



Foundation 03/2025

UNIDROIT Foundation
EE ICLR 11th Workshop
Cambridge, 23 January 2025

11th Meeting of the Economic Evaluation of International Commercial Law Reform Project

Main Conclusions and next steps

1. The 11th session for the project on Economic Evaluation of International Commercial Law Reform (EE ICLR Project) took place in Cambridge and online on 23 January 2025. For the agenda and the list of participants, please see [Annex 1](#) and [Annex 2](#).

Item 1: Opening of the Session

2. The Directors of the Cape Town Convention Academic Project, *Professor Louise Gullifer* (Cambridge University), *Professor Jeffrey Wool* (President of the UNIDROIT Foundation and Secretary-General of the Aviation Working Group) and *Professor Ignacio Tirado* (UNIDROIT Secretary-General), opened the session and welcomed all the participants.

3. The Directors expressed their appreciation for Professor Jordi Paniagua's contribution to the project, and acknowledged the significant progress made on the project.

Item 2: Summary of intersessional work and introduction to the revised Draft Guide and Case Study

4. The *UNIDROIT Secretariat* (*Ms Theodora Kostoula*) and *Professor Paniagua* (Consultant to the project) presented the intersessional work undertaken between the 10th and 11th Workshops, which included an intersessional meeting, and several changes made to the Framework and the Guide. A notable change concerned the shift to "transaction costs" as the main target of the ICLR, while broader economic frictions (introduced in Annex 2 of the Guide) should be acknowledged where relevant and directly linked to the law reform. It was clarified that the scope of the Guide should focus on private law, while the section related to timing and the nature of law reforms was further elaborated and largely revised. Other main changes included a new "adoption phase" within the *ex-ante* evaluation stage, to reflect the phase between the approval of the instrument and the wide adoption by states.

5. The Factors were further refined. A significant change was made to Factor E, which was scaled down (from -10,10 to -1,10), to represent 10% of the weight of other Factors. In addition, tables with bands and qualitative assessments were introduced in the scoring methodology, using a traffic-light model to provide both quantitative economic outcomes and qualitative interpretation.

6. The "Benchmark" section was further improved and expanded to provide guidance on how to select and prioritise Benchmarks, on the use of multiple and diverse Benchmarks,

and the creation of constructed Benchmarks based on multiple existing Benchmarks rather than from scratch.

7. A new Case Study on a hypothetical legislative proposal for Dispute Resolution for Cross-border Transactions of Micro, Small and Medium Entities (MSMEs) and Rural Communities was developed to test the Framework and Guide at the conceptual stage of the proposed law reform, helping understand how the constructed Benchmark could be designed.

Item 3: Framework and Factors

Factors

8. *The participants* discussed the changes to the Framework and the Factors which included expanding Factor D through indicative questions, and scaling down Factor E.

9. *The participants* further discussed Factor C, noting that it required a stricter focus on quantifiable effects to avoid abstract comparisons.

Scoring Methodology

10. *Professor Paniagua* introduced the ES bands and rating tables, which would allow for a qualitative interpretation of the ES. He explained that the scoring system functioned similarly to traffic light indicators used by many scoring agencies, including the World Bank. Under this system, scores between 7 and 10 were categorised in the green zone, indicating a highly favourable impact. Scores between 5 and 7 fell into a lighter green zone, reflecting moderate but positive effects. Lower scores (0–1) or negative values indicated adverse effects relative to the Benchmark.

11. *Professor Gullifer* noted that this approach facilitated easy comparison across different projects, making it particularly useful for prioritisation.

12. *A participant* sought clarification on the formula, raising concerns about the reliability of economic scoring, especially in the early stages of evaluations where strong Benchmarks or robust data may be lacking. The participant suggested using qualitative rating bands instead of precise numerical scores to provide more flexibility and account for potential imprecision. *Another participant* with an economics background highlighted the importance for decision-makers of distinguishing between small variations, such as an 8.1 versus an 8.3. Even minor differences could influence project prioritisation, as stakeholders might question why one project scored 8.3 while another scored 8.0. The ability to explain these differences, such as higher implementation costs, added value to the decision-making process by contextualising scores.

13. *Professor Paniagua* reiterated that a key strength of the Framework was its ability to provide quantification, particularly for comparing different law reforms. He suggested that numerical scoring and qualitative rating bands were not mutually exclusive; the quantitative score offered precision, while the qualitative rating provided a broader interpretative Framework. He further explained that both qualitative and quantitative evaluations were incorporated into the Framework, noting that the final score was not only a numerical value but also converted into a qualitative rating, such as "AA". *Professor Wool* expressed support for this approach, affirming the importance of including multiple assessment perspectives.

14. *Professor Gullifer* pointed out the importance of numerical scores for decision-makers. As long as the same methodology was consistently applied across projects, numerical scores provided a meaningful basis for comparison. While the practical difference

between an 8.5 and an 8.1 might be limited, applying a consistent methodology would enhance the reliability of comparisons between projects.

15. *The participants* suggested clarifying that economic impact was just one factor among many to consider, depending on the policy context.

16. *The participants* also discussed how the scoring methodology had become a formal step in the workflow. *Professor Paniagua* highlighted a key addition: the introduction of a "correspondence" mechanism to increase transparency. For example, if a Factor met 75% of the relevant criteria, it would receive a score of 7.5. He emphasised that this approach ensured transparency and provided a clear interpretation of scores. To illustrate this, he referenced the Case Study on MSMEs, where the economic score was calculated as 8 with an 85% degree of certainty. The 15% reduction in certainty was due to limited data on law reforms targeting SMEs. He further explained that since a study showed the UNCITRAL Model Law increased international trade flows by 7%, the expected impact for a score of 8 would be 80% of that 7% increase, equating to a 4% increase.

17. *A participant* questioned how the formula integrated numerical values, particularly when some Factors were absent. *The participant* posed a hypothetical scenario where a law reform scored perfectly (10/10) on Factor A but had no impact on Factors B and C. *The participants* discussed whether the Framework might undervalue highly technical reforms with no indirect or network effects, noting that international organisations might still adopt such reforms despite a potentially lower overall score. *Ms Kostoula* acknowledged this concern but noted that it was rare for any Factor to be absent, particularly Factor B, which would almost always be present in international law reforms due to their global nature.

18. *Professor Paniagua* clarified that scores were relative to the Benchmark. If a law reform had no indirect effects but the Benchmark also lacked indirect effects, Factor B could still receive a score of 10 due to this alignment. This ensured that the reform would not be penalised for lacking certain effects if the Benchmark set the same expectation. He further elaborated that if a law reform, by definition, lacked indirect effects, such as a reform targeting a narrow constituency, a perfect Benchmark for that type of reform would also exhibit low network effects. If the network effects corresponded one-to-one, the score for Factor B would be 10. He emphasised that the Framework evaluated reforms in relation to the Benchmark rather than in absolute terms. If both the reform and the Benchmark lack direct effects, their alignment justified a score of 10.

19. *Professor Wool* noted that the concept of converting the scoring to numbers was significant. He explained that in scenarios where a high score resulted from a one-to-one mapping against a Benchmark with a low B or C, this would be valid within the Framework. However, when converted to numbers, the impact of that part would be zero. He argued against revising the entire Framework.

20. *The participants* discussed how to handle cases where certain Factors, such as Factors B and C were absent. It was proposed to explain the rule of three and reframe the calculation methodology.

21. *Several participants* cautioned against changing the equal weighting of the three Factors, as this would alter the tool's function as a comparative Framework. Introducing different weights would complicate the methodology and make it harder for readers to interpret the results.

22. The default calculation would still be based on division by three, but the formula could be adjusted when one Factor was not relevant. An alternative formula, such as " $(A + B) * D / 2$," was suggested to account for missing Factors while maintaining consistency in

results. This approach should be outlined in the quantification section of the Guide. *Professor Paniagua* added that the Framework was flexible enough to accommodate both the absence of Factors and the inclusion of unforeseen elements. For instance, the formula could be expanded in the future if additional Factors were identified. He suggested that if a fourth Factor were introduced, such as Factor F, the formula should be adjusted to " $(A+B+C+F)/4$."

23. *Professor Anna Veneziano* raised concerns about removing a Factor when it was very low. She noted that in most cases, any Factor might still have some relevance even if it was minor.

24. *It was suggested (i) explaining that quantification of economic outcomes relied on the ceteris paribus assumption; (ii) specifying that the "division by 3" should apply only when all three Factors were present. Where one Factor was absent, the Framework could be reformulated accordingly (e.g., division by 1 or 2); adding the ceteris paribus assumption.*

Item 4: Benchmark and ex-post analysis

25. The discussion shifted to the selection and role of Benchmarks.

26. *Ms Kostoula* and *Professor Paniagua* emphasised the importance of Benchmarks describing them as past or existing legal reforms used as reference points to evaluate the potential impact of proposed changes. The concept, introduced during the previous workshop by *Professor Paniagua*, aimed to identify best practices and highlight successful models without necessarily representing the ideal legal solution. Drawing from the Case Study, they stressed the need to use multiple Benchmarks to cover different legal aspects, assess various challenges, and provide a more comprehensive evaluation. Especially at the conceptual stage, multiple and diverse Benchmarks could be used to reflect different legislative options, especially when the content of future legal rules remained uncertain. For example, different Benchmarks might focus on arbitration versus other dispute resolution mechanisms. The Framework allowed for the application of various Benchmarks to different scenarios to provide a well-rounded assessment of proposed reforms.

27. *Ms Kostoula* clarified two key points following discussions from the international meeting: (i) Benchmarks did not represent the ideal law but economic performance. In other words, Benchmarks were selected on economic terms, functioning as tools to identify best-in-class performers; (ii) A constructed Benchmark was not created from scratch but assembled from existing Benchmarks, ensuring it remained grounded in real-world application.

28. *Professor Gullifer* highlighted that these were both crucial points, emphasising that the Framework's strength lies in using Benchmarks without assuming that a Benchmark reflects an ideal legal model. This approach enabled a more realistic assessment of law reforms by separating economic performance from theoretical legal strength. *Professor Paniagua* added that an ideal law based on theoretical principles might not generate the expected economic impact. Conversely, a law that was theoretically weaker could prove more impactful in practice. Therefore, real-world economic impact should take precedence over theoretical strength when selecting Benchmarks. He also explained that in cases where no direct Benchmark existed or none fully applied, a Benchmark can be constructed by combining elements from multiple existing ones.

29. *Professor Gullifer* raised two additional points. First, she stressed the hope that, over time, the methodology developed in the Guide might be increasingly used for ex-post studies, which in turn could inform ex-ante evaluations. Second, even failed law reforms could provide valuable insights. While identifying "best-in-class" Benchmarks and building on successful models was essential, reviewing past mistakes and understanding why some reforms failed could also improve future legislation. In light of this, she questioned whether

failed or less successful law reforms could be factored in the benchmarking process. *Several participants* noted that qualitative studies could help explain why reforms failed and provide context for their economic impact.

30. *The participants* then discussed criteria for selecting the best-in-class Benchmark. Benchmarks should be relevant and data-backed rather than speculative. When no suitable Benchmark existed, constructing one from multiple sources may be necessary.

31. The discussion highlighted the importance of distinguishing between empirical and theoretical studies in benchmarking. *Ms Kostoula* clarified that the benchmarking approach includes both quantifiable and non-quantifiable studies, but quantifiable data should be prioritised. *Some participants* questioned the utility of non-quantified studies in benchmarking. *Professor Wool* noted that some references might provide useful insights but might not be strong enough to serve as Benchmarks. Such sources should be categorised appropriately. However, it was agreed that while a Benchmark should ideally have *ex-post*, quantified economic studies, qualitative case studies could still provide useful context. Theoretical, non-quantifiable studies should support empirical analysis rather than replace it.

32. Benchmarks should be chosen based on economic evidence rather than legal classification. Selection would involve identifying relevant issues and drawing functional analogies across legal fields. *Ex-post* studies should be conducted rigorously, following the Guide's methodological Framework, and relying on high-quality empirical evidence.

33. *One participant* noted the triangulation method used in the Case Study, combining different sources such as SME interventions, digital tools, and dispute resolution data to construct more relevant Benchmarks. It was added that Benchmarks must align with market inefficiencies and policy objectives. If a Benchmark did not reflect the policy goal (e.g., SME dispute resolution requiring cost-effectiveness), it should not be used.

34. *The participants* further discussed whether the scoring system should allow Benchmarks to be exceeded if a law reform outperformed the reference point. *Professor Gullifer* noted that while a Benchmark score represented the minimum expected economic impact, actual benefits could exceed that estimate. *Some participants* argued that best-in-class Benchmarks should be assigned a flexible score, allowing for improvements beyond the reference point.

35. It was explained that the degree of certainty was intended to account for cases where a Benchmark did not fit fully. Moreover, the degree of certainty should accommodate higher-than-expected performance and could account for cases where a reform was expected to outperform its Benchmark. For example, if a Benchmark scored at 10 but the degree of certainty was 80%, the expected impact range could be between 8 and 12. *Professor Gullifer* stressed the importance of this clarification, noting that most law reforms aim to exceed existing models rather than merely replicate them.

36. *Professor Gullifer* concluded that the methodology should not be changed, as it was already well-developed and robust. However, she agreed that this issue should be acknowledged in the Guide, noting that in real-world cases, decision-makers might consider the possibility that a reform could have a greater economic impact than its Benchmark.

37. *The participants agreed to provide (i) further guidance on selecting Benchmarks; (ii) emphasise the importance of quantitative data over qualitative studies; (iii) include examples in text boxes to illustrate how Benchmarks were selected; (iv) clarify that the search for Benchmarks with similar policy goals was relevant not only to Factor C but also to other Factors; (v) explain that Benchmarks were chosen based on economic*

considerations, not as optimal legal standards; (vi) clarify that constructed Benchmarks were not created from scratch but were based on a combination of existing Benchmarks.

Ex-post analysis

38. *Professor Paniagua* outlined two key aspects of the *ex-post* analysis in the Guide (Annex 2). First, while the Guide primarily focused on *ex ante* evaluations, it acknowledged that *ex-post* evaluations could be conducted at later stages. Second, the Annex aimed to establish a connection between *ex-ante* and *ex-post* evaluations, particularly in relation to Benchmark selection and best practices which highly relied on *ex-post* studies. Those studies should rely on rigorous economic methodologies, such as causal inference techniques, randomised control trials, instrumental variables, and difference-in-differences analysis, as described in that section. The section further clarified the distinction between “causal inference” and “conceptual causal links,” emphasising the importance of using precise terminology.

39. *Removing the repetition of this distinction in the Workflow chapter (VI) and Annex 2 was suggested.*

Item 5: Degree of Certainty

40. *The participants* discussed the concept of the degree of certainty, focusing on how to assign a specific percentage and its impact on the precision of scores. *Professor Paniagua* explained that the degree of certainty was intended to be determined by the evaluation team and the value reflected an expert assessment of probability. The decision of the team should be based on statistical reasoning, similar to confidence intervals used in statistical analysis. The degree of certainty reflected uncertainty about whether those results would apply to a specific context. If such uncertainty existed, it could be quantified, for example, as 85%. This would reflect a reasonably high level of certainty that the Benchmark outcomes would apply to the proposed reform.

41. *The participants* agreed on the need for a detailed methodology for determining certainty levels to prevent arbitrary adjustments. This should include guidance on what the certainty levels represented and why certain values, such as 85%, 90%, or 50%, were assigned.

42. *Professor Paniagua* outlined two key considerations for determining the degree of certainty. First, the certainty score should influence the range of possible outcomes. For instance, if a certainty level of 85% was assigned, the resulting score should allow for an upward or downward adjustment of 1.5 points. This meant that a score of 9 would translate to a range of approximately 7.5 to 10.5. Second, guidance should clarify how different levels of certainty would affect the final score, with reference to established rating categories (e.g., AAA).

43. *Professor Gullifer* emphasised that the level of certainty primarily measured the proximity of the Benchmark to the subject being evaluated. It reflected the strength of the comparison rather than the quality of the reform itself. If a suitable Benchmark was not available, the certainty level would be lower, which would impact the final score’s range but not necessarily suggest that the reform was beneficial or not.

44. *Professor Paniagua* further elaborated, explaining that a certainty level of 100% would mean the score directly reflects the Benchmark outcome without adjustments. If the certainty level was 50%, the assigned score would require an upward and downward adjustment of 50. In cases where the law reform was suboptimal or poorly implemented, providing range would serve as a safeguard for implementing the Framework. To ensure a

more accurate evaluation, *Professor Paniagua* suggested assigning a range rather than a definitive score. For example, instead of assigning a fixed score of 8, a range of 6 to 10 could be provided.

45. The discussion then focused on whether certainty should be represented as numerical ranges or qualitative labels. *Some participants* favoured numerical ranges for their precision and flexibility, while others preferred simpler qualitative labels such as "high" or "low" to reduce complexity.

46. *Professor Paniagua* supported the use of ranges, noting that they would clarify how certainty levels affected scoring and could accommodate scenarios where reforms outperformed Benchmarks. For example, if a score of 9 had an 80% certainty level, the final range could be between 8 and 11. He suggested that while the range may not need to appear in the final output, it should be reflected in the evaluation process.

47. *A participant* pointed out that on the other hand, numerical scores typically derived ranges from statistical models, such as confidence intervals, which provided a clear basis for interpretation. It was explained that this was distinct from simply introducing a range without a structured methodology.

48. *Professor Gullifer* highlighted that introducing ranges into the Framework would add complexity and require adjustments to the methodology. She suggested that interpreting results within ranges, rather than formally adjusting scores, would be more practical. *Professor Wool* supported using qualitative terms like "high," "medium," and "low" to express certainty levels, arguing that precise percentages might give a false impression of accuracy.

49. *Professor Gullifer* added that certainty levels depend on the type and stage of the law reform process. Certainty would likely increase as a reform moved closer to implementation since more data would be available. In contrast, early-stage reforms would involve higher uncertainty due to unknown Factors.

50. *Some participants* referenced the use of percentage-based estimates in the project's Case Studies (e.g., a 7% increase in trade), noting that absolute values were not always possible. *Professor Paniagua* explained that economic effects could be meaningfully expressed in percentages and, where possible, converted to monetary values if total figures were available. However, relying on absolute values might not be practical in markets with volatile currencies.

51. To maintain consistency, *the participants* proposed a classification system using either percentage-based ranges (e.g., 80%–90%) or qualitative terms ("very high," "high," "medium," "low"), or a combination of both. This hybrid approach would offer clarity and flexibility for users.

52. *The participants* further agreed that certainty levels should always be determined relative to the Benchmark.

53. The discussion then shifted to the relationship between the degree of certainty and externalities. *Ms Myrte Thijssen* observed that externalities, such as environmental or social impacts, could influence certainty levels and asked for clarification on how these Factors should be treated in the Framework.

54. *Professor Paniagua* clarified that the term "degree of certainty" had been replaced with "level of certainty" to better reflect alignment with the Benchmark. He explained that a perfect match between the Benchmark and the reform would result in a high level of certainty. However, discrepancies introduced by externalities would lower certainty. He emphasised that while externalities, or unintended consequences, should not directly

influence the economic score, they should be acknowledged in the certainty assessment. For example, if a Benchmark law had minimal environmental impact, but the proposed reform had significant environmental consequences, the level of certainty would be lower.

55. *The participants* agreed that evaluators would not be expected to conduct full environmental impact assessments. Instead, they should compare the externalities of the Benchmark with those of the proposed reform based on existing *ex-post* studies. For instance, if a proposed maritime law reform could affect sea levels, but the Benchmark law was land-based, the level of certainty would be lower due to the mismatch in environmental impacts. However, if both the Benchmark and the reform had similar externalities, the one-to-one alignment would be maintained, keeping the level of certainty high. *Professor Veneziano* suggested adding clarification in the Guide to explain that only externalities assessed in relation to the Benchmark should be considered when determining certainty.

56. *The participants* agreed to (i) rename the concept of "degree of certainty" to "level of certainty" to better reflect the qualitative approach; (ii) develop structured guidance on assigning certainty levels, by introducing a scoring methodology (e.g., ranges or qualitative labels such as "very high," "high," "medium," "low") and explaining what each level implied; (iii) clarify that the level of certainty would be determined in relation to the Benchmark; (iv) explaining how the level of certainty accounted for the absence of best-in-class Benchmarks; (v) clarify the treatment of externalities and their role in determining the level of certainty.

Item 5: "Context of Use"

57. *Professor Paniagua* suggested clarifying the term "timing" to avoid confusion between the timing of effects of the law reform and the timing phase in which the evaluation was taking place. Replacing it with "sequence" or "timing aspects" was suggested.

58. *Ms Kostoula* explained that the section had been adjusted by replacing "promotional law" with "facilitating law," in line with previous discussions where it was agreed that "promotional" carried a different connotation. Additionally, relevant academic literature had been added. *Professor Tirado* suggested further reflecting and revising this section, raising concerns regarding the terminology. He added that within the same legal framework, there could be both promotional and reactive rules. He suggested revising the wording to reflect this distinction more accurately.

59. *Changing the section title "Timing" to "Timing Aspects" and conducting further research on the concepts of "facilitating law" and "promotive law" within this context was suggested.*

Item 7: Evaluation actors

60. A discussion ensued regarding the selection and role of evaluation and Benchmark teams, recognising their significant influence on the final outcomes.

61. *Professor Paniagua* stressed the importance of maintaining separation between the teams responsible for constructing and evaluating the Benchmark to avoid bias. While logistical and financial constraints could make this difficult, he argued that, ideally, two distinct teams should be assigned these roles.

62. *Several participants* supported the need for multidisciplinary teams consisting of both lawyers and economists. *Ms Kostoula* highlighted that a similar approach was followed in the Case Study, where team members involved in both constructing and evaluating the Benchmarks worked on different aspects independently. For example, *Ms Kostoula* primarily focused on the baseline scenario and legal framework, while *Professor Paniagua* handled the economic analysis. They later collaborated to integrate findings and conduct a peer review.

She acknowledged that full separation might not always be feasible due to resource limitations but proposed that the Guide highlight both the importance of team independence and the value of collaboration to strike a balance between impartiality and active engagement.

63. *A participant* suggested that rather than rigidly enforcing team independence, greater focus should be placed on transparency and consensus-building through external peer review, internal feedback, and response mechanisms, similar to those used by the World Bank. It was noted that independence alone did not guarantee rigorous evaluation, as motivation and engagement also play critical roles in ensuring quality outcomes.

64. To further enhance quality control, *another participant* proposed implementing an auditing process for evaluations conducted by internal teams. This would help prevent bias, such as inflating scores to justify a law reform. Establishing an independent oversight board or involving external experts could help improve the credibility of the evaluation process.

65. *Professor Veneziano* highlighted that in international organisations, decision-making was often guided by governance structures that helped reduce the risk of internal bias. However, she acknowledged concerns about figures being selectively adjusted to support predetermined conclusions.

66. *The participants agreed on the importance of fostering collaboration and synergy between evaluation actors. They proposed that the evaluation methodology include a recommendation to incorporate peer review and external validation. While such mechanisms would not be mandatory, they would enhance transparency and credibility.*

Item 8: Scope

67. *The participants* discussed the importance of the Guide for *ex-post* evaluations. *Professor Tirado* added that the Guide served as an effective initial framework that countries could later supplement with their own tailored analyses. Such analyses may still be important, for example to support the implementation of a law reform at country level. *Professor Veneziano* suggested incorporating a clarification in the introduction, noting that the methodology was not meant to replace country-specific or sector-specific analyses. *Professor Tirado* added that this clarification would also function as a form of disclaimer for the project. He explained that legislators needed to recognise the challenge of developing a universal *ex-ante* methodology that would work perfectly for every country and situation. While the project provided a useful framework, national policymakers would still need to conduct their own assessments using local data.

68. While certain countries conducted traditional economic analyses when adopting legal instruments, these were often limited in scope. *Professor Gullifer* highlighted that the methodology developed in this project represented an important step, as it took into account the broader economic impact by incorporating Factors such as network benefits and broader economic implications. Other key differences included the use of Benchmarks, and a broader perspective of costs and benefits, beyond administrative considerations. *Another participant* added that while the Framework focused on economic efficiency, it did not exclude other considerations.

69. *A participant* referred to the scope of the Guide on private commercial law expressing concerns for situations where the distinction between private and public law elements was less clear. It was suggested that, while the Framework primarily addressed private law, mixed reforms could still be relevant and, where certain public law elements were included in a reform, the feasibility of their inclusion should be considered. *Ms Kostoula* agreed that, if a law reform also had effects on regulatory matters, similar to the approach followed for

economic frictions, those aspects should be considered, provided they were directly connected to the reform.

70. *It was suggested to reflect this discussion in the Guide.*

Item 9: Documents for publication

71. *The participants* discussed the publication of the Guide and the associated documents to be produced.

72. *Professor Wool* suggested developing two distinct versions of the Guide. Based on the discussions, the first version would be the full and official version, containing all references, footnotes, acknowledgments, and supporting Annexes.

73. The second version would function as an electronic template for direct use by the evaluation actors. This version would include an executive summary of the Guide, allowing users to quickly reference key points without needing to consult the full document. The executive summary would reflect the core elements of the Guide while maintaining its overall structure in a more accessible format. Its structure should follow the workflow described in the full version of the Guide, enabling evaluation actors to input data directly without unnecessary navigation through additional content. The primary objective was to ensure practicality, keeping the document lean, easy to use, and free of excessive information. The electronic version should specify the time span of the economic evaluation, which would depend on each specific case rather than adhering to a fixed period. While time periods were necessary for producing numerical assessments, they should remain flexible, and the Framework should not dictate a specific time horizon. However, evaluators should report their assumptions transparently to ensure clarity and consistency in economic analyses.

74. *The participants* further discussed the Annexes, including Annex 2 on economic frictions and Annex 4 on *ex-post* economic evaluation. A *participant* suggested removing the footnotes related to the legal context, such as on harmonisation, from the main text and placing them in a separate law-related Annex. This would allow legal considerations to be presented separately, similar to the approach already taken for the economic content of current Annex 2. *The participants* agreed that the text should be operational rather than academic. This implied maintaining footnotes only where strictly necessary for the Framework's function. A cross-reference in the main text would direct readers to Annex 1 for legal clarifications. It was suggested integrating most of them in the main text when directly linked to the analysis or placed in Annex 1 where less necessary.

75. In addition, it was suggested that one or more Case Studies be published to serve as examples for evaluation actors. The Case Study on SMEs in particular was considered to be the most complete and appropriate one for better understanding the application of the Guide. The Case Study itself, or a link to the to external source where it would be published, should be included as Annex 5. The final decision on the suitability of the Case Studies would rest with UNIDROIT.

76. Based on this, Annexes should include the following: Annex 1 covering legal aspects, such as harmonisation, the discourse on private and public law distinctions, and the scope of the Guide; Annex 2 addressing economics-related aspects, such as those initially contained in the footnotes; Annex 3 on *ex-post* evaluation; Annex 4 compiling academic and institutional references; and Annex 5 on Case Study, either included in the document or linked externally.

77. *It was agreed to develop two versions of the Guide: (i) a full version, similar to the current one and containing all references and material, and (ii) a second, shorter version functioning as a workflow tool. It was further agreed to develop the Annexes by designing a*

new Annex 1 with legal footnotes, while Annex 2 should include the economics-related footnotes. It was also suggested refining the latest Case Study and considering making it publicly available, either as Annex 5, or as a standalone document.

Item 10: Next steps

78. The main conclusions and next steps for the EE ICLR project include:

1) Updating the Preliminary Draft Guide as per below:

a) Framework and Quantification

- Explaining that quantification of economic outcomes relies on the *ceteris paribus* assumption.
- Specifying that the “division by 3” applies only when all three Factors are present. Where one Factor is absent, the Framework could be reformulated accordingly (e.g., division by 1 or 2).

b) Benchmark

- Providing further guidance on selecting a Benchmark, emphasising the importance of quantitative data over qualitative studies.
- Including examples in text boxes to illustrate how Benchmarks are selected.
- Clarifying that the search for Benchmarks with similar policy goals is relevant not only to Factor C but also to other Factors.
- Noting that the law reform under evaluation may be superior to the selected Benchmark, and that Benchmarks are chosen based on economic considerations, not as optimal legal standards.
- Clarifying that constructed Benchmarks are not created from scratch but are based on a combination of existing Benchmarks.
- Providing more detailed guidance on how to construct a Benchmark.

c) Degree of certainty

- Renaming this concept to “level of certainty”.
- Offering further guidance on how to assign it.
- Introducing a scoring methodology ((e.g., ranges or qualitative labels such as “very high,” “high,” “medium,” “low”) and explaining what each level implies.
- Explaining how the level of certainty accounts for the absence of best-in-class Benchmarks.
- Clarifying the treatment of externalities and their role in determining the level of certainty.

d) Evaluation actors

- Emphasising the importance of collaboration and synergies between evaluation teams.
- Outlining different forms of collaboration, including peer review, quality control, and external validation mechanisms.

e) “Context of Use”

- Changing the section title “Timing” to “Timing Aspects.”
- Conducting further research on the concepts of “facilitating law” and “promotive law” within this context.
- Reflecting on the distinction between the different *ex-ante* phases.

2) Further developing the Annexes:

- Removing most footnotes from the main text and including them in two separate annexes.
- Annex 1: Footnotes related to the legal context, such as on harmonisation.
- Annex 2: Footnotes related to economics.
- Case Study to be made available separately on the website (either as Annex 5, or as a standalone document).

3) Developing two versions of the Guide:

- **Full version:** Containing all references, footnotes, and acknowledgments.
- **Practical version:** A user-friendly electronic version including an executive summary and the workflow.

4) Case Studies:

- Considering whether Case Studies should be made publicly available under the authors' names, with a disclaimer for UNIDROIT.
- The Case Study on SMEs is noted as the most complete and practical for applying the Guide.

5) New Case Study:

- Exploring the possibility of creating a new Case Study based on an upcoming UNIDROIT instrument, for presentation at the Consultation Workshop, if feasible.

79. The next steps include launching a targeted Consultation for relevant stakeholders, organising a Consultation Workshop to discuss feedback received and test the application of the Guide, and publishing the Guide by the end of 2025.

80. *The CTCAP Directors thanked all the participants for their time and valuable contributions to the discussions, noting that the project was making significant progress and approaching its finalisation.*

ANNEX 1

**Economic Evaluation of
International Commercial Law Reform
project**



**11th Workshop on
Economic Evaluation of International Commercial Law Reform**

**Draft Agenda
23 January 2024**

*University of Cambridge
and online via Zoom
all times are in GMT*

Chairs: Professor Louise Gullifer, Professor Jeffrey Wool, Professor Ignacio Tirado (CTCAP Directors)

09:00 - 09:10	Opening of the Session <i>CTCAP Directors</i>
09:10 - 09:30	Summary of intersessional work, introduction to the revised Draft Guide and Case Study <i>UNIDROIT Secretariat and Prof. Jordi Paniagua</i>
09:30 - 11:15	Agenda Item No. 1: Discussion on the Benchmark, quantification and <i>ex-post</i> analysis (Annex 2)
11:15 - 11:30	Coffee break
11:30 - 12:15	Agenda Item No. 2: Discussion on the Framework and its scoring system
12:15 - 13:15	Agenda Item No. 3: Discussion on the Workflow and Evaluation Actors
13:15 - 14:00	Lunch break
14:00 - 15:00	Agenda Item No. 4: Discussion on the Factors
15:00 - 15:45	Agenda Item No. 5: Discussion on the Introduction and Annex 1
15:45 - 16:00	Coffee break
16:00 - 17:00	Agenda Item No. 6: Discussion on the contexts of use (timing and nature of law reform) and <i>ex-post</i> analysis (Annex 2)
17:00 - 17:30	Summary and next steps
17:30	Closing Remarks <i>CTCAP Directors</i>

List of participants

1. Jeffrey Wool (UNIDROIT Foundation, Aviation Working Group)
2. Louise Gullifer (Cambridge University)
3. Ignacio Tirado (UNIDROIT)
4. Andrew Myburgh (IFC)
5. Jordi Paniagua (University of Valencia, University of Notre Dame)
6. Mathias Siems (European University Institute)
7. Oren Sussman (University of Oxford)
8. Anna Veneziano (UNIDROIT)
9. Myrte Thijssen (UNIDROIT)
10. Theodora Kostoula (UNIDROIT)
11. Kateryna Bovsunovska (UNIDROIT)
12. Dogukan Kalinoglu (UNIDROIT, intern)