PANDA: Prompt Transfer Meets Knowledge Distillation for Efficient Model Adaptation

Anonymous Author(s)

A APPENDIX

Here, we provide more detailed results of our paper submitted to the WWW2023 conference. Appendix A.1 shows the full results of our proposed PANDA and vanilla PoT [1] on a total of 189 source-target cross-task pairs. Appendix A.2 shows the prompt transferability predicted by different metrics. Lastly, we provide the transfer performance at the first epoch in Appendix A.3. Note that we conduct experiments on five PLMs with different model sizes for all studies. We show all results on these PLMs for reference.

A.1 Full results of our PANDA and vanilla PoT approach

We report all results of our study across model sizes of PLMs. Specifically, Table 1, Table 2, Table 3, Table 4 and Table 5 list results of BERT-large, BERT-base, BERT-medium, BERT-small and BERT-tiny respectively. Our PANDA approach achieves consistent and significant performance improvements compared to the vanilla prompt transfer.

A.2 Prompt transferability predicted by our metric and other metrics

Here, we provide more heatmap results of our predicted prompt transferability across all 21 tasks on different PLMs. we provide more heatmap results of our predicted prompt transferability across all 21 tasks on different PLMs. Specifically, Figure 1, Figure 2, Figure 3 ,Figure 4 and Figure 5 show the results on BERT-large, BERT-base, BERT-medium, BERT-small and BERT-tiny respectively. More interestingly, our predicted similarities tend to drop as the scales of PLMs decrease, while the cosine similarities of prompt embeddings have the opposite tendency. One possible reason is that prompt-tuning works worse on smaller PLMs, but the paramete

A.3 Transfer performance at the first epoch

As mentioned in the main body of paper, we calculate the Spearman's rank correlation scores between ranks of predicted prompt transferability and transfer performance at the first epoch. Here, we report the detailed transfer performance across all model sizes for references. Specifically, besides the cross-task transfer performance, we also list the performance of randomly initialized prompts. Table 6, Table 7, Table 8, Table 9 and Table 10 show the results on BERT-large, BERT-base, BERT-medium, BERT-small and BERT-tiny, respectively.

REFERENCES

[1] Tu Vu, Brian Lester, Noah Constant, Rami Al-Rfou, and Daniel Cer. 2022. SPoT: Better Frozen Model Adaptation through Soft Prompt Transfer. In *ACL*.

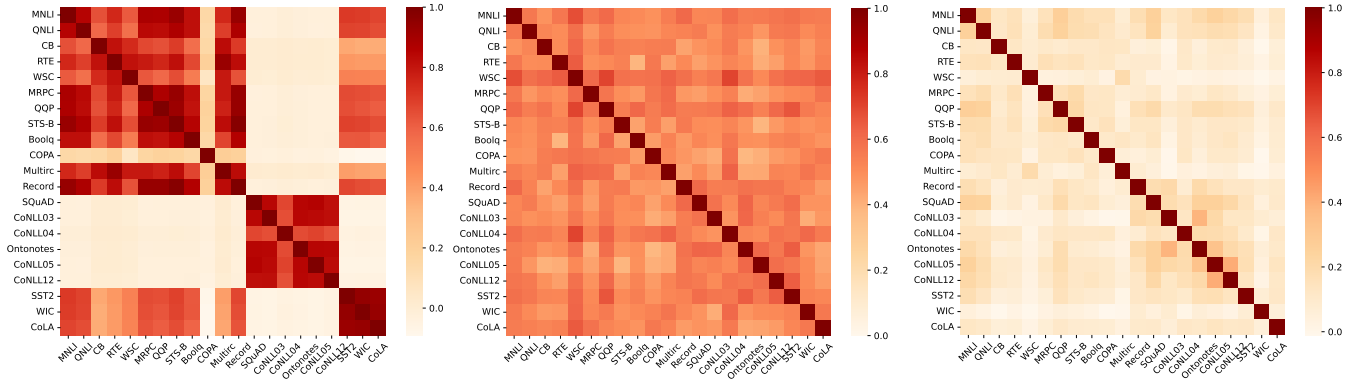

Figure 1: Left: heatmap of our results; Medium: results predicted by "ON'; Right: results predicted by "E_{avg}". BERT-large is used.

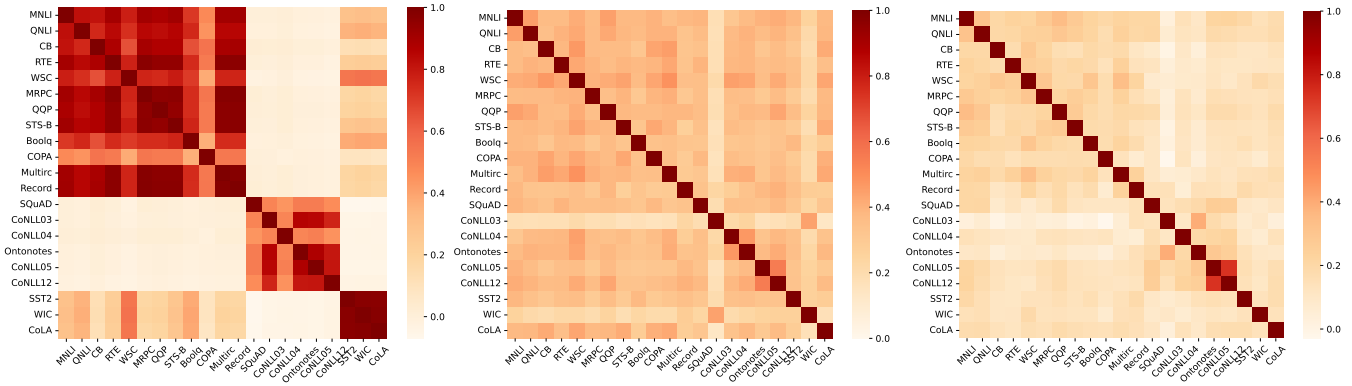

Figure 2: Left: heatmap of our results; Medium: results predicted by "ON'; Right: results predicted by "E_{avg}". BERT-base is used.

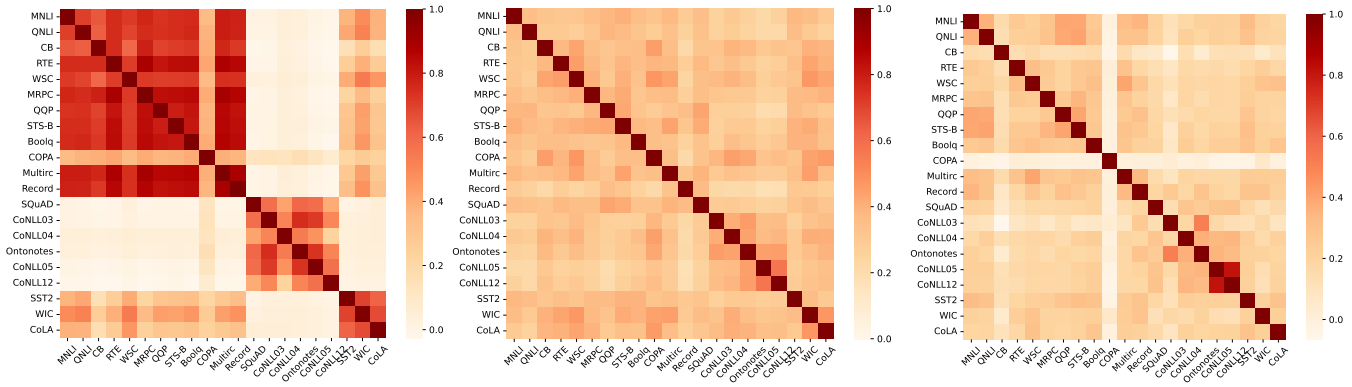

Figure 3: Left: heatmap of our results; Medium: results predicted by "ON'; Right: results predicted by "E_{avg}". BERT-medium is used.

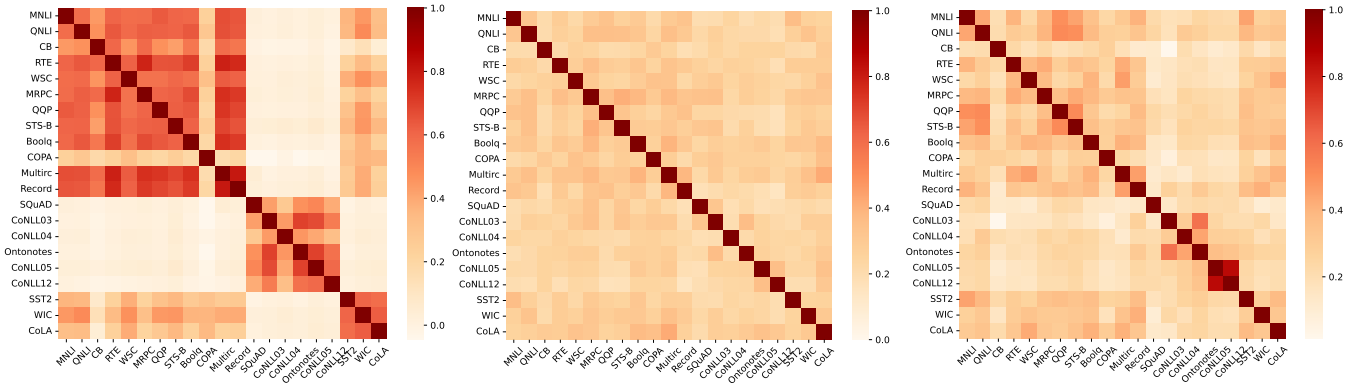

Figure 4: Left: heatmap of our results; Medium: results predicted by "ON"; Right: results predicted by "E_{avg}". BERT-small is used.

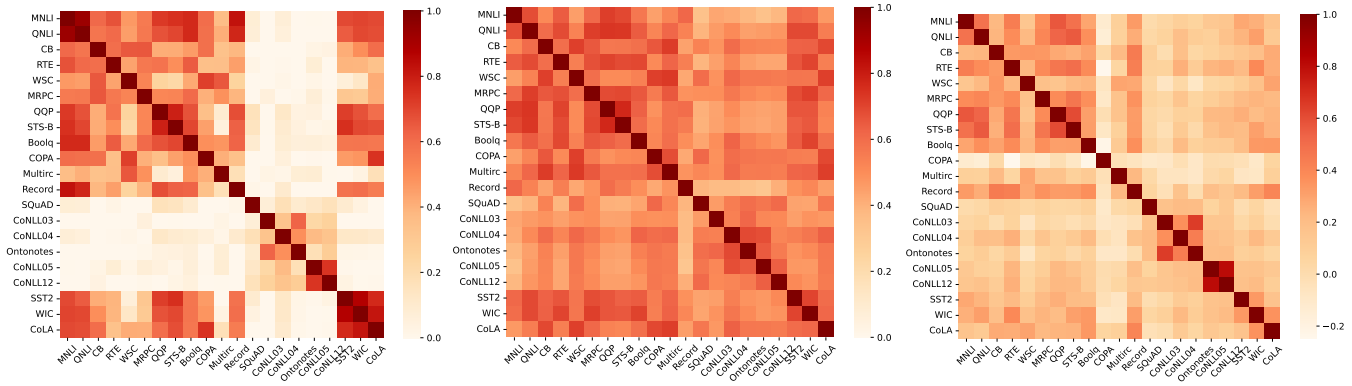

Figure 5: Left: heatmap of our results; Medium: results predicted by "ON"; Right: results predicted by "E_{avg}". BERT-tiny is used.

Table 1: Results (%) of cross-task prompt transfer on BERT-large. The red-colored row shows the results of full-tuning BERT-base model, while orange-colored ones denote prompt tuning without any prompt transfer. Notably, positive transfers are in green and "Avg." denotes the average performance of all target tasks. Numbers in the subscript indicate relative improvements compared to the vanilla prompt transfer, and we highlight the average performance that is higher than full-tuning performance in light grey.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04	Avg.
Full-tune	94.6	69.0	68.3	75.8	74.9	60.6	88.0	90.0	85.6	78.5
Baseline	87.5	76.0	64.4	76.2	66.9	63.8	86.8	90.5	85.5	77.5
(a) Transfer with Vanilla Prompt Transfer approach										
BoolQ	87.5	68.0	66.3	73.3	65.0	59.3	86.5	87.9	85.6	**75.5**
CB	87.5	74.0	64.4	73.3	66.0	64.3	88.5	90.3	85.0	**77.0**
RTE	82.1	79.0	65.4	76.2	67.9	62.1	85.3	90.2	85.0	**77.0**
WIC	85.7	74.0	65.4	76.5	66.9	54.2	84.6	87.4	84.4	**75.5**
WSC	89.3	77.0	64.4	73.6	67.9	64.5	87.3	90.7	85.5	**77.8**
COPA	87.5	76.0	64.4	76.2	67.7	62.4	86.3	90.3	84.2	**77.2**
MultiRC	87.5	75.0	65.4	74.4	67.9	60.4	88.0	90.3	84.5	**77.0**
ReCoRD	78.6	63.0	65.4	53.8	51.7	0.0	77.7	85.0	82.7	**62.0**
MNLI	96.4	71.0	67.3	80.9	66.5	58.9	88.2	91.0	83.0	**78.1**
COLA	89.3	77.0	65.4	70.4	67.1	63.8	85.5	90.0	85.0	**77.1**
SST2	92.9	74.0	64.4	71.8	66.8	60.1	87.0	89.6	84.3	**76.8**
QNLI	89.3	76.0	65.4	76.2	70.4	63.7	88.5	90.7	83.5	**78.2**
MRPC	89.3	78.0	67.3	74.4	68.3	64.1	86.8	90.3	84.3	**78.1**
STSB	85.7	79.0	66.3	77.3	66.9	62.6	87.3	90.7	84.7	**77.8**
QQP	82.1	80.0	65.4	80.1	64.6	56.3	87.0	90.5	83.7	**76.6**
SQuAD	87.5	74.0	66.3	71.8	51.7	6.0	87.3	89.3	82.5	**68.5**
CoNLL03	73.2	64.0	63.5	60.3	51.9	0.0	71.3	16.4	84.8	**53.9**
CoNLL04	82.1	75.0	66.3	76.5	67.1	64.3	87.5	87.9	85.5	**76.9**
CoNLL05	87.5	65.0	64.4	69.3	68.3	61.3	88.7	88.4	83.8	**75.2**
CoNLL12	89.3	62.0	67.3	63.2	67.4	58.7	90.4	88.5	83.6	**74.5**
Ontonotes	78.6	65.0	66.3	56.7	54.1	59.3	82.4	84.5	86.1	**70.3**
(b) Transfer with Our PANDA *approach*										
BoolQ	89.3	75.0	64.4	76.9	68.7	63.7	88.0	90.6	86.3	**78.1**$_{2.6}$
CB	87.5	76.0	67.3	71.1	69.3	64.6	88.7	90.6	85.5	**77.8**$_{0.8}$
RTE	92.9	73.0	66.6	76.2	68.7	64.2	87.3	90.7	85.8	**78.4**$_{1.4}$
WIC	91.1	78.0	66.3	77.6	66.9	64.3	87.5	90.5	86.0	**78.7**$_{3.2}$
WSC	89.3	75.0	64.4	76.2	69.5	63.8	89.7	90.7	84.7	**78.1**$_{0.3}$
COPA	91.1	76.0	64.4	76.9	69.4	62.4	86.8	90.6	84.8	**78.0**$_{0.8}$
MultiRc	91.1	76.0	63.5	77.3	68.0	62.3	89.0	90.6	85.6	**78.2**$_{1.1}$
ReCoRD	87.5	76.0	66.3	77.3	68.5	62.4	87.5	90.7	84.9	**77.9**$_{15.9}$
MNLI	92.9	77.0	67.3	78.0	68.8	66.3	88.5	90.6	85.4	**79.4**$_{1.3}$
COLA	94.6	78.0	66.3	75.8	68.5	65.3	88.0	90.6	84.9	**79.1**$_{2.1}$
SST2	92.9	77.0	68.3	76.5	70.1	64.8	88.5	90.7	86.3	**79.5**$_{2.7}$
QNLI	92.9	77.0	66.3	77.3	70.8	63.9	87.5	90.8	86.6	**79.2**$_{1.0}$
MRPC	91.1	75.0	67.3	76.5	68.5	64.2	88.0	90.7	86.3	**78.6**$_{0.5}$
STSB	92.9	76.0	67.3	75.8	69.0	64.0	88.7	90.5	85.5	**78.9**$_{1.0}$
QQP	94.6	77.0	66.3	76.2	69.4	62.6	87.0	90.7	86.0	**78.9**$_{2.2}$
SQuAD	89.3	75.0	66.3	75.5	69.3	63.1	87.3	88.9	85.7	**77.8**$_{9.3}$
CoNLL03	91.1	72.0	68.3	76.9	67.4	63.6	86.5	90.6	85.6	**78.0**$_{24.1}$
CoNLL04	87.5	73.0	68.3	75.1	65.0	64.1	90.0	90.7	86.2	**77.8**$_{0.9}$
CoNLL05	87.5	79.0	65.4	77.6	69.6	63.7	87.5	90.8	84.8	**78.4**$_{3.2}$
CoNLL12	87.5	76.0	66.3	74.4	68.5	63.7	87.5	90.8	85.0	**77.7**$_{3.3}$
Ontonotes	89.3	74.0	66.3	76.2	69.1	64.2	88.0	90.8	85.7	**78.2**$_{7.8}$

Table 2: Results (%) of cross-task prompt transfer on BERT-base.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04	Avg.
Full-tune	85.1	67.0	63.5	68.4	71.1	54.3	89.5	88.9	85.0	74.8
Baseline	80.4	68.0	64.4	73.6	65.8	57.9	87.4	88.9	84.7	74.6
(a) Transfer with Vanilla Prompt Transfer approach										
Boolq	87.5	70.0	64.4	66.4	63.6	58.8	88.0	88.7	83.6	**74.6**
CB	83.9	67.0	65.4	70.4	61.6	61.1	87.0	89.0	85.4	74.5
RTE	82.1	69.0	66.3	74.0	66.1	58.1	86.0	88.4	84.6	75.0
WIC	78.6	65.0	64.4	65.3	68.5	57.3	85.8	88.0	82.8	**72.9**
WSC	82.1	64.0	66.3	69.0	64.7	60.9	87.7	88.7	83.3	74.1
COPA	87.5	65.0	65.4	72.2	63.0	59.3	87.3	88.8	83.4	74.7
Multirc	78.6	65.0	64.4	65.3	68.5	57.3	85.8	88.0	82.8	**72.9**
Record	82.1	64.0	66.3	69.0	64.7	60.9	87.7	88.7	83.3	**74.1**
MNLI	85.7	73.0	64.4	76.2	66.1	56.5	87.0	89.2	82.4	75.6
COLA	85.7	69.0	65.4	66.4	64.3	62.3	85.5	88.6	84.1	**74.6**
SST2	82.1	71.0	64.4	67.9	65.8	55.0	86.3	88.4	83.1	**73.8**
QNLI	89.3	68.0	65.4	71.5	67.4	53.6	88.2	89.2	83.4	75.1
MRPC	85.7	71.0	64.4	71.5	65.4	58.6	88.0	89.0	83.8	75.3
STSB	85.7	71.0	65.4	73.6	64.6	57.5	87.7	89.1	82.9	75.3
QQP	80.4	72.0	64.4	73.3	63.9	54.7	87.7	89.0	82.4	**74.2**
SQuAD	83.9	63.0	64.4	72.2	64.7	53.6	89.2	88.1	84.8	**73.8**
CoNLL03	73.2	66.0	64.4	57.0	60.0	51.6	71.8	84.2	84.3	**68.1**
CoNLL04	85.7	67.0	66.3	69.0	66.1	56.8	86.8	88.9	85.6	74.7
CoNLL05	87.5	64.0	63.5	67.9	61.6	53.5	86.5	88.0	82.7	**72.8**
CoNLL12	89.3	66.0	63.5	64.3	66.1	53.9	88.5	88.5	84.2	**73.8**
Ontonotes	76.8	66.0	63.5	62.1	63.6	52.9	81.6	87.7	87.0	**71.2**
(b) Transfer with Our PANDA approach										
Boolq	82.1	72.0	66.3	72.6	65.0	58.5	88.5	88.8	84.7	$75.4_{0.8}$
CB	78.6	72.0	66.3	74.0	64.9	59.6	87.0	88.8	84.5	$75.1_{0.5}$
RTE	80.4	71.0	66.3	71.5	66.6	61.6	87.7	88.6	84.7	$75.4_{0.4}$
WIC	82.1	69.0	66.3	73.6	65.7	58.6	87.3	88.7	85.2	$75.2_{2.3}$
WSC	83.9	71.0	66.3	71.5	67.4	59.4	88.2	88.7	84.6	$75.7_{1.6}$
COPA	85.7	67.0	66.3	72.2	66.9	60.2	87.5	88.7	84.2	$75.4_{0.8}$
Multirc	82.1	69.0	66.3	72.6	65.8	59.3	87.7	88.7	84.9	$75.2_{2.3}$
Record	82.1	72.0	66.3	71.5	65.5	58.8	87.3	88.7	84.7	$75.2_{1.1}$
MNLI	82.1	73.0	66.3	72.9	66.3	60.3	88.2	89.0	84.8	$75.9_{0.3}$
COLA	85.7	67.0	66.3	72.2	66.9	60.2	87.5	88.7	84.2	$75.4_{0.8}$
SST2	87.5	71.0	66.3	72.2	67.7	61.1	87.7	88.9	84.2	$76.3_{2.5}$
QNLI	83.9	70.0	66.3	72.2	67.7	58.9	87.5	89.0	85.1	$75.6_{0.5}$
MRPC	80.4	72.0	66.3	73.3	66.5	59.3	87.3	88.7	84.7	$75.4_{0.1}$
STSB	80.4	73.0	66.3	74.4	66.1	57.5	87.5	89.0	83.9	$75.3_{0.1}$
QQP	87.5	71.0	65.4	71.8	66.3	59.8	87.5	88.9	84.3	$75.8_{1.6}$
SQuAD	83.9	69.0	65.4	72.9	66.0	58.8	86.3	88.8	85.2	$75.1_{1.4}$
CoNLL03	83.9	70.0	67.3	72.9	66.6	59.1	87.3	88.9	85.2	$75.7_{7.6}$
CoNLL04	83.9	70.0	65.4	70.0	66.1	59.1	87.7	88.6	83.9	$75.0_{0.3}$
CoNLL05	83.9	69.0	67.3	74.0	66.0	59.6	89.2	88.9	84.9	$75.9_{3.1}$
CoNLL12	82.1	71.0	66.3	72.6	65.4	59.4	87.3	88.9	84.1	$75.2_{1.4}$
Ontonotes	82.1	71.0	65.4	73.3	66.6	58.3	87.5	89.0	84.5	$75.3_{4.1}$

Table 3: Results (%) of cross-task prompt transfer on BERT-medium.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04	Avg.
Full-tune	**81.5**	**66.0**	**65.6**	**63.0**	**66.5**	**42.5**	**86.6**	**86.4**	**81.2**	**71.0**
Baseline	**80.4**	**65.0**	**63.5**	**67.1**	**64.1**	**43.0**	**84.1**	**87.8**	**79.7**	**70.5**
(a) Transfer with Vanilla Prompt Transfer approach										
Boolq	**83.9**	65.0	**65.4**	64.3	63.6	41.9	83.8	87.4	79.7	**70.6**
CB	**82.1**	60.0	**64.4**	65.7	63.9	**48.3**	82.6	87.3	78.8	**70.3**
RTE	76.8	63.0	**63.5**	**70.0**	61.8	**45.6**	83.1	87.3	79.0	**70.0**
WIC	76.8	65.0	**64.4**	62.1	63.6	42.7	83.6	**87.9**	79.0	**69.5**
WSC	**82.1**	**66.0**	63.5	**69.3**	62.4	**43.3**	**87.0**	**88.0**	78.9	**71.2**
COPA	**82.1**	**66.0**	**65.4**	**69.0**	**64.3**	**45.5**	**85.3**	**87.9**	79.9	**71.7**
Multirc	**82.1**	65.0	**65.4**	65.3	61.8	42.8	82.6	87.5	79.2	**70.2**
Record	73.2	60.0	**64.4**	62.8	60.0	38.3	80.4	87.3	78.6	**67.2**
MNLI	**87.5**	61.0	**66.3**	**71.5**	60.5	41.2	**86.5**	**88.8**	78.3	**71.3**
COLA	75.0	63.0	**65.4**	62.8	62.2	**47.4**	**86.8**	86.9	79.8	**69.9**
SST2	78.6	64.0	64.4	**68.2**	59.1	42.6	**86.3**	87.8	78.8	**70.0**
QNLI	80.4	60.0	**65.4**	64.6	61.9	38.2	**85.3**	**87.9**	79.0	**69.2**
MRPC	**82.1**	63.0	64.4	65.3	63.9	41.6	**84.8**	**87.9**	78.6	**70.2**
STSB	**85.7**	**66.0**	**65.4**	67.1	63.5	42.7	83.8	87.7	79.7	**71.3**
QQP	**89.3**	61.0	**66.3**	**69.0**	61.0	37.7	**85.0**	**88.1**	78.7	**70.7**
SQuAD	71.4	64.0	**66.3**	**67.5**	63.0	40.3	**85.0**	**88.2**	78.0	**69.3**
CoNLL03	80.4	58.0	63.5	62.8	62.1	37.3	78.2	86.8	79.0	**67.6**
CoNLL04	78.6	62.0	63.5	63.5	**65.2**	**43.9**	83.8	87.2	**80.0**	**69.7**
CoNLL05	76.8	62.0	**64.4**	64.3	62.1	43.0	86.8	87.6	78.3	**69.5**
CoNLL12	76.8	**66.0**	**65.4**	64.3	61.4	**43.9**	**86.0**	87.5	77.6	**69.9**
Ontonotes	75.0	64.0	**64.4**	61.0	61.1	40.9	80.4	87.2	**81.3**	**68.4**
(b) Transfer with Our PANDA approach										
Boolq	**85.7**	63.0	**65.4**	**69.7**	**65.4**	**44.7**	**86.0**	87.8	79.2	**71.9**$_{1.3}$
CB	**85.7**	62.0	**65.4**	**69.0**	**65.8**	**45.0**	**86.5**	**87.9**	**79.8**	**71.9**$_{1.6}$
RTE	**83.9**	63.0	**64.4**	**68.6**	63.8	**43.7**	**86.3**	**87.9**	79.2	**71.2**$_{1.2}$
WIC	**82.1**	65.0	**64.4**	**68.2**	**65.4**	**45.0**	**87.3**	**87.9**	79.6	**71.7**$_{2.2}$
WSC	**83.9**	65.0	**65.4**	**71.1**	**66.6**	**44.5**	**86.8**	**87.8**	79.4	**72.3**$_{1.1}$
COPA	**83.9**	**67.0**	**66.3**	**69.3**	**64.9**	**45.6**	**86.3**	**87.9**	79.4	**72.3**$_{0.6}$
Multirc	**83.9**	64.0	**64.4**	**69.3**	**64.4**	**43.9**	**86.3**	87.8	79.2	**71.5**$_{1.3}$
Record	**85.7**	61.0	**64.4**	**69.0**	**64.9**	**43.7**	**86.0**	87.8	79.6	**71.3**$_{4.1}$
MNLI	**83.9**	**66.0**	**64.4**	**69.3**	**65.0**	**43.2**	**86.5**	**87.9**	79.1	**71.7**$_{0.4}$
COLA	**83.9**	**66.0**	**64.4**	**67.9**	**65.0**	**46.1**	**86.8**	87.8	79.3	**71.9**$_{2.0}$
SST2	**83.9**	**66.0**	**64.4**	**69.3**	**64.6**	**45.3**	**86.5**	87.8	79.2	**71.9**$_{1.9}$
QNLI	**87.5**	**68.0**	**65.4**	**69.7**	**64.6**	**44.0**	**86.3**	87.8	79.3	**72.5**$_{3.3}$
MRPC	**83.9**	64.0	**64.4**	**69.3**	**65.2**	**43.7**	**87.3**	**87.9**	79.6	**71.7**$_{1.5}$
STSB	**85.7**	**67.0**	**65.4**	**69.0**	**65.8**	**44.3**	**86.8**	87.8	79.7	**72.4**$_{1.1}$
QQP	**83.9**	64.0	**64.4**	**69.0**	**65.4**	**43.6**	**86.8**	87.8	79.2	**71.6**$_{0.9}$
SQuAD	**85.7**	63.0	**65.4**	**69.0**	64.1	**43.8**	**86.0**	87.8	79.2	**71.6**$_{2.3}$
CoNLL03	**82.1**	63.0	**64.4**	**69.0**	**64.3**	**43.8**	**86.3**	87.7	**79.9**	**71.2**$_{3.6}$
CoNLL04	78.6	63.0	**65.4**	**69.7**	**66.3**	**44.7**	**86.5**	87.8	79.1	**71.2**$_{1.5}$
CoNLL05	**85.7**	61.0	**65.4**	**70.4**	**64.4**	**44.0**	**87.7**	87.8	79.5	**71.8**$_{2.3}$
CoNLL12	**83.9**	65.0	**65.4**	**69.3**	**64.9**	**43.6**	**86.3**	**87.9**	79.4	**71.7**$_{1.9}$
Ontonotes	**83.9**	63.0	**65.4**	**69.7**	**65.5**	**43.8**	**85.8**	**88.0**	79.3	**71.6**$_{3.2}$

Table 4: Results (%) of cross-task prompt transfer on BERT-small.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04	Avg.
Full-tune	78.6	66.0	65.4	65.3	62.9	38.4	84.5	84.2	72.8	68.7
Baseline	76.8	64.0	63.5	65.7	63.3	30.7	81.6	86.5	75.6	67.5
(a) Transfer with Vanilla Prompt Transfer approach										
Boolq	83.9	61.0	64.4	62.1	61.8	30.3	79.2	86.6	75.2	**67.2**
CB	80.4	62.0	65.4	66.8	63.2	31.4	83.3	86.0	75.5	68.2
RTE	76.8	64.0	66.3	63.5	63.6	28.5	78.7	86.1	74.7	**66.9**
WIC	76.8	64.0	63.5	64.6	62.2	31.3	78.9	86.6	73.7	**66.8**
WSC	78.6	65.0	66.3	63.9	63.3	30.2	81.1	86.4	75.5	67.8
COPA	82.1	64.0	64.4	67.5	62.1	30.2	82.6	86.6	75.0	68.3
Multirc	73.2	61.0	64.4	66.8	61.6	26.8	79.9	86.4	75.5	**66.2**
Record	71.4	61.0	65.4	61.0	58.6	7.4	75.7	85.9	72.6	**62.1**
MNLI	83.9	64.0	65.4	71.5	60.8	25.0	82.1	86.6	73.4	68.1
COLA	73.2	67.0	63.5	63.5	59.9	35.7	81.9	86.5	74.8	**67.3**
SST2	78.6	59.0	65.4	63.5	61.3	35.0	80.9	86.0	74.5	**67.1**
QNLI	80.4	65.0	64.4	65.7	62.7	26.7	80.4	86.5	75.4	**67.5**
MRPC	82.1	58.0	65.4	63.5	61.1	30.7	82.1	86.3	74.6	**67.1**
STSB	83.9	64.0	65.4	66.4	63.2	29.9	80.1	86.1	75.6	68.3
QQP	83.9	58.0	66.3	65.3	61.8	25.6	79.9	86.9	74.0	**66.9**
SQuAD	80.4	62.0	65.4	68.2	62.2	24.9	81.1	86.5	74.3	**67.2**
CoNLL3	76.8	65.0	63.5	61.0	62.7	27.4	75.5	86.3	75.2	**65.9**
CoNLL4	82.1	61.0	63.5	60.6	60.8	31.1	79.7	86.1	76.2	**66.8**
CoNLL5	78.6	64.0	64.4	63.2	58.9	21.8	82.1	85.8	74.9	**66.0**
CoNLL12	73.2	63.0	68.3	61.4	61.0	28.2	82.8	85.8	74.9	**66.5**
Ontonotes	73.2	61.0	64.4	59.2	60.3	28.0	75.0	85.9	78.4	**65.0**
(b) Transfer with Our PANDA approach										
Boolq	78.6	65.0	65.4	63.2	61.8	35.0	83.8	86.5	75.6	$68.3_{1.2}$
CB	78.6	66.0	65.4	64.6	62.2	35.5	83.8	86.4	75.4	$68.7_{0.4}$
RTE	78.6	67.0	65.4	66.1	63.3	35.0	83.3	86.6	75.1	$68.9_{2.0}$
WIC	76.8	68.0	65.4	63.5	61.9	35.1	82.8	86.6	75.7	$68.4_{1.6}$
WSC	80.4	66.0	65.4	63.2	61.4	35.2	83.3	86.5	75.5	$68.5_{0.7}$
COPA	78.6	65.0	65.4	64.3	61.6	35.3	83.3	86.4	75.5	$68.4_{0.1}$
Multirc	78.6	66.0	65.4	64.6	62.2	35.2	84.6	86.5	75.2	$68.7_{2.5}$
Record	78.6	66.0	65.4	63.9	62.7	34.7	84.8	86.8	75.4	$68.7_{6.6}$
MNLI	80.4	66.0	65.4	66.1	63.5	35.5	83.8	86.5	75.7	$69.2_{1.1}$
COLA	80.4	65.0	65.4	64.3	62.9	34.5	83.3	86.7	75.7	$68.7_{1.4}$
SST2	78.6	66.0	65.4	66.4	63.2	36.6	83.8	86.7	75.4	$69.1_{2.0}$
QNLI	82.1	66.0	66.3	66.4	62.4	35.0	83.6	86.6	75.5	$69.3_{1.9}$
MRPC	78.6	67.0	65.4	63.5	62.4	34.5	83.6	86.6	75.9	$68.6_{1.5}$
STSB	78.6	67.0	65.4	63.5	62.4	34.5	83.6	86.6	75.9	$68.6_{0.3}$
QQP	78.6	67.0	65.4	65.3	63.5	34.1	83.1	86.7	75.8	$68.8_{2.0}$
SQuAD	78.6	66.0	65.4	63.5	61.0	35.0	84.1	86.5	75.9	$68.4_{1.2}$
CoNLL03	78.6	66.0	65.4	66.1	63.6	35.7	84.3	86.7	76.0	$69.2_{3.2}$
CoNLL04	78.6	67.0	65.4	62.8	62.7	34.9	83.8	86.5	75.7	$68.6_{1.8}$
CoNLL05	82.1	70.0	65.4	65.3	62.7	35.2	83.6	86.6	75.8	$69.6_{3.7}$
CoNLL12	78.6	65.0	65.4	63.9	62.5	35.9	84.3	86.6	75.5	$68.6_{2.1}$
Ontonotes	78.6	66.0	65.4	63.2	62.4	35.6	83.6	86.7	76.1	$68.6_{3.6}$

Table 5: Results (%) of cross-task prompt transfer on BERT-tiny.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04	Avg.
Full-tune	73.2	61.0	64.8	64.2	60.9	5.0	77.1	78.3	62.6	60.8
Baseline	71.4	62.0	63.4	65.0	57.7	0.0	72.8	81.8	57.6	59.1
(a) Transfer with Vanilla Prompt Transfer approach										
Boolq	73.2	62.0	66.3	60.3	58.0	7.4	72.3	82.1	57.7	59.9
CB	71.4	67.0	67.3	63.5	59.9	1.1	73.5	82.3	57.0	60.3
RTE	73.2	55.0	67.3	63.9	58.6	6.6	74.8	82.6	57.8	60.0
WIC	67.9	60.0	64.4	61.7	60.0	6.7	74.0	82.1	57.2	59.3
WSC	71.4	61.0	65.4	62.8	59.1	0.1	72.5	81.8	57.0	**59.0**
COPA	71.4	64.0	66.3	65.3	57.4	6.9	73.0	82.1	56.7	60.3
Multirc	69.6	62.0	66.3	65.0	59.1	0.0	72.5	81.5	57.3	59.3
Record	50.0	61.0	63.5	52.7	50.0	0.0	68.4	10.0	52.9	**45.4**
MNLI	71.4	56.0	64.4	66.1	53.0	6.6	77.7	83.0	56.2	59.4
COLA	71.4	62.0	66.3	64.3	58.2	4.1	72.1	82.1	56.3	59.6
SST2	69.6	59.0	65.4	58.8	58.2	2.1	74.0	80.9	56.9	**58.3**
QNLI	69.6	58.0	65.4	65.0	59.9	0.0	75.5	82.8	56.2	59.2
MRPC	71.4	66.0	64.4	62.8	57.5	0.0	73.3	82.0	57.3	59.4
STSB	73.2	59.0	63.5	64.6	56.3	5.2	74.8	82.2	57.6	59.6
QQP	71.4	59.0	67.3	63.2	58.2	6.6	75.5	82.4	55.9	59.9
SQuAD	75.0	60.0	63.5	59.9	55.5	2.8	74.0	80.9	56.8	**58.7**
CoNLL03	75.0	61.0	63.5	53.1	58.0	0.0	72.8	80.1	58.9	**58.0**
CoNLL04	73.2	56.0	63.5	60.3	58.8	4.6	74.0	82.2	58.8	**59.0**
CoNLL05	71.4	59.0	63.5	61.7	54.4	0.0	73.5	81.5	55.4	**57.8**
CoNLL12	73.2	59.0	63.5	60.3	59.1	0.0	74.3	81.6	54.3	**58.4**
Ontonotes	71.4	61.0	63.5	58.1	56.9	0.5	71.8	80.6	60.3	**58.2**
(b) Transfer with Our PANDA approach										
Boolq	75.0	62.0	66.3	65.0	58.5	5.9	73.0	82.0	55.8	$60.4_{0.5}$
CB	75.0	65.0	67.3	65.7	59.7	8.0	73.5	82.0	56.6	$61.4_{1.1}$
RTE	73.2	62.0	67.3	65.0	58.3	8.0	73.0	82.0	55.9	$60.5_{0.5}$
WIC	73.2	60.0	68.3	65.0	58.8	3.6	73.0	82.0	55.6	$59.9_{0.6}$
WSC	73.2	62.0	66.3	65.0	58.9	7.4	73.0	82.0	56.2	$60.4_{1.4}$
COPA	75.0	63.0	66.3	65.0	58.2	6.2	73.0	82.0	55.8	$60.5_{0.2}$
Multirc	73.2	62.0	66.3	65.3	59.6	4.6	73.5	82.0	55.8	$60.3_{1.0}$
Record	75.0	62.0	67.3	64.6	58.8	6.0	73.0	82.0	55.8	$60.5_{15.1}$
MNLI	73.2	62.0	67.3	65.0	58.5	4.8	73.0	82.1	55.9	$60.2_{0.8}$
COLA	75.0	63.0	67.3	64.6	59.4	8.0	73.0	82.0	56.1	$60.9_{1.3}$
SST2	73.2	62.0	66.3	65.0	59.1	3.1	73.3	82.0	55.8	$60.0_{1.7}$
QNLI	73.2	62.0	66.3	65.0	58.5	8.0	73.0	82.0	56.2	$60.5_{1.3}$
MRPC	73.2	62.0	66.3	65.0	58.3	5.2	73.0	82.0	55.9	$60.1_{0.7}$
STSB	73.2	61.0	66.3	65.0	59.4	3.5	73.0	82.0	55.6	$59.9_{0.3}$
QQP	73.2	62.0	67.3	65.7	59.2	5.3	73.0	82.0	55.9	$60.4_{0.5}$
SQuAD	75.0	61.0	67.3	64.6	59.6	2.1	83.0	82.0	56.0	$61.2_{2.5}$
CoNLL03	75.0	62.0	67.3	64.6	60.2	6.6	73.0	82.0	55.5	$60.7_{2.6}$
CoNLL04	75.0	60.0	66.3	65.0	58.9	5.9	73.0	82.0	57.8	$60.4_{1.4}$
CoNLL05	75.0	61.0	66.3	65.0	59.1	5.2	73.0	82.0	55.4	$60.2_{2.4}$
CoNLL12	73.2	61.0	66.3	65.0	58.5	4.5	73.0	82.0	55.9	$59.9_{1.6}$
Ontonotes	73.2	63.0	66.3	64.6	58.8	4.1	73.0	82.0	55.4	$60.0_{1.8}$

Table 6: Results (%) of cross-task prompt transfer at the first epoch on BERT-large. The orange-colored row shows the results of randomly initialized prompts. Notably, positive transfers are in green.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04
Random	42.9	63.0	61.5	50.5	50.0	45.0	52.5	78.5	67.9
Boolq	75.0	44.0	48.1	58.1	50.5	2.6	77.0	83.8	56.3
CB	80.4	62.0	59.6	53.1	49.8	48.1	61.0	84.6	66.5
RTE	60.7	70.0	62.5	74.0	50.0	11.2	77.0	86.3	63.8
WIC	67.9	55.0	59.6	55.6	66.0	29.7	69.9	82.6	57.0
WSC	50.0	53.0	63.5	52.7	50.0	27.6	31.6	83.3	68.9
COPA	50.0	75.0	62.5	54.5	54.9	29.0	71.8	82.0	68.8
Multirc	69.6	48.0	54.8	50.2	50.0	7.0	68.6	82.9	64.1
Record	41.1	55.0	55.8	47.3	51.7	0.0	68.4	8.1	0.0
MNLI	85.7	71.0	49.0	79.8	56.0	0.7	80.4	88.4	57.1
COLA	67.9	59.0	62.5	49.5	60.5	54.8	72.8	83.8	69.3
SST2	73.2	72.0	58.7	55.6	51.7	50.9	71.1	84.3	67.7
QNLI	71.4	66.0	63.5	59.6	66.3	44.4	82.6	88.8	65.6
MRPC	58.9	72.0	33.7	56.7	50.0	0.0	84.8	88.5	64.7
STSB	71.4	75.0	59.6	67.1	59.7	44.5	83.1	90.4	64.0
QQP	69.6	76.0	63.5	63.9	51.9	2.5	79.2	89.4	56.4
SQuAD	67.9	68.0	63.5	47.3	50.0	5.9	80.4	31.8	54.5
CoNLL03	55.4	48.0	63.5	50.2	50.5	0.0	69.1	10.6	70.8
CoNLL04	48.2	64.0	59.6	48.0	59.2	41.3	68.6	74.3	80.6
CoNLL05	57.1	49.0	53.8	59.9	50.0	0.0	71.8	77.6	60.9
CoNLL12	62.5	54.0	52.9	47.3	63.2	49.2	74.5	83.6	67.9
Ontonotes	67.9	56.0	61.5	47.7	52.4	27.2	69.6	20.4	74.3

Table 7: Results (%) of cross-task prompt transfer at the first epoch on BERT-base.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04
Random	41.1	55.0	53.8	55.2	53.1	37.6	31,6	83.5	69.6
Boolq	75.0	59.0	63.5	59.2	49.8	28.9	74.5	84.4	68.5
CB	73.2	59.0	62.5	64.3	51.9	43.0	71.1	83.8	69.9
RTE	75.0	43.0	61.5	70.0	54.2	31.8	76.0	84.7	70.6
WIC	62.5	45.0	63.5	51.3	63.8	27.5	71.6	82.3	63.8
WSC	69.6	61.0	65.4	56.0	49.8	36.1	70.3	83.0	70.1
COPA	55.4	62.0	65.4	50.9	51.1	18.0	76.0	84.4	69.6
Multirc	73.2	57.0	63.5	56.7	51.9	36.5	74.5	84.6	70.8
Record	64.3	49.0	63.5	48.4	49.7	16.7	70.6	82.5	56.0
MNLI	76.8	54.0	60.6	70.8	50.6	18.4	80.1	87.9	65.6
COLA	60.7	58.0	65.4	55.6	57.7	58.4	74.3	81.6	71.2
SST2	66.1	59.0	61.5	52.3	56.7	37.9	75.0	84.4	69.4
QNLI	67.9	68.0	63.5	50.9	54.4	37.8	79.9	86.5	71.7
MRPC	64.3	63.0	55.8	61.0	55.0	30.3	85.0	85.4	70.0
STSB	75.0	70.0	63.5	49.5	54.4	38.9	83.3	88.3	70.2
QQP	64.3	65.0	63.5	65.3	53.1	32.8	81.1	88.1	68.8
SQuAD	62.5	50.0	63.5	52.7	60.3	31.5	79.2	84.7	70.6
CoNLL03	62.5	57.0	63.5	47.3	50.0	7.4	69.1	5.3	72.0
CoNLL04	41.1	51.0	65.4	47.3	54.4	45.3	69.6	84.0	81.7
CoNLL05	51.8	51.0	63.5	41.9	50.0	1.3	74.3	82.2	65.6
CoNLL12	58.9	54.0	63.5	57.8	50.8	23.2	69.4	80.4	68.6
Ontonotes	41.1	55.0	63.5	47.3	50.0	0.0	71.8	76.7	74.1

Table 8: Results (%) of cross-task prompt transfer at the first epoch on BERT-medium.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04
Random	**71.0**	**56.0**	**59.6**	**53.8**	**54.0**	**21.5**	**71.8**	**80.7**	**65.8**
Boolq	71.4	58.0	63.5	52.7	56.6	15.5	73.5	82.8	66.2
CB	76.8	54.0	60.6	55.2	53.6	13.9	70.3	81.4	67.2
RTE	67.9	55.0	61.5	57.0	52.7	14.6	72.8	83.6	66.6
WIC	67.9	52.0	63.5	53.8	63.0	27.5	71.1	81.6	65.5
WSC	67.9	55.0	60.6	56.7	52.5	19.2	71.1	81.1	67.9
COPA	67.9	61.0	58.7	52.7	55.0	20.6	71.3	81.1	66.1
Multirc	71.4	65.0	63.5	58.8	52.0	18.1	73.8	82.6	68.3
Record	53.6	57.0	63.5	48.4	53.8	0.0	70.8	81.8	60.4
MNLI	67.9	52.0	63.5	65.0	55.5	0.0	78.7	86.4	61.2
COLA	67.9	63.0	63.5	56.3	52.2	43.2	72.1	82.9	67.4
SST2	67.9	54.0	63.5	55.6	54.2	14.8	74.5	82.4	64.9
QNLI	67.9	52.0	63.5	54.5	54.7	10.5	78.2	84.8	66.6
MRPC	73.2	52.0	62.5	54.2	51.9	3.1	82.8	84.2	64.8
STSB	66.1	54.0	63.5	62.8	53.4	0.0	77.7	87.4	66.6
QQP	71.4	61.0	60.6	58.5	55.0	2.9	76.2	84.9	62.2
SQuAD	69.6	53.0	64.4	64.3	49.8	1.8	77.5	86.3	62.4
CoNLL03	67.9	50.0	63.5	56.3	51.4	1.8	69.1	73.4	73.6
CoNLL04	71.4	55.0	63.5	50.5	53.1	26.8	71.1	82.4	75.4
CoNLL05	67.9	45.0	63.5	51.6	53.4	3.0	67.4	79.7	59.8
CoNLL12	71.4	51.0	63.5	53.4	50.2	6.4	70.1	81.9	63.0
Ontonotes	69.6	55.0	63.5	56.0	55.6	2.1	68.9	80.0	72.6

Table 9: Results (%) of cross-task prompt transfer at the first epoch on BERT-small.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04
Random	**69.6**	**54.0**	**53.8**	**52.3**	**54.5**	**10.3**	**70.6**	**78.6**	**64.2**
Boolq	69.6	43.0	63.5	55.2	52.7	6.7	70.6	81.2	62.5
CB	71.4	52.0	52.9	50.9	57.2	8.2	72.3	80.3	63.8
RTE	71.4	61.0	63.5	56.3	51.4	0.3	74.0	79.4	65.1
WIC	69.6	59.0	63.5	54.9	61.3	3.9	70.6	80.5	60.8
WSC	69.6	56.0	53.8	55.6	53.1	6.3	70.6	78.5	64.9
COPA	71.4	53.0	59.6	58.1	52.5	1.8	71.1	80.9	64.8
Multirc	67.9	50.0	63.5	49.5	49.7	0.0	68.4	77.2	63.6
Record	58.9	54.0	63.5	47.3	49.2	0.0	68.4	22.2	40.6
MNLI	73.2	63.0	63.5	67.1	49.5	8.0	78.2	84.1	58.5
COLA	67.9	67.0	61.5	52.7	51.3	24.6	68.9	80.1	66.4
SST2	67.9	56.0	63.5	51.3	56.3	3.1	70.3	80.2	59.9
QNLI	71.4	50.0	63.5	56.0	51.9	4.6	73.8	82.7	61.1
MRPC	69.6	53.0	61.5	51.6	49.2	0.4	81.4	81.9	61.0
STSB	73.2	64.0	63.5	57.8	53.4	0.0	74.8	85.4	63.0
QQP	69.6	57.0	63.5	58.1	50.5	0.0	72.1	84.8	58.5
SQuAD	66.1	53.0	63.5	57.4	50.8	0.0	72.3	83.7	56.6
CoNLL03	69.6	58.0	56.7	57.4	55.8	0.0	69.6	75.9	68.8
CoNLL04	69.6	60.0	53.8	54.9	54.9	0.9	68.9	81.2	72.1
CoNLL05	66.1	55.0	63.5	51.6	51.1	0.4	69.4	76.5	58.9
CoNLL12	67.9	37.0	60.6	50.2	50.3	4.8	69.4	74.9	60.1
Ontonotes	67.9	48.0	59.6	59.2	49.7	0.0	70.3	78.6	69.0

Table 10: Results (%) of cross-task prompt transfer at the first epoch on BERT-tiny.

Tasks	CB	COPA	WSC	RTE	WIC	COLA	MRPC	STSB	CoNLL04
Random	**64.3**	**50.0**	**61.5**	**50.4**	**51.2**	**0.0**	**69.1**	**72.9**	**40.5**
Boolq	64.3	52.0	63.5	53.1	49.7	0.0	68.6	72.6	41.0
CB	64.3	48.0	65.4	54.9	51.9	1.1	68.9	73.6	40.4
RTE	53.6	50.0	63.5	52.7	52.0	0.0	68.6	74.6	42.6
WIC	53.6	54.0	61.5	56.3	55.8	4.4	69.6	71.8	39.0
WSC	64.3	54.0	65.4	58.5	51.6	0.1	69.1	73.5	40.5
COPA	64.3	51.0	64.4	57.4	51.7	0.0	68.9	72.8	40.8
Multirc	62.5	50.0	65.4	57.0	51.4	0.0	69.1	72.3	40.6
Record	41.1	51.0	63.5	47.3	50.0	0.0	68.4	0.4	0.0
MNLI	41.1	50.0	61.5	52.7	50.5	0.0	70.3	76.2	31.8
COLA	66.1	46.0	65.4	55.6	50.5	4.1	68.9	72.6	40.8
SST2	58.9	52.0	62.5	51.3	49.7	0.0	69.1	66.3	26.8
QNLI	57.1	49.0	63.5	53.1	51.4	0.0	70.6	78.6	37.6
MRPC	62.5	46.0	63.5	50.5	51.9	0.0	72.5	76.4	34.7
STSB	57.1	50.0	63.5	50.2	54.4	0.0	69.9	80.6	40.3
QQP	60.7	52.0	55.8	57.8	50.2	0.0	69.6	78.5	40.2
SQuAD	41.1	52.0	63.5	53.8	50.0	0.0	68.4	63.0	38.5
CoNLL03	73.2	46.0	63.5	49.1	50.0	0.0	68.4	43.4	54.5
CoNLL04	58.9	49.0	63.5	52.7	51.3	0.0	68.4	67.6	54.2
CoNLL05	66.1	52.0	63.5	48.4	48.6	0.0	68.4	67.5	38.8
CoNLL12	73.2	49.0	63.5	49.1	52.0	0.0	69.4	66.4	37.7
Ontonotes	71.4	52.0	63.5	56.0	49.8	0.5	68.4	35.0	54.3