Exposé on Master's Thesis

Name: João Rodrigo Olenscki
Matr. Nr.: 03751449
EMail: joao.olenscki@tum.de
Supervisor: Cheng Lyu, M.Sc.; Prof. Constantinos Antoniou, Ph.D.
Institute: Chair of Transportation Systems Engineering
Title: Improving spatio-temporal traffic prediction through transfer learning
Date: November 16, 2023

1 Topic

According to [30], a smart city can be defined as a well-coordinated system that integrates advanced technological infrastructure, relying on sophisticated data processing. The primary objectives of such integration are to enhance city governance efficiency, improve citizen satisfaction, foster business prosperity, and promote environmental sustainability. Within a smart city, the management of various individual systems that constitute the urban environment is not solely reliant on the data collected within the city; it also relies on different adaptable models that can learn and evolve to suit the specific needs and characteristics of the city. In this context, the development of traffic prediction models emerges as a pivotal component for establishing the foundational framework of smart city management.

Recent advances in deep learning have led to significant advancements in prediction tasks related to traffic, such as crowd flow [10, 32], traffic flow [20, 27], public transit flow [4, 14], travel demands [9], and traffic speeds [31]. While formidable in their predictive power, these models come with a substantial data appetite. This data requirement poses a challenge for initiating new intelligent networks because meaningful inferences remain elusive despite the considerable investment needed to establish the sensor network without access to substantial data history. This difficulty is known in the field as the "cold-start" problem.

To address the aforementioned challenge, novel techniques rooted in transfer learning [18] have been introduced. These approaches enable training predictive traffic models for cities constrained by limited data by taking advantage of patterns observed in cities with abundant data resources. The fundamental concept behind these models involves the application of Multi-task Learning, a type of Inductive Transfer Learning, which entails initializing the network in the source city and implementing fine-tuning to adapt the network to the unique characteristics of the target city.

2 State of Science and Technology

2.1 Traffic Forecasting

Traffic Forecasting is a long-lasting field of study in Traffic Engineering, conceived in the 1950s [2, 3]. Initially centered on traffic simulation, the area observed a significant upward trend in recent years. By its very nature, a traffic network constitutes a vast and complex system where events occurring at various junctures within the road grid can exert profound influence over the entire traffic flow. These inherent complexities render it an ideal subject for examination by cutting-edge machine-learning algorithms, such as LSTM, GRU, and CNN.

Traffic network problems inherently comprise two domains: the temporal and spatial domains. In the early stages of deep learning model development for the field of traffic, a natural division emerged, allocating separate components to address each of these domains. For example, CNN has conventionally been harnessed primarily for spatial feature extraction, which involves identifying elements' physical locations or arrangements within a given dataset. Conversely, RNN has specialized in temporal feature extraction, focusing on discerning patterns that evolve or change over time.

The pioneering work of [16] marked one of the earliest instances of employing CNNs for traffic prediction tasks. Subsequently, this methodology proliferated across various model frameworks and underwent integration with other architectural paradigms. An illustrative example of this evolutionary trajectory emerges from the study conducted by [13], which employed a multiple GCN framework. In this network type, the input to the convolutional layers consists of the city's graph representation.

Similarly, [34] was among the early proponents of employing LSTM cells for short-term traffic prediction. This approach gained substantial traction, integrating cells into various models, as exemplified by the ConvLSTM module introduced by [29]. In this architecture, spatial features were extracted at each time frame and subsequently fed into an LSTM chain. This innovative approach allowed for the extraction of spatial and temporal features concurrently.

As a complex spatio-temporal problem, traffic forecasting offers a range of strategies, spanning from problem formulation to data structuring, encompassing data type selection. Regarding data representation, many authors [24, 25, 28] choose to apply a grid in the city and treat each 1-by-1 region autonomously, computing variables (inflow and outflow, for instance) inside these boundaries. In this approach, the instantaneous snapshot of the city could be compared to an image in the context of image classification algorithms, with each pixel being equivalent to a region.

An alternative to this structure is transforming the raw data, typically provided as a matrix or tensor, into a graph-based representation, with each geographical region converted into a distinct node and an adjacency matrix defining the connections between neighboring regions [8, 11, 13, 15, 17, 22, 23, 26, 33]. This method can better capture the network's complexity and regions' nuanced connectivity. In contrast to the grid-based approach, where regions may share a border without necessarily sharing a connecting road, the graph representation effectively accounts for such subtleties.

A third strategy, applied in the works of [6, 12], involves defining individual road segments as graph nodes. This approach offers significantly greater detail and precision in modeling traffic data as the data sources are condensed in a relatively small area. As a drawback, it also requires the installation of many more sensors to produce the data.

2.2 Transfer Learning

Transfer learning techniques [18] were introduced as valuable tools for addressing problems where existing knowledge or expertise from one domain could be employed to enhance learning and performance in another domain. These techniques prove incredibly beneficial when the latter domain suffers from a shortage of data. These techniques are widely used in NLP problems such as sentiment and document classification. Furthermore, the computer vision field also benefited greatly from applying these approaches, as they are extensively used for image classification.

As many cities started to prepare themselves to transition to become "smart cities," they stumbled upon the "cold start" problem. This problem refers to the lack of data that a city faces after installing the sensor network and before acquiring enough data to justify deep learning models, and it is not unique to smart cities but resonates with the broader field of machine learning [1]. In such a situation, despite all investments made in predicting traffic, managers would be required to wait for at least three months until they can make reasonable predictions with deep learning models. In these cases, based on the assumption that despite being different, some cities can share a common framework of behavior and have similar data distributions, transfer learning acts in favor of acquiring knowledge from cities with abundant data and using this knowledge to understand and predict cities with scarce data.

Furthermore, Transfer Learning techniques are also suitable for intra-city transfer, i.e., transferring the learning from a domain (for instance, bike sharing flow) to another in the same city (for example, pedestrian flow). This can observed in the work of [25], in which the author used a taxi trip dataset in New York to learn about bike sharing in the same city.

Generally, a transfer learning algorithm consists of three parts: feature extraction, in which spatio-temporal features are, through various ways, obtained; domain adaptation, in which the knowledge is transferred; and a predictor. The feature extraction step is also present in deep learning approaches to the traffic forecasting problem. It can be achieved by many different architectures, as discussed in Section 2.1. On the other hand, the domain adaptation step is mainly present in the transfer learning networks and aims to generate transferable latent features between the domains. With it, one aims to learn domain-invariant knowledge about the problem.

3 Objective of the Thesis

The main objective of this work is to analyze and explore state-of-the-art models for traffic prediction to enhance their accuracy. As a matter of organization, the following secondary objectives were drawn:

- Analyze current state-of-the-art models and identify cells and architectures that could be used when building a novel model;
- Based on the results of the first objective, propose a novel model to be built, which should follow these requirements:
  - capable of learning from multiple cities at the same time;
  - capable of taking external data (such as weather data, POI locations, holidays' calendar) as a supplement; and
  - with a pre-processing step that includes a data augmentation cell.
- Analyze how impactful each feature derived from a requirement is to the model.

Concurrently, the following research questions were raised:

- Q1. Is it possible to encompass more than two cities as sources in a transfer learning process?
- Q2. What's the impact of the number of sources on the model's accuracy?
- Q3. Is there a limit on the number of sources?
- Q4. Is data augmentation possible in the traffic prediction field?

4 Work Plan and necessary Resources

During the development of this thesis, the following activities will be performed, as stated and detailed in Figure 1:

- Bibliographic revision
- Development of test framework

- Analysis of possible architectures
- Evaluation and optimization
- Thesis writing

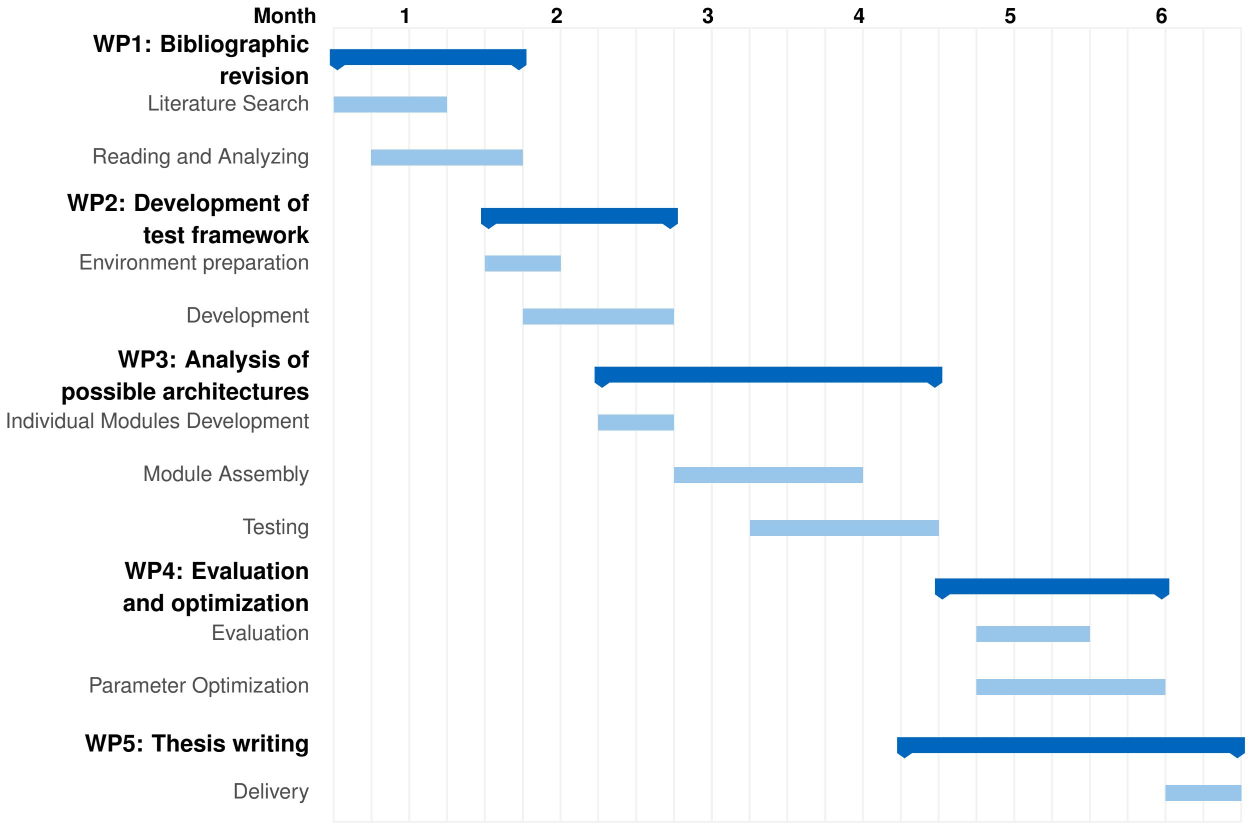

Figure 1: Time schedule.

As for the necessary resources, the models in the thesis will be designed and created with PyTorch [19] and its' numerous submodules [7, 21]. The data that will feed the models is the NeurIPS 2021 competition [5] organized by the Institute of Advanced Research in Artificial Intelligence (IARAI).

5 Literature

[1] Ali, Z., Kefalas, P., Muhammad, K., Ali, B., and Imran, M. "Deep learning in citation recommendation models survey". In: *Expert Systems with Applications* 162 (Dec. 2020), p. 113790. DOI: 10.1016/j.eswa.2020.113790. URL: https://doi.org/10.1016/j.eswa.2020.113790.

[2] Beckmann, M. J., McGuire, C. B., and Winsten, C. B. *Studies in the Economics of Transportation*. Santa Monica, CA: RAND Corporation, 1955.

[3] Bevis, H. W. "A Model for Predicting Urban Travel Patterns". In: *Journal of the American Institute of Planners* 25.2 (May 1959), pp. 87–89. DOI: 10.1080/01944365908978310. URL: https://doi.org/10.1080/01944365908978310.

[4] Chai, D., Wang, L., and Yang, Q. "Bike flow prediction with multi-graph convolutional networks". In: *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems*. 2018, pp. 397–400.

[5] Eichenberger, C., Neun, M., Martin, H., Herruzo, P., Spanring, M., Lu, Y., Choi, S., Konyakhin, V., Lukashina, N., Shpilman, A., Wiedemann, N., Raubal, M., Wang, B., Vu, H. L., Mohajerpoor, R., Cai, C., Kim, I., Hermes, L., Melnik, A., Velioglu, R., Vieth, M., Schilling, M., Bojesomo, A., Marzouqi, H. A., Liatsis, P., Santokhi, J., Hillier, D., Yang, Y., Sarwar, J., Jordan, A., Hewage, E., Jonietz, D., Tang, F., Gruca, A., Kopp, M., Kreil, D., and Hochreiter, S. "Traffic4cast at NeurIPS 2021 - Temporal and Spatial Few-Shot Transfer Learning in Gridded Geo-Spatial Processes". In: *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*. Ed. by Kiela, D., Ciccone, M., and Caputo, B. Vol. 176. Proceedings of Machine Learning Research. PMLR, June 2022, pp. 97–112. URL: https://proceedings.mlr.press/v176/eichenberger22a.html.

[6] Elmi, S. and Tan, K.-L. "Travel Time Prediction in Missing Data Areas: Feature-based Transfer Learning Approach". In: 2020, pp. 1088–1095. DOI: 10.1109/HPCC-SmartCity-DSS50907.2020.00196. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85105246596&doi=10.1109%2fHPCC-SmartCity-DSS50907.2020.00196&partnerID=40&md5=2071db3ecbac45353dd2132a1528fa7

[7] Fey, M. and Lenssen, J. E. "Fast Graph Representation Learning with PyTorch Geometric". In: *ICLR Workshop on Representation Learning on Graphs and Manifolds*. 2019.

[8] Geng, X., Li, Y., Wang, L., Zhang, L., Yang, Q., Ye, J., and Liu, Y. "Spatiotemporal Multi-Graph Convolution Network for Ride-Hailing Demand Forecasting". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.01 (July 2019), pp. 3656–3663. DOI: 10.1609/aaai.v33i01.33013656. URL: https://doi.org/10.1609/aaai.v33i01.33013656.

[9] Geng, X., Li, Y., Wang, L., Zhang, L., Yang, Q., Ye, J., and Liu, Y. "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. 01. 2019, pp. 3656–3663.

[10] Jin, W., Lin, Y., Wu, Z., and Wan, H. "Spatio-temporal recurrent convolutional networks for city-wide short-term crowd flows prediction". In: *Proceedings of the 2nd International Conference on Compute and Data Analysis*. 2018, pp. 28–35.

[11] Jin, Y., Chen, K., and Yang, Q. "Selective Cross-City Transfer Learning for Traffic Prediction via Source City Region Re-Weighting". In: 2022, pp. 731–741. DOI: 10.1145/3534678.3539250. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85137142449&doi=10.1145%2f3534678.3539250&partnerID=40&md5=7fece17da0d95984f76371135da8a047.

[12] Li, J., Xie, N., Zhang, K., Guo, F., Hu, S., and Chen, X. ("Network-scale traffic prediction via knowledge transfer and regional MFD analysis". In: *Transportation Research Part C: Emerging Technologies* 141 (2022). DOI: 10.1016/j.trc.2022.103719. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85131098775&doi=10.1016%2fj.trc.2022.103719&partnerID=40&md5=42cc6ec5698a2ce553c7b1a3c6b3025e.

[13] Lin, J., Chen, Y., Zheng, H., Ding, M., Cheng, P., and Hanzo, L. "A Data-driven Base Station Sleeping Strategy Based on Traffic Prediction". In: *IEEE Transactions on Network Science and Engineering* (2021). DOI: 10.1109/TNSE.2021.3109614. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85114716577&doi=10.1109%2fTNSE.2021.3109614&partnerID=40&md5=ae21355623e27ac918b3b717e6f05869.

[14] Liu, Y., Liu, Z., and Jia, R. "DeepPF: A deep learning based architecture for metro passenger flow prediction". In: *Transportation Research Part C: Emerging Technologies* 101 (2019), pp. 18–34.

[15] Lu, B., Gan, X., Zhang, W., Yao, H., Fu, L., and Wang, X. "Spatio-Temporal Graph Few-Shot Learning with Cross-City Knowledge Transfer". In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2022. DOI: 10.1145/3534678.3539281. URL: https://doi.org/10.1145/3534678.3539281.

[16] Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., and Wang, Y. "Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction". In: *Sensors* 17.4 (Apr. 2017), p. 818. DOI: 10.3390/s17040818. URL: https://doi.org/10.3390/s17040818.

[17] Ouyang, X., Yang, Y., Zhou, W., Zhang, Y., Wang, H., and Huang, W. "CityTrans: Domain-Adversarial Training with Knowledge Transfer for Spatio-Temporal Prediction across Cities". In: *IEEE Transactions on Knowledge and Data Engineering* (2023), pp. 1–14. DOI: 10.1109/TKDE.2023.3283520. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85161563633&doi=10.1109%2fTKDE.2023.3283520&partnerID=40&md5=f8360c1d3d50b9aac67d4fbbf3d28868.

[18] Pan, S. J. and Yang, Q. "A survey on transfer learning". In: *IEEE Transactions on knowledge and data engineering* 22.10 (2009), pp. 1345–1359.

[19] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. Curran Associates, Inc., 2019, pp. 8024–8035. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[20] Polson, N. G. and Sokolov, V. O. "Deep learning for short-term traffic flow prediction". In: *Transportation Research Part C: Emerging Technologies* 79 (2017), pp. 1–17.

[21] Rozemberczki, B., Scherer, P., He, Y., Panagopoulos, G., Riedel, A., Astefanoaei, M., Kiss, O., Beres, F., Lopez, G., Collignon, N., and Sarkar, R. "PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models". In: *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*. 2021, pp. 4564–4573.

[22] Tang, Y., Qu, A., Chow, A. H. F., Lam, W. H. K., Wong, S. C., and Ma, W. "Domain Adversarial Spatial-Temporal Network: A Transferable Framework for Short-term Traffic Forecasting across Cities". In: (2022). DOI: 10.48550/ARXIV.2202.03630. URL: https://arxiv.org/abs/2202.03630.

[23] Wang, L., Chai, D., Liu, X., Chen, L., and Chen, K. "Exploring the Generalizability of Spatio-Temporal Traffic Prediction: Meta-Modeling and an Analytic Framework". In: *IEEE Transactions on Knowledge and Data Engineering* 35.4 (Apr. 2023), pp. 3870–3884. DOI: 10.1109/tkde.2021.3130762. URL: https://doi.org/10.1109/tkde.2021.3130762.

[24] Wang, L., Geng, X., Ma, X., Liu, F., and Yang, Q. "Cross-City Transfer Learning for Deep Spatio-Temporal Prediction". In: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Aug. 2019. DOI: 10.24963/ijcai.2019/262. URL: https://doi.org/10.24963/ijcai.2019/262.

[25] Wang, S., Miao, H., Li, J., and Cao, J. "Spatio-Temporal Knowledge Transfer for Urban Crowd Flow Prediction via Deep Attentive Adaptation Networks". In: *IEEE Transactions on Intelligent Transportation Systems* 23.5 (2022), pp. 4695–4705. DOI: 10.1109/TITS.2021.3055207. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85100868963&doi=10.1109%2fTITS.2021.3055207&partnerID=40&md5=dddaae60893f302b2d084dd455c1d772.

[26] Wei, Y., Zheng, Y., and Yang, Q. "Transfer Knowledge between Cities". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016. DOI: 10.1145/2939672.2939830. URL: https://doi.org/10.1145/2939672.2939830.

[27] Wu, Y., Tan, H., Qin, L., Ran, B., and Jiang, Z. "A hybrid deep learning based traffic flow prediction method and its understanding". In: *Transportation Research Part C: Emerging Technologies* 90 (2018), pp. 166–180.

[28] Yao, H., Liu, Y., Wei, Y., Tang, X., and Li, Z. "Learning from Multiple Cities: A Meta-Learning Approach for Spatial-Temporal Prediction". In: *The World Wide Web Conference*. ACM, May 2019. DOI: 10.1145/3308558.3313577. URL: https://doi.org/10.1145/3308558.3313577.

[29] Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., and Li, Z. "Deep multi-view spatial-temporal network for taxi demand prediction". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 32. 1. 2018.

[30] Yin, C., Xiong, Z., Chen, H., Wang, J., Cooper, D., and David, B. "A literature survey on smart cities." In: *Sci. China Inf. Sci.* 58.10 (2015), pp. 1–18.

[31] Yu, B., Yin, H., and Zhu, Z. "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting". In: *arXiv preprint arXiv:1709.04875* (2017).

[32] Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., and Li, T. "Predicting citywide crowd flows using deep spatio-temporal residual networks". In: *Artificial Intelligence* 259 (2018), pp. 147–166.

[33] Zhang, L., Geng, X., Qin, Z., Wang, H., Wang, X., Zhang, Y., Liang, J., Wu, G., Song, X., and Wang, Y. "Multi-Modal Graph Interaction for Multi-Graph Convolution Network in Urban Spatiotemporal Forecasting". In: *Sustainability* 14.19 (Sept. 2022), p. 12397. DOI: 10.3390/su141912397. URL: https://doi.org/10.3390/su141912397.

[34] Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., and Liu, J. "LSTM network: a deep learning approach for short-term traffic forecast". In: *IET Intelligent Transport Systems* 11.2 (Feb. 2017), pp. 68–75. DOI: 10.1049/iet-its.2016.0208. URL: https://doi.org/10.1049/iet-its.2016.0208.

6 Structure of the Thesis

1. Introduction
   (a) Motivation
   (b) Research Questions
   (c) Contribution
   (d) Outline
2. Literature Review
   (a) Traffic Forecasting
   (b) Transfer Learning
   (c) Domain Adaptation
3. Methodology
   (a) Data Analysis and Exploration
   (b) Loss Function
   (c) Model Outline
   (d) Feature Extraction Network
      - Offline Training
      - Online Training
   (e) Embedding Network
   (f) Prediction Network

(g) Experiments

4. Results

   (a) Experiments Results

   (b) Next steps

5. Conclusion

Review

Content presentation created by:

______, ______	______, ______	______, ______
Place, date	*Place, date*	*Place, date*
______________	______________	______________
Student	First examiner, Supervising scientific staff at the chair	Second examiner, Center of Key Competencies