: Association for Computational Linguistics.

Kumar, V., Hua, Y., Ramakrishnan, G., Qi, G., Gao, L., Li, Y.-F. (2019). Difficulty-Controllable multi-hop question generation from knowledge graphs. *The semantic web – ISWC 2019* (pp. 382–398). Springer International Publishing.

Kumar, V., Joshi, N., Mukherjee, A., Ramakrishnan, G., Jyothi, P. (2019, July). Cross-Lingual training for automatic question generation. *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 4863–4872). Florence, Italy: Association for Computational Linguistics.

Kwiatkowski, T., Palomaki, J., Redfield, O., Collins, M., Parikh, A., Alberti, C., ... Others (2019). Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*, *7*, 453–466.

Lai, G., Xie, Q., Liu, H., Yang, Y., Hovy, E. (2017, September). RACE: Large-scale ReAding comprehension dataset from examinations. *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 785–794). Copenhagen, Denmark: Association for Computational Linguistics.

Lamba, D., & Hsu, W.H. (2021). Answer-Agnostic question generation in privacy policy domain using Sequence-to-Sequence and transformer models. *kdd.cs.ksu.edu*.

Lee, D.B., Lee, S., Jeong, W.T., Kim, D., Hwang, S.J. (2020, July). Generating diverse and consistent QA pairs from contexts with Information-Maximizing hierarchical conditional VAEs. *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 208–224). Online: Association for Computational Linguistics.

Lee, J., Liang, B., Fong, H. (2021, August). Restatement and question generation for counsellor chatbot. *Proceedings of the 1st workshop on NLP for positive impact* (pp. 1–7). Online: Association for Computational Linguistics.

Lelkes, A.D., Tran, V.Q., Yu, C. (2021, April). Quiz-Style question generation for news stories. *Proceedings of the web conference 2021* (pp. 2501–2511). New York, NY, USA: Association for Computing Machinery.

Li, J., Gao, Y., Bing, L., King, I., Lyu, M.R. (2019, November). Improving question generation with to the point context. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 3216–3226). Hong Kong, China: Association for Computational Linguistics.

Lim, S., Kim, M., Lee, J. (2019, September). KorQuAD1.0: Korean QA dataset for machine reading comprehension.

https://arxiv.org/abs/1909.07005 [cs.CL]

Ling, Y., Cai, F., Chen, H., de Rijke, M. (2020, April). Leveraging context for neural question generation in open-domain dialogue systems. *Proceedings of the web conference 2020* (pp. 2486–2492). New York, NY, USA: Association for Computing Machinery.

Liu, B., Wei, H., Niu, D., Chen, H., He, Y. (2020, April). Asking questions the human way: Scalable Question-Answer generation from text corpus. *Proceedings of the web conference 2020* (pp. 2032–2043). New York, NY, USA: Association for Computing Machinery.

Liu, B., Zhao, M., Niu, D., Lai, K., He, Y., Wei, H., Xu, Y. (2019, May). Learning to generate questions by LearningWhat not to generate. *The world wide web conference* (pp. 1106–1118). New York, NY, USA: Association for Computing Machinery.

Liu, C., Liu, K., He, S., Nie, Z., Zhao, J. (2019). Generating questions for knowledge basesvia incorporating diversified contexts and Answer-Aware loss. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 2431–2441). Stroudsburg, PA, USA: Association for Computational Linguistics.

Liu, D., Gong, Y., Fu, J., Yan, Y., Chen, J., Lv, J., ... Zhou, M. (2020, November). Tell me how to ask again: Question data augmentation with controllable rewriting in continuous space. *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 5798–5810). Online: Association for Computational Linguistics.

Liu, T., Fang, Q., Ding, W., Li, H., Wu, Z., Liu, Z. (2021, November). Mathematical word problem generation from commonsense knowledge graph and equations. *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 4225–4240). Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Liu, Z., Huang, K., Huang, D., Zhao, J. (2020). Semantics-reinforced networks for question generation. *ECAI 2020* (pp. 2078–2084). IOS Press.

Lopez, L.E., Cruz, D.K., Cruz, J.C.B., Cheng, C. (2021). Simplifying Paragraph-Level question generation via transformer language models. *PRICAI 2021: Trends in artificial intelligence* (pp. 323–334). Springer International Publishing.

Ma, X., Zhu, Q., Zhou, Y., Li, X. (2020, April). Improving question generation with Sentence-Level semantic matching and answer position inferring. *AAAI*, *34* (05), 8464–8471.

Mahdavi, S., An, A., Davoudi, H., Delpisheh, M., Gohari, E. (2020). Question-Worthy sentence selection for question generation. *Canadian conference on AI* (pp. 388–400). inago.com.

Majumder, B.P., Rao, S., Galley, M., McAuley, J. (2021, June). Ask what's missing and what's useful: Improving clarification question generation using global knowledge. *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 4300–4312). Online: Association for Computational Linguistics.

Matsumoto, T., Hasegawa, K., Yamakawa, Y., Mitamura, T. (2018, November). Textual entailment based question generation. *Proceedings of the workshop on intelligent interactive systems and language generation (2IS&NLG)* (pp. 15–19). Tilburg, the Netherlands: Association for Computational Linguistics.

Mishra, S.K., Goel, P., Sharma, A., Jagannatha, A., Jacobs, D., Daumé, H., III. (2020, April). Towards automatic generation of questions from long answers.
https://arxiv.org/abs/2004.05109 [cs.CL]

Murakhovs'ka, L., Wu, C.-S., Niu, T., Liu, W., Xiong, C. (2021, October). MixQG: Neural question generation with mixed answer types.
https://arxiv.org/abs/2110.08175 [cs.CL]

Nakanishi, M., Kobayashi, T., Hayashi, Y. (2019, November). Towards answer-unaware conversational question generation. *Proceedings of the 2nd workshop on machine reading for question answering* (pp. 63–71). Hong Kong, China: Association for Computational Linguistics.

Nema, P., Mohankumar, A.K., Khapra, M.M., Srinivasan, B.V., Ravindran, B. (2019). *Let's ask again: Refine network for automatic question generation.*

Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., Deng, L. (2016, January). *MS MARCO: A human generated MAchine reading COmprehension dataset.*

Pan, L., Xie, Y., Feng, Y., Chua, T.-S., Kan, M.-Y. (2020, July). Semantic graphs for generating deep questions. *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 1463–1475). Online: Association for Computational Linguistics.

Qi, P., Zhang, Y., Manning, C.D. (2020, April). Stay hungry, stay focused: Generating informative and specific questions in Information-Seeking conversations.
https://arxiv.org/abs/2004.14530 [cs.CL]

Qiu, J., & Xiong, D. (2019, November). Generating highly relevant questions. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 5983–5987). Hong Kong, China: Association for Computational Linguistics.

Qu, F., Jia, X., Wu, Y. (2021, September). Asking questions like educational experts: Automatically generating Question-Answer pairs on Real-World examination data. *Proceedings of the 2021 conference on empirical methods in natural language processing (EMNLP).*

Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P. (2016, November). SQuAD: 100,000+ questions for machine comprehension of text. *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 2383–2392).

Rao S B, P., Agnihotri, M., Babu Jayagopi, D. (2021). Improving asynchronous interview interaction with follow-up question generation. *Int. j. interact. multimed. artif. intell.*, *6*(5), 79.

Reddy, S., Chen, D., Manning, C.D. (2019). CoQA: A conversational question answering challenge. *Transactions of the Association for Computational Linguistics*, *7*, 249–266.

Reddy, S., Raghu, D., Khapra, M.M., Joshi, S. (2017, April). Generating natural language Question-Answer pairs from a knowledge graph using a RNN based question generation model. *Proceedings of the 15th conference of the European chapter of the association for computational linguistics: Volume 1, long papers* (pp. 376–385). Valencia, Spain: Association for Computational Linguistics.

Rice University (1999). *OpenStax.* https://openstax.org/. (Accessed: 2022-6-1)

Scialom, T., Piwowarski, B., Staiano, J. (2019). *Self-Attention architectures for Answer-Agnostic neural question generation.*

Scialom, T., & Staiano, J. (2020, December). Ask to learn: A study on curiosity-driven question generation. *Proceedings of the 28th international conference on computational linguistics* (pp. 2224–2235).

Barcelona, Spain (Online): International Committee on Computational Linguistics.

Sekulić, I., Aliannejadi, M., Crestani, F. (2021, July). Towards Facet-Driven generation of clarifying questions for conversational search. *Proceedings of the 2021 ACM SIGIR international conference on theory of information retrieval* (pp. 167–175). New York, NY, USA: Association for Computing Machinery.

Shen, L., Meng, F., Zhang, J., Feng, Y., Zhou, J. (2021, August). GTM: A generative triple-wise model for conversational question generation. *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 3495–3506). Online: Association for Computational Linguistics.

Shimmei, M., & Matsuda, N. (2021). Learning association between learning objectives and key concepts to generate pedagogically valuable questions. *Artificial intelligence in education* (pp. 320–324). Springer International Publishing.

Shinoda, K., Sugawara, S., Aizawa, A. (2021, August). Improving the robustness of QA models to challenge sets with variational Question-Answer pair generation. *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing: Student research workshop* (pp. 197–214). Online: Association for Computational Linguistics.

Song, L., Wang, Z., Hamza, W., Zhang, Y., Gildea, D. (2018, June). Leveraging context information for natural question generation. *Proceedings of the 2018 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 2 (short papers)* (pp. 569–574). New Orleans, Louisiana: Association for Computational Linguistics.

Srivastava, M., & Goodman, N. (2021, August). Question generation for adaptive education. *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 2: Short papers)* (pp. 692–701). Online: Association for Computational Linguistics.

Stasaski, K., Rathod, M., Tu, T., Xiao, Y., Hearst, M.A. (2021). Automatically generating Cause-and-Effect questions from passages. *Proceedings of the 16th workshop on innovative use of NLP for building educational applications* (pp. 158–170).

Steuer, T., Filighera, A., Meuser, T., Rensing, C. (2021, October). I do not understand what I cannot define: Automatic question generation with Pedagogically-Driven content selection. https://arxiv.org/abs/2110.04123 [cs.CL]

Steuer, T., Filighera, A., Rensing, C. (2020). Remember the facts? investigating Answer-Aware neural question generation for text comprehension. *Artificial intelligence in education* (pp. 512–523). Springer International Publishing.

Stewart, I., & Mihalcea, R. (2021, October). How well do you know your audience? reader-aware question generation. https://arxiv.org/abs/2110.08445 [cs.CL]

Su, M.-H., Wu, C.-H., Huang, K.-Y., Hong, Q.-B., Huang, H.-H. (2018). Follow-up question generation using pattern-based seq2seq with a small corpus for interview coaching. *INTERSPEECH* (pp. 1006–1010). isca-speech.org.

Subramanian, S., Wang, T., Yuan, X., Zhang, S., Trischler, A., Bengio, Y. (2018). *Neural models for key phrase extraction and question generation.*

Sultan, M.A., Chandel, S., Astudillo, R.F., Castelli, V. (2020). On the importance of diversity in question generation for QA. *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 5651–5656).

Sun, X., Liu, J., Lyu, Y., He, W., Ma, Y., Wang, S. (2018). Answer-focused and position-aware neural question generation. *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3930–3939).

Syed, R., Collins-Thompson, K., Bennett, P.N., Teng, M., Williams, S., Tay, D.W.W., Iqbal, S. (2020, April). Improving learning outcomes with gaze tracking and automatic question generation. *Proceedings of the web conference 2020* (pp. 1693–1703). New York, NY, USA: Association for Computing Machinery.

Trischler, A., Wang, T., Yuan, X., Harris, J., Sordoni, A., Bachman, P., Suleman, K. (2016, November). NewsQA: A machine comprehension dataset. https://arxiv.org/abs/1611.09830 [cs.CL]

Tsatsaronis, G., Balikas, G., Malakasiotis, P., Partalas, I., Zschunke, M., Alvers, M.R., ... Paliouras, G. (2015, April). An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. *BMC Bioinformatics*, *16*, 138.

Tuan, L.A., Shah, D., Barzilay, R. (2020, April). Capturing greater context for question generation. *AAAI*, *34*(05), 9065–9072.

Wang, L., Xu, Z., Lin, Z., Zheng, H., Shen, Y. (2020, December). Answer-driven deep question generation based on reinforcement learning. *Proceedings of the 28th international conference on computational linguistics* (pp. 5159–5170). Barcelona, Spain (Online): International Committee on Computational Linguistics.

Wang, S., Wei, Z., Fan, Z., Huang, Z., Sun, W., Zhang, Q., Huang, X. (2020, November). PathQG: Neural question generation from facts. *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 9066–9075). Online: Association for Computational Linguistics.

Wang, W., Feng, S., Wang, D., Zhang, Y. (2019). Answer-guided and semantic coherent question generation in open-domain conversation. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 5066–5076). aclweb.org.

Wang, Y., Liu, C., Huang, M., Nie, L. (2018, July). Learning to ask questions in open-domain conversational systems with typed decoders. *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 2193–2203). Melbourne, Australia: Association for Computational Linguistics.

Wang, Y., Liu, X., Shi, S. (2017, September). Deep neural solver for math word problems. *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 845–854). Copenhagen, Denmark: Association for Computational Linguistics.

Wang, Y., Rong, W., Zhang, J., Zhou, S., Xiong, Z. (2020, October). Multi-turn dialogue-oriented pretrained question generation model. *Complex & Intelligent Systems*, *6*(3), 493–505.

Wang, Z., Lan, A., Baraniuk, R. (2021, November). Math word problem generation with mathematical consistency and problem context constraints. *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 5986–5999). Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Wang, Z., Lan, A.S., Nie, W., Waters, A.E., Grimaldi, P.J., Baraniuk, R.G. (2018, June). QG-net: a data-driven question generation model for educational content. *Proceedings of the fifth annual ACM conference on learning at scale* (p. 7). ACM.

Wang, Z., Rao, S., Zhang, J., Qin, Z., Tian, G., Wang, J. (2020). Diversify question generation with continuous content selectors and question type modeling. *Proceedings of the 2020 conference on empirical methods in natural language processing: Findings* (pp. 2134–2143). aclweb.org.

Willis, A., Davis, G., Ruan, S., Manoharan, L., Landay, J., Brunskill, E. (2019, June). Key phrase extraction for generating educational Question-Answer pairs. *Proceedings of the sixth (2019) ACM conference on learning @ scale* (pp. 1–10). New York, NY, USA: Association for Computing Machinery.

Wu, K., Hong, Y., Zhu, M., Tang, H., Zhang, M. (2019, November). Separate answer decoding for multi-class question generation. *2019 international conference on asian language processing (IALP)* (pp. 325–330). ieeexplore.ieee.org.

Wu, X., Jiang, N., Wu, Y. (2020, July). A question type driven and copy loss enhanced frameworkfor Answer-Agnostic neural question generation. *Proceedings of the fourth workshop on neural generation and translation* (pp. 69–78). Online: Association for Computational Linguistics.

Xie, Y., Pan, L., Wang, D., Kan, M.-Y., Feng, Y. (2020, December). Exploring Question-Specific rewards for generating deep questions. *Proceedings of the 28th international conference on computational linguistics* (pp. 2534–2546). Barcelona, Spain (Online): International Committee on Computational Linguistics.

Xin, J., Hao, W., Dawei, Y., Yunfang, W. (2021, August). Enhancing question generation with commonsense knowledge. *Proceedings of the 20th chinese national conference on computational linguistics* (pp. 976–987). Huhhot, China: Chinese Information Processing Society of China.

Yang, Y., Yih, W.-T., Meek, C. (2015, September). WikiQA: A challenge dataset for Open-Domain question answering. *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 2013–2018). Lisbon, Portugal: Association for Computational Linguistics.

Yang, Z., Hu, J., Salakhutdinov, R., Cohen, W. (2017, July). Semi-Supervised QA with generative Domain-Adaptive nets. *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1040–1050). Vancouver, Canada: Association for

Computational Linguistics.

Yang, Z., Qi, P., Zhang, S., Bengio, Y., Cohen, W., Salakhutdinov, R., Manning, C.D. (2018). HotpotQA: A dataset for diverse, explainable multi-hop question answering. *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 2369–2380). Brussels, Belgium: Association for Computational Linguistics.

Yao, B., Wang, D., Wu, T., Hoang, T., Sun, B., Li, T.J.-J., . . . Xu, Y. (2022, May). It is AI's turn to ask humans a question: Question-Answer pair generation for children's story books. *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 731–744). Association for Computational Linguistics.

Yih, W.-T., Richardson, M., Meek, C., Chang, M.-W., Suh, J. (2016, August). The value of semantic parse labeling for knowledge base question answering. *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 201–206). Berlin, Germany: Association for Computational Linguistics.

Yin, X., Zhou, L., Small, K., May, J. (2021, August). Summary-Oriented question generation for informational queries. *Proceedings of the 1st workshop on document-grounded dialogue and conversational question answering (DialDoc 2021)* (pp. 81–97). Online: Association for Computational Linguistics.

Yu, X., & Jiang, A. (2021, April). Expanding, retrieving and infilling: Diversifying Cross-Domain question generation with flexible templates. *Proceedings of the 16th conference of the european chapter of the association for computational linguistics: Main volume* (pp. 3202–3212). Online: Association for Computational Linguistics.

Yuan, W., He, T., Dai, X. (2021, April). Improving neural question generation using deep linguistic representation. *Proceedings of the web conference 2021* (pp. 3489–3500). New York, NY, USA: Association for Computing Machinery.

Yuan, W., Yin, H., He, T., Chen, T., Wang, Q., Cui, L. (2022, January). Unified question generation with continual lifelong learning. https://arxiv.org/abs/2201.09696 [cs.CL]

Zamani, H., Dumais, S., Craswell, N., Bennett, P., Lueck, G. (2020, April). Generating clarifying questions for information retrieval. *Proceedings of the web conference 2020* (pp. 418–428). New York, NY, USA: Association for Computing Machinery.

Zhang, S., & Bansal, M. (2019, November). Addressing semantic drift in question generation for Semi-Supervised question answering. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 2495–2509). Hong Kong, China: Association for Computational Linguistics.

Zhang, Z. (2020). *Cue word guided question generation with BERT model fine-tuned on natural question dataset* (Unpublished doctoral dissertation). The University of Waikato.

Zhang, Z., & Zhu, K. (2021, April). Diverse and specific clarification question generation with keywords. *Proceedings of the web conference 2021* (pp. 3501–3511). New York, NY, USA: Association for Computing Machinery.

Zhong, V., Xiong, C., Socher, R. (2017, August). Seq2SQL: Generating structured queries from natural language using reinforcement learning. https://arxiv.org/abs/1709.00103 [cs.CL]

Zhou, M., Huang, M., Zhu, X. (2018, August). An interpretable reasoning network for Multi-Relation question answering. *Proceedings of the 27th international conference on computational linguistics* (pp. 2010–2022). Santa Fe, New Mexico, USA: Association for Computational Linguistics.

Zhou, Q., & Huang, D. (2019). Towards generating math word problems from equations and topics. *Proceedings of the 12th international conference on natural language generation* (pp. 494–503).

Zhou, Q., Yang, N., Wei, F., Tan, C., Bao, H., Zhou, M. (2017). Neural question generation from text: A preliminary study. *Natural language processing and chinese computing* (pp. 662–671). Springer International Publishing.

Zhou, W., Zhang, M., Wu, Y. (2019, November). Question-type driven question generation. *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 6032–6037). Hong Kong, China: Association for Computational Linguistics.

Zhu, H., Dong, L., Wei, F., Wang, W., Qin, B., Liu, T. (2019, July). Learning to ask unanswerable questions for machine reading comprehension. *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 4238–4248). Florence, Italy: Association for Computational Linguistics.