

# Article

## The Legitimacy of Final Statements and Reports of National Contact Points

An Empirical Analysis of (Final) Statements and Reports of the UK, US and Dutch National Contact Point of the Organisation for Economic Co-operation and Development (2001-2016)

Sander van 't Foort, Vivan IJzerman, Jasmin Lagziel and Tineke Lambooy\*

### 1 Introduction

Non-judicial grievance mechanisms play an important role in enhancing corporate social responsibility (CSR). Even when judicial mechanisms are effective, they cannot always carry the burden of addressing all alleged abuses. Moreover, judicial remedy is not always the favoured approach for the parties concerned.<sup>1</sup> One of the CSR instruments with a built-in non-judicial grievance mechanism is the Guidelines for Multinational Enterprises (MNE Guidelines or Guidelines) of the Organisation for Economic Co-operation and Development (OECD).<sup>2</sup> These Guidelines contain recommendations encouraging responsible business conduct and cover topics such as disclosure, human rights, employment

and industrial relations, environment, competition and taxation. Even though the Guidelines provide voluntary principles and standards,<sup>3</sup> both enterprises and their stakeholders consider the Guidelines as morally binding.<sup>4</sup> Whenever an MNE<sup>5</sup> allegedly does not observe the Guidelines, interested parties can appeal to a so-called National Contact Point (NCP) in order to address alleged misconduct.

The importance of non-judicial grievance mechanisms is stressed by the UN Guiding Principles on Business and Human Rights (Guiding Principles). Effective and well-functioning grievance mechanisms provide poten-

37

\* Nyenrode Business Universiteit. The authors would like to thank Prof. Dr. R. Jeurissen for his guidance and invaluable feedback on this article. The research was conducted from November 2014 until May 2016. In December 2016, the findings of this article were presented during the OECD's roundtable on 40 years of the OECD Guidelines for Multinational Enterprises at the OECD in Paris.

1. Ruggie J. (2011). Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises; Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework, UNHRC (A/HRC/17/31), pp. 33-34.
2. OECD (2012). Annual Report on the OECD Guidelines for Multinational Enterprises 2013: Responsible Business Conduct in Action, OECD Publishing. Retrieved from <<http://dx.doi.org/10.1787/mne-2013-en>>, p. 147.

3. OECD (2011). OECD Guidelines for Multinational Enterprises, OECD Publishing. Retrieved from: <<http://dx.doi.org/10.1787/9789264115415-en>>, p. 13.
4. Van Eyk S.C. (1995). *The OECD Declaration and Decisions Concerning Multinational Enterprises. An Attempt to Tame the Shrew*. Nijmegen: Ars Aequi Libri, pp. 121-122 and 135. The MNE Guidelines represent a 'firm expectation of MNE behaviour'. See OECD (1982). Mid-term report on the 1976 declaration and decisions, p. 57.
5. Strictly speaking, the scope of the specific instance is not restricted to international privately held large enterprises, but also includes state-owned enterprises and partly state-owned enterprises, small- to medium-sized enterprises and domestic enterprises (OECD (1979)). Report of the Committee on International Investment and Multinational Enterprises on the Review of the 1976 Declaration and Decisions on International Investment and Multinational Enterprises (C(79)102(final), par. 39; Rees C. & Vermijs D. (2008). *Mapping Grievance Mechanisms in the Business and Human Rights Arena*. Cambridge, MA: Corporate Social Responsibility Initiative, John F. Kennedy School of Government, Harvard University, p. 99; OECD (2000). Aide-Mémoire Information Meeting with Non-Members on the OECD Declaration for International Investment and Multinational Enterprises and the Guidelines for Multinational Enterprises, par. 11; OECD (2011), pp. 17-18).

tially affected parties access to remedy.<sup>6</sup> In order to enhance the effectiveness of non-judicial grievance mechanisms, the Guiding Principles contain several criteria that should be adhered to, among which is the criterion of ‘legitimacy’.<sup>7</sup>

NCPs have been criticised in the literature for their lack of legitimacy.<sup>8</sup> It is, however, difficult to determine what this criticism entails, since it is not specified why or how the legitimacy of NCPs or their decisions is lacking. The question of legitimacy is the key topic of this article, as the authors have assessed the legitimacy of the decisions of NCPs. Specifically, this article discusses to what extent the decisions of the UK NCP, US NCP and Dutch NCP can be regarded as legitimate.

Firstly, this article defines the relevant central concepts (Section 2), followed by a description of the methodology applied (Section 3) and an overview of the results of the empirical assessment of the decisions of UK NCP, Dutch NCP and US NCP (Section 4). In the penultimate section, the level of legitimacy of the decisions of the three NCPs will be compared (Section 5), and the final section contains conclusions and recommendations to possibly strengthen the legitimacy of the NCP decisions (Section 6).

## 2 Central Concepts

Before further elaborating on the concept of legitimacy and how this concept plays a role in the decisions of NCPs, this section briefly describes the central concepts of the study. This section describes NCPs and the MNE Guidelines (Section 2.1), NCP’s specific instances (Section 2.2) and NCP’s (final) statements and reports (Section 2.3).

### 2.1 NCPs and the MNE Guidelines

NCPs act within the framework of the MNE Guidelines, which were developed by the OECD 40 years ago.<sup>9</sup> All 34 OECD member countries and 13 non-member countries endorsed the MNE Guidelines and installed or are in the process of installing an NCP as a supervisory mechanism to ensure that the recommendations

contained in the MNE Guidelines are respected.<sup>10</sup> Whenever an MNE allegedly acts contrary to the recommendations contained in the MNE Guidelines, any interested person or party can notify an NCP to initiate a specific instance (*see* Section 2.2).

### 2.2 Specific Instances

Dealing with specific instances could be considered as the *modus operandi* of NCPs. A specific instance is the grievance mechanism that an NCP has put in place to discuss alleged non-observance of the MNE Guidelines with the parties involved. The main goal of the specific instance is to assist parties in resolving their conflicts in a non-judicial and consensual manner.<sup>11</sup> A specific instance comprises three phases: (i) initial assessment; (ii) good offices; and (iii) conclusion.<sup>12</sup>

During the first phase (‘initial assessment’) an NCP has to conduct an assessment on the basis of which it has to decide whether an issue ‘merits further examination’.<sup>13</sup> An NCP has to take into account various factors when conducting its initial assessment, for example, whether the issue is sufficiently substantiated. When an issue does merit further examination, phase two commences (‘good offices’). When an issue does not merit further examination, phase two will be skipped and the specific instances jumps to phase three (‘conclusion’). When phase two begins, an NCP offers its good offices to help resolve the issue at stake, for instance, by soliciting advice from relevant experts, consulting other NCPs and offering access to mediation or conciliation. At the final stage of the specific instance, the NCP publishes the results of the procedures after consulting the involved parties. The NCP can issue a (final) statement or a report depending on the specific circumstances of the case (*see* Section 2.3).<sup>14</sup>

### 2.3 Statements and Reports

An NCP can issue different types of decisions (*see* Figure 1). When a case does not merit further examination, a statement has to be issued containing at least the issues raised and the reasons for the NCP’s decision. A state-

6. Ruggie (2011), p. 23.

7. Ruggie (2011), pp. 26-27.

8. Lukas K., Plank L. & Staritz C. (2010). Securing Labour Rights in Global Production Networks, Legal Instruments and Policy Options, Vienna Chamber of Labour, p. 37. Cf. Robinson S. (2014). International Obligations, State Responsibility and Judicial Review Under the OECD Guidelines for Multinational Enterprises Regime. *Utrecht Journal of International and European Law*, 30(68), 75, <<http://dx.doi.org/10.5334/ujiel.cd>>. See for other critical remarks: The TUAC (2015). The OECD Guidelines for Multinational Enterprises Recommendations for Responsible Business Conduct in a Global Context. TUAC Submission to the Working Party on Responsible Business Conduct, par. 6; OECD Watch (2015). Remedy Remains Rare. An analysis of 15 years of NCP cases and their contribution to improve access to remedy for victims of corporate misconduct, p. 5.

9. Declaration on International Investment and Multinational Enterprises (C(76)99/FINAL).

10. OECD. National Contact Points. Retrieved from: <<http://mneguidelines.oecd.org/ncps/>> (lastly visited May 1, 2017).

11. OECD (2014). OECD Guidelines for Multinational Enterprises. Responsible Business Conduct Matters, p. 13.

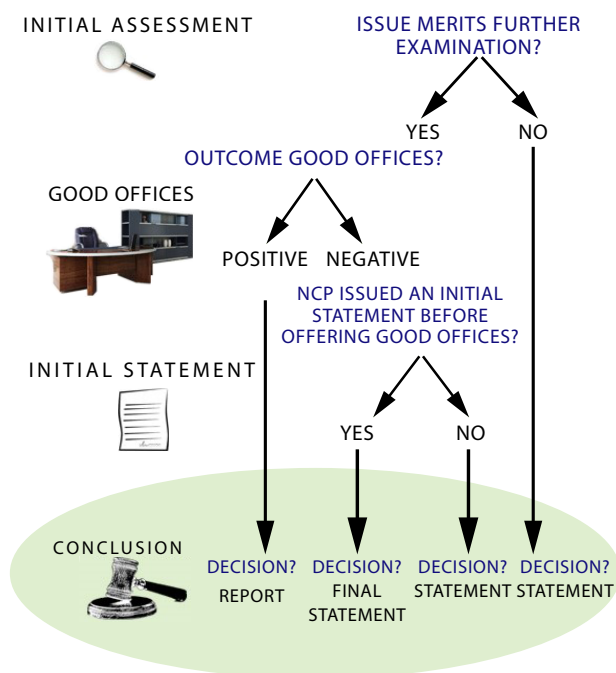
12. The TUAC discerns a fourth ‘examination phase’. During this phase an NCP examines if the MNE Guidelines were correctly observed. The ‘examination phase’ is not grounded on any official OECD document and will therefore be excluded from this section (The TUAC (2012). The OECD Guidelines for Multinational Enterprises. Recommendations for Responsible Business Conduct in a Global Context. Trade Union Guide, pp. 40 and 44). OECD Watch renames the last two phases as the ‘mediation phase’ and ‘final statement phase’. Both overlap for a great deal with the ‘good offices phase’ and ‘conclusion phase’. OECD Watch adds a ‘follow-up phase’ to the three existing phases. During the follow-up phase NCPs or other stakeholders must monitor the implementation of the decision (OECD Watch (2013). Calling for Corporate Accountability, A Guide to the 2011 OECD Guidelines for Multinational Enterprises, pp. 34 and 36).

13. In specific instances initiated by the TUAC only 22% of all NCPs publish their initial assessment (The TUAC, National Contact Point Comparison. Retrieved from: <[www.tuacoecdmguidelines.org/NCPcomparisonAll.asp](http://www.tuacoecdmguidelines.org/NCPcomparisonAll.asp)> (last visited 9 May 2016)).

14. OECD (2011), pp. 72-74 and 82-85.

ment is also issued when no agreement has been reached or when one of the parties is not prepared to partake in the specific instance. In these cases a statement can include recommendations on the implementation of the MNE Guidelines and must at least describe the parties involved, the dates on which the issues were raised and the issues themselves, the reasons for meriting further examination and the processes initiated by the NCP. Reasons for the absence of any agreement can also be included, where appropriate. If parties have reached an agreement, a report will be issued, which describes the issues raised, the processes initiated by the NCP, when an agreement was reached and the results of the proceedings. Information on the content of the agreement may also be included with consent of the involved parties. It is also possible to issue a statement at the end of the initial assessment ('initial statement').<sup>15</sup> Consequently, two statements can be issued during one specific instance procedure: an initial statement stating that the issue merits further examination and a statement about the end result of the specific instance, which could be either positive ('report') or negative ('final statement').<sup>16</sup> Only in the latter case the initial statement renders a statement into a final statement (see Figure 1). A report preceded by an initial statement remains a report.

Figure 1 Types of NCP decisions<sup>17</sup>



15. In practice, only a few NCPs publish an initial statement (Schliemann C. (2012). Procedural Rules for the Implementation of the OECD Guidelines for Multinational Enterprises – A Public International Law Perspective, *German Law Journal*, 13(01), 71).

16. OECD (2011), pp. 73 and 84-85.

17. In the 'anonymous v. international manufacturer' decision, an agreement was reached between parties before the initial assessment was concluded. This decision does not fit in Figure 1, but still has been classified as a report since the outcome was positive.

### 3 Theoretical Framework

This section describes the theoretical basis on which this research is grounded. Section 3.1 defines the concept of legitimacy. Sections 3.2 and 3.3 describe how the concept of legitimacy relates to NCP decisions and explain how it is operationalised in various factors and indicators.

#### 3.1 The Concept of Legitimacy

Legitimacy is not a clear-cut concept.<sup>18</sup> It has multiple definitions depending on its object. For example, the legitimacy of a decision of an NCP may differ from the legitimacy of an NCP as an institution. The context in which the concept of legitimacy is used may differ as well.

Within political sciences, Weber defines legitimacy as: 'a citizen[?]s willingness to comply with a system of rule, out of not selfishness, expedience, or habit, but rather a considered belief in the moral validity of that rule'.<sup>19</sup> In legal sciences, Weston defines legitimacy as

a property of a rule or a rule-making institution which itself exerts a pull toward compliance on those addressed [...] because those addressed believe that the rule or institution has come into being and operates in accordance with generally accepted principles of right process.<sup>20</sup>

Another legal scholar, Rawls, argues that legitimacy is vested in the exercise of power in accordance with a constitution. Rawls states that the:

exercise of [...] power is [...] justifiable only when it is exercised in accordance with a constitution [,] the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational.<sup>21</sup>

Each of these definitions may be useful within their own context, but as such have not been applied within the context of NCPs.

For want of any definition about legitimate (final) statements and reports, the concept of legitimacy is translated to the context of (final) statements and reports by using the three aforementioned definitions and also Guiding Principle 31 of the Protect, Respect and Remedy Framework (better known as the 'Ruggie Framework') (see Section 3.2). In this research the legitimacy of (final) statements and reports is defined as

18. Malsch M. (2008). Lay elements in the criminal justice system of the Netherlands. In J. Shapland (Ed.), *Justice, Community and Civil Society. A Contested Terrain*, Devon: Willan Publishing, pp. 108-109.

19. Gilley B. (2009). *The Right to Rule: How States Win and Lose Legitimacy*. New York: Columbia University Press, p. 3.

20. Weston B.H. (1991). Security Council Resolution 678 and Persian Gulf Decision Making: Precarious Legitimacy. *The American Journal of International Law*, 85(3), 516.

21. Rawls J. (2005). *Political Liberalism* (Expanded edition). New York: Columbia University Press, p. 217.

‘a stakeholder’s obedience of and trust in (final) statements and/or reports that wield power in accordance with the MNE Guidelines’.

A (final) statement’s and report’s ability to deserve obedience is derived from Weston’s ‘pull towards compliance’ and Weber’s ‘willingness to comply’. When applying Rawls’ definition to (final) statements and reports, wielding power in accordance with the MNE Guidelines resonates with ‘exercising power in accordance with a constitution’, whereby the MNE Guidelines can be considered as ‘the constitution’ on which (final) statements are grounded. The ability to enable trust from stakeholders affected by a decision is derived from Guiding Principle 31.

### 3.2 NCPs and Their Relation to the Concept of Legitimacy in the Context of the Guiding Principles

In 2011, the United Nations Human Rights Council endorsed the Ruggie Framework. This normative framework on business and human rights was established by the former UN Special Representative of the Secretary-General on Human Rights and Transnational Corporations and Other Business Enterprises, Ruggie. The framework consists of three pillars: (i) the state’s duty to protect human rights; (ii) the corporate’s responsibility to respect human rights; and (iii) access to remedy. These three pillars are operationalised through 31 Guiding Principles. Guiding Principle 31 sets forth several effectiveness criteria for non-judicial grievance mechanisms. Although these criteria were designed for human rights-oriented non-judicial grievance mechanisms, Ruggie suggested to incorporate these criteria in the 2011 update of the MNE Guidelines to apply to specific instances and NCPs, which have a broader scope than just human rights.<sup>22</sup> One of the factors that may determine the effectiveness of non-judicial grievance mechanisms, such as specific instances, is ‘legitimacy’.<sup>23</sup> According to Principle 31, legitimate non-judicial grievance mechanisms, ‘[should enable] trust from the stakeholder groups for whose use they are intended’ and ‘[should account] for the fair conduct of grievance processes’.<sup>24</sup> Neither the Ruggie Framework nor the MNE Guidelines<sup>25</sup> further define this concept, or provide for concrete and practical tools on how to effectuate legitimate grievance mechanisms.

22. Except to the criterion ‘a source of continuous learning’ (Ruggie J. (2010). 10th OECD Roundtable on Corporate Social Responsibility Updating the Guidelines for Multinational Enterprises, Discussion Paper, pp. 5-6).

23. Besides legitimate, non-judicial grievance mechanisms should – according to the Ruggie Framework – be accessible, predictable, equitable, transparent, rights-compatible and a source of continuous learning (Ruggie (2011), p. 33).

24. Ruggie (2011), p. 26.

25. Not all criteria, as suggested by Ruggie, were incorporated into the MNE Guidelines. ‘Legitimacy’ and ‘a source of continuous learning’ have been left out in the MNE Guidelines (Ruggie (2011), p. 26).

### 3.3 Operationalisation of the Concept of Legitimacy

Due to a lack of literature specifically pertaining to the legitimacy of (final) statements or reports, this article turns to literature about the legitimacy of court decisions and decisions made by alternative dispute resolution (ADR)<sup>26</sup> institutions. Literature about the legitimacy of court decisions may be of particular relevance, because final statements are said to be analytically similar to court decisions.<sup>27</sup> As NCPs are ADR institutions offering mediation and conciliation services, literature about the legitimacy of decisions made by ADR institutions may also be relevant for NCP decisions.

Based on literature about the legitimacy of court decisions and decisions made by ADR institutions, a selection of factors was made and an initial analysis and trail coding encompassing 74 decisions of the UK NCP, Dutch NCP and US NCP was performed.<sup>28</sup> The initial analysis was conducted to rule out factors that were virtually always met<sup>29</sup> or could not be operationalised in a manner that ensured their validity and/or reliability.<sup>30</sup> As a result, multiple factors were discarded, merged and added.<sup>31</sup> Five factors remained and were used to evaluate all decisions of the UK NCP, Dutch NCP and US NCP.

Each factor was operationalised by one or more indicators. Each indicator could be either met or not met. If all indicators were met, then a decision would receive the highest rating possible. If no indicator was met, then a decision would receive the lowest rating possible. We assessed each (final) statement and report on the basis of the indicators. The assessment was initially conducted independently of each other. Afterwards, we came together to discuss the results and reached consensus, in case of disagreement, on the final rating.

The broad concept of legitimacy is operationalised by using certain factors. Although many different factors can be used to measure legitimacy, we have selected the following five factors to apply to the decisions of NCPs:

26. ADR can be defined as ‘any method of resolving disputes other than by litigation’. ADR includes ‘early neutral evaluation, negotiation, conciliation, mediation and arbitration’ (Legal Information Institute, Alternative Dispute Resolution, legal dictionary, Cornell University Law School. Retrieved from <www.law.cornell.edu/wex/alternative\_dispute\_resolution> (last visited 28 December 2016).

27. Van Doren W., Marx A. & Wouters J. (2014). Dispute Settlement as a Regulatory Intermediary: the Case of National Contact Points under the OECD Guidelines for Multinational Enterprises, Leuven Centre for Global Governance Studies University of Leuven, p. 22.

28. The complete methodology of the initial analysis is on file with the first author and can be shared upon request.

29. The usage of narratives, a decision’s internal consistency (*i.e.* the consistency within one decision), ‘adherence’ (*i.e.* whether a (final) statement and report are issued by a competent authority) and decision’s support (*i.e.* dissenting or concurring opinions contained in a (final) statement or report) was almost always met.

30. The results of the initial analysis were scrutinised by all four authors as well as the following three professors: Prof. Dr. P.M. Langbroek (Utrecht University), Prof. Dr. R. Nieuwenkamp (University of Amsterdam), Prof. Dr. R. Jeurissen (Nyenrode Business Universiteit). Dr. J. Wilde-Ramsing (SOMO) also provided feedback.

31. A comprehensive overview of all methodological steps taken is on file with the authors of this article and can be shared upon request.

(i) completeness; (ii) comprehensibility; (iii) structure; (iv) justification; and (v) impartiality. In the literature, these factors emerged as key factors of the legitimacy of decisions and we regard these factors as best suited to objectively analyse (final) statements and reports.

The first factor, completeness, was selected since this is one of the factors that is often taken into account in empirical studies of court decisions.<sup>32</sup> Furthermore, the MNE Guidelines distinguish several elements that must be included in a (final) statement and report<sup>33</sup> and indicate that value is attributed to the completeness of decisions.<sup>34</sup>

The second factor, comprehensibility, was selected since this factor is also often taken into account in empirical studies of court decisions.<sup>35</sup> Additionally, research shows that readable, therefore comprehensible, decisions are regarded as more persuasive and prestigious than less readable alternatives.<sup>36</sup>

The third factor, structure, has proven to positively affect the legitimacy of decisions.<sup>37</sup> Langbroek *et al.* distinguish in their research on judicial decisions certain characterisations of decisions that we have used to evaluate an NCP decision's structure.<sup>38</sup>

The fourth factor, justification of a decision, is, in line with the arguments of former president of the Dutch Supreme Court, Corstens, an essential factor for legitimacy. Corstens considered that different considerations must be taken into account and weighed for a decision to be legitimate.<sup>39</sup> Other scholars have also acknowledged the importance of reasoning for the legitimacy of court decisions.<sup>40</sup> Empirical research conducted amongst 700 participants recorded how participants responded to reasoning applied by arbitrators and judges in their decisions. The researchers found that decisions that were accompanied by multiple reasons received the highest legitimacy scores.<sup>41</sup>

The fifth and last factor, impartiality, determines whether an NCP grounds its decision on reasons, which

reflect both sides of the dispute. Research found that non-biased decisions received a higher legitimacy rating than biased decisions.<sup>42</sup> Likewise, Corstens argues that courts may not have any bias.<sup>43</sup>

Annex 1 contains an elaboration of the five factors. Each factor is linked to various indicators that can help to measure the factors. The list of indicators used is not exhaustive; other indicators can also support the same factors. However, we selected these indicators because they are based on literature and the MNE Guidelines and are concrete enough to be measured. Table 1 enumerates all factors and indicators. A complete account of the indicators, including their explanation and sources, can be found in Annex I.

## 4 Methodology

This section describes the methodology applied in this research. The method of analysis (Section 4.1), the sample selection (Section 4.2) and the reliability and validity of this research (Section 4.3) will be discussed consecutively.

### 4.1 Method of Data Analysis

In order to answer the main research question, a content analysis was performed. In content analysis, texts are analysed within their context through analytical constructs (*see* Table 1 for the analytical construct of this research). An answer to the research question is reached through inferences drawn from texts after analysing their contents. These inferences are warranted by the application of an analytical construct (*see* Table 1) that is grounded in literature, plus the assurance that the analysis has been performed reliably (*see* Section 4.3 for inter-rater reliability).<sup>44</sup>

### 4.2 Sample Selection

In order to limit the scope of the research, this article confines the investigation to the legitimacy of the decisions of NCPs in specific instance procedures (*see* Section 2.2). All publicly available decisions of three NCPs were analysed, 82 in total.<sup>45</sup> All decisions were published between December 2002 and May 2016. The UK NCP and US NCP have by far dealt with the most cases of all NCPs (approximately 50 cases each). The Dutch NCP dealt with considerably less cases (23 cases), but remains one of the most active and leading NCPs in the

32. See Loth M.A. (2009). Courts in quest for legitimacy: a comparative approach. In N. Huls, M. Adams & J. Bornhoff (Eds.), *The Legitimacy of Highest Courts' Rulings: Judicial Deliberations and Beyond*. The Hague: T.C.M. Asser Press, pp. 267-288 and Malsch M., Efstatiades C. & Nijboer H. (2005). Justification of court decisions in criminal cases: continental Western European countries compared, Report NSCR 2005-5, pp. 9-20. We requested and received a hard copy of the report from M. Malsch.
33. See, e.g. the issues raised, the date on which the issues were raised with the NCP and recommendations on the implementation of the MNE Guidelines as appropriate (OECD (2011), p. 73).
34. OECD (2011), p. 73.
35. Malsch *et al* (2005), pp. 9-20.
36. Murphy Romig J. (2012). Improving Legal Writing – Quantifiably. *Georgia Bar Journal*, 18(1), 64.
37. Malsch *et al* (2005), pp. 9-20.
38. Langbroek Ph., *et al* (2014). Een onderzoek naar de tekstkenmerken en de publieken van rechterlijke motiveringen, Montagne centrum, p. 113.
39. Corstens G. (2013). Objective law and subjective judges, Conference of the International Association of Tax judges, pp. 6-9.
40. Hol A.M. (2009). Internationalisation and legitimacy of decisions by the highest courts. In A.S. Muller & M.A. Roth (Eds.), *Highest Courts and the Internationalization of Law: Challenges and Changes*, The Hague: Hague Academic Press, p. 78.
41. Simon D. & Scurich N. (2011). Lay Judgments of Judicial Decision Making. *Journal of Empirical Legal Studies*, 8(4), pp. 709-727.

42. Simon & Scurich (2011), pp. 709-727.

43. Corstens (2013), pp. 6-9.

44. Krippendorf K. (2004). *Content Analysis. An Introduction to Its Methodology*. London: Sage, pp. 29-31 and 38.

45. UK NCP (44 decisions), Dutch NCP (19 decisions) and US NCP (19 decisions).

Table 1 Operationalisation of the concept of legitimacy in factors and indicators

Factor	Indicator
<b>Completeness of the decision</b>	The decision includes the provisions of the MNE Guidelines.
	The decision includes the allegations raised.
	The decision includes the response of the MNE on the allegations.
	The decision includes information about the parties.
	The decision includes the date on which the issues were first raised with the NCP.
	In case of a report: the decision includes the content of the agreement reached or the results of the agreement.
	In case of a (final) statement: the decision includes (a) recommendation(s).
	In case of a (final) statement: the decision includes (a) determination(s) about adherence to the MNE Guidelines.
<b>Comprehensibility of the decision</b>	Use of NCP jargon.
	The decision explains the provisions of the MNE Guidelines.
	Readability/text complexity.
<b>Decision's structure</b>	The decision does not repeat sentences.
	In case of a report: procedure, facts, considerations of the parties and the results of the negotiations are distinguishable.
	In case of a final statement: procedure, facts, considerations of the parties, reasons of the NCP are distinguishable.
<b>Justification of the decision</b>	The NCP weighs the different considerations involved.
	The NCP accompanies its decision with multiple reasons.
	The NCP includes supporting facts for the decision.
	The NCP puts forwards supporting evidence for the decision.
<b>Impartiality</b>	In case of a final statement: the NCP grounds its decision on reasons, which reflect both sides of the dispute.

42

world.<sup>46</sup> All three NCPs were selected because they issued the most decisions compared to the other NCPs, making it possible to conduct an in-depth content analysis of their decisions. Taken together, the three NCPs almost represent one-third of all decisions made by NCPs. These were the only three that produced a considerable number of decisions, which could be used to conduct a quantitative content analysis. An additional argument for selecting the Dutch NCP and the UK NCP is the level of compatibility of their decisions with legal verdicts.<sup>47</sup> Literature about the legitimacy of legal verdicts may thus apply to the (final) statements and reports of the Dutch NCP and UK NCP.

46. The Dutch NCP is one of the most active and leading NCPs, when taking into account the number of cases dealt with, its active involvement in peer reviews and its innovative institutional structure. See for instance: Accountability Counsel (2013). A Case Study of the Dutch NCP, p. 1; *Kamerstukken II 2010/11*, 26 485, nr. 101, 8 ('Letter from the State Secretary of Economic Affairs, Agriculture and Innovation'); *Kamerstukken II 2013/14*, 26 485, nr. 190, p. 3.

47. Van Doren *et al* (2014), p. 22.

### 4.3 Reliability and Validity

To increase the reliability of this research, only publicly available documents are used. All (final) statements and reports that are analysed can be found in online repositories or were retrieved from the websites of the three NCPs.<sup>48</sup> The reliability of this research is also increased by operationalising each legitimacy factor into one or more indicators. In this way, possible interpretation differences between researchers are avoided as much as possible.

The inter-rater reliability, that is, the extent to which the ratings of the three researchers coincided,<sup>49</sup> was cal-

48. Three databases were used: the OECD database. Retrieved from: <<https://mneguidelines.oecd.org/database/>>; the database of OECD Watch. Retrieved from: <<http://oecdwatch.org/cases>>; and the database of the TUAC. Retrieved from: <[www.tuacoecd.mneguidelines.org/cases.asp](http://www.tuacoecd.mneguidelines.org/cases.asp)>. The websites of the three NCPs can be found here: OECD, National Contact Points. Retrieved from: <<https://mneguidelines.oecd.org/npcs/>> (last visited 1 May 2017).

49. Gwet K.L. (2014). *Handbook of Inter-rater Reliability. The Definitive Guide to Measuring the Extent of Agreement Among Raters*. Gaithersburg: Advanced Analytics, p. 4.

culated by applying Krippendorff's alpha (KALPHA). KALPHA was selected, because it is conceived as the golden standard for computing the inter-rater reliability in content analysis.<sup>50</sup> KALPHA is relatively conservative, meaning that high scores are not easily reached when compared to other instruments of measurement.<sup>51</sup> A KALPHA of 0.00 signifies complete absence of reliability, whereas a KALPHA of 1.00 signifies perfect reliability.<sup>52</sup> A value of 0.80 or higher is regarded as a 'good norm' for reliability.<sup>53</sup> In this research, the computed KALPHA has a value of 0.82.<sup>54</sup>

The internal validity<sup>55</sup> of this research is strengthened through theory triangulation, that is, different theoretical lenses<sup>56</sup> and an extensive body of literature were used, as well as researchers triangulation, *viz.* the complete analysis was conducted by three researchers.<sup>57</sup> The most prevalent threat to the internal validity of this research design is that literature about the legitimacy of court decisions and decisions of ADR institutions is used to assess the legitimacy of (final) statements and reports. Although final statements are said to be analytically similar to court decisions,<sup>58</sup> and NCPs are ADR institutions, no empirical research has established a correlation or causal relationship between court or ADR decisions and NCP decisions proving their similarity. Future rigorous empirical research will be needed to establish which factors contribute to the legitimacy of (final) statements and reports. However, this research helps to signal possible factors to test their legitimacy. We note that the list of factors is not exhaustive and no statements are made as to whether one factor may have a greater impact on the legitimacy of decisions than another. Hence, no ranking amongst the factors is applied.

The external validity is high, because no intention exists to generalise the research results so that they apply to NCPs other than the UK NCP, Dutch NCP and US NCP. Generalisability within the UK NCP and Dutch NCP seems less problematic, as almost all of their decisions were published and analysed. In case of the UK

NCP, five<sup>59</sup> out of 49 decisions were not publicly available. In case of the Dutch NCP, four<sup>60</sup> out of 23 decisions were not publicly available. The generalisability of the US NCP's decisions is more problematic, because all 29 decisions<sup>61</sup> that were made before 2011 were not published. The conclusions reached regarding the US NCP are therefore restricted to decisions after 2010.

A final remark regards the indicators that were taken into account. The content analysis of NCP decisions was limited to the appraisal of the various indicators. Other issues that may have affected the results, such as parallel procedures, the institutional position of the NCP and the availability of (financial) resources, were not taken into account.

## 5 Results

This section presents the results of the content analysis of the decisions of the UK NCP (Section 4.1), Dutch NCP (Section 4.2) and US NCP (Section 4.3). Each section contains a separate discussion of reports and (final) statements. A separate discussion of (final) statements and reports is warranted, since no argumentation is applied in reports so that the factors 'justification of the decision' and 'impartiality' could not be taken into account when assessing reports. When interpreting the graphs in the following sections, we note that the MNE Guidelines were lastly revised in 2011. Low scores before the last revision may therefore not be attributable to the NCPs, since not all indicators were included in previous editions of the MNE Guidelines.

### 5.1 UK NCP

#### 5.1.1 Final Statements

Figure 2 displays the scores of 31 (final) statements of the UK NCP during the period 2001–2016. Over the

50. De Swert K. (2012). Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha, University of Amsterdam, p. 1; Hayes A.F. & Krippendorff K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, 1(1), 88.
51. Zhao X., Liu J.S. & Deng K. (2013). Assumptions behind intercoder reliability indices. In C.T. Salmon (Eds.), *Communication Yearbook* 36. Routledge: New York (pre-published paper), pp. 59-61 and 81-83.
52. Hayes & Krippendorff (2007), p. 82.
53. De Swert (2012), p. 5.
54. KALPHA was computed by applying the KALPHA macro that is specially designed for SPSS. Calculations were verified by using the following online tool: <<http://158.182.41.59/reliability/index.jsp?firstInitFlag=0>>. The percentage of agreement amongst the three raters was 86.6 per cent.
55. De Vaus D. (2009). *Research Design in Social Research*. London: Sage, pp. 27-28; Saunders M., Lewis P. & Thornhill A. (2012). *Research Methods for Business Students*. Edinburgh Gate: Pearson Education Limited, p. 193.
56. For instance, when defining the concept of legitimacy (Section 3.1).
57. Cf. Lee N. & Lings I. (2013). *Doing Business Research. A Guide to Theory and Practice*, Los Angeles: Sage, p. 239.
58. Van Doren *et al* (2014), p. 22.

59. (Final) statements and reports of the following cases are not publicly available: RAID v. Binani (withdrawn); RAID v. Alex Stewart (Assayers) Ltd (withdrawn); RAID v. Tremalt (withdrawn); RAID v. Ridgepoint (withdrawn); and Imerys v. Transport and General Workers Union (OECD Watch, RAID v. Binani. Retrieved from: <[http://oecdwatch.org/cases/Case\\_18](http://oecdwatch.org/cases/Case_18)>; OECD Watch, RAID v. Alex Stewart (Assayers) Ltd. Retrieved from: <[http://oecdwatch.org/cases/Case\\_43](http://oecdwatch.org/cases/Case_43)>; OECD Watch, RAID v. Tremalt Retrieved from: <[http://oecdwatch.org/cases/Case\\_42](http://oecdwatch.org/cases/Case_42)>; OECD Watch, RAID v. Ridgepoint Retrieved from: <[http://oecdwatch.org/cases/Case\\_44-en](http://oecdwatch.org/cases/Case_44-en)>; The TUAC, Imerys v. Transport and General Workers Union (T&G) Retrieved from: <[www.tuacoedmneguidelines.org/CaseDescription.asp?id=52](http://www.tuacoedmneguidelines.org/CaseDescription.asp?id=52)> (last visited 9 May, 2016)).
60. (Final) statements and reports of the following cases are not publicly available: Wärtsilä v. FNV; Hewlett Packard v. FNV; Smead Europe v. FNV; and Ryanair v. FNV (The TUAC, Wärtsilä v. FNV. Retrieved from: <[www.tuacoedmneguidelines.org/CaseDescription.asp?id=86](http://www.tuacoedmneguidelines.org/CaseDescription.asp?id=86)>; The TUAC, Hewlett Packard v. FNV. Retrieved from: <[www.tuacoedmneguidelines.org/CaseDescription.asp?id=76](http://www.tuacoedmneguidelines.org/CaseDescription.asp?id=76)>; The TUAC, Smead Europe v. FNV. Retrieved from: <[www.tuacoedmneguidelines.org/CaseDescription.asp?id=101](http://www.tuacoedmneguidelines.org/CaseDescription.asp?id=101)>; The TUAC, Ryanair v. FNV and FNV Bondgenoten. Retrieved from: <[www.tuacoedmneguidelines.org/CaseDescription.asp?id=102](http://www.tuacoedmneguidelines.org/CaseDescription.asp?id=102)> (last visited 9 May 2016)).
61. A list of unpublished decisions is available upon request.

Figure 2 Level of legitimacy (final) statements for UK NCP

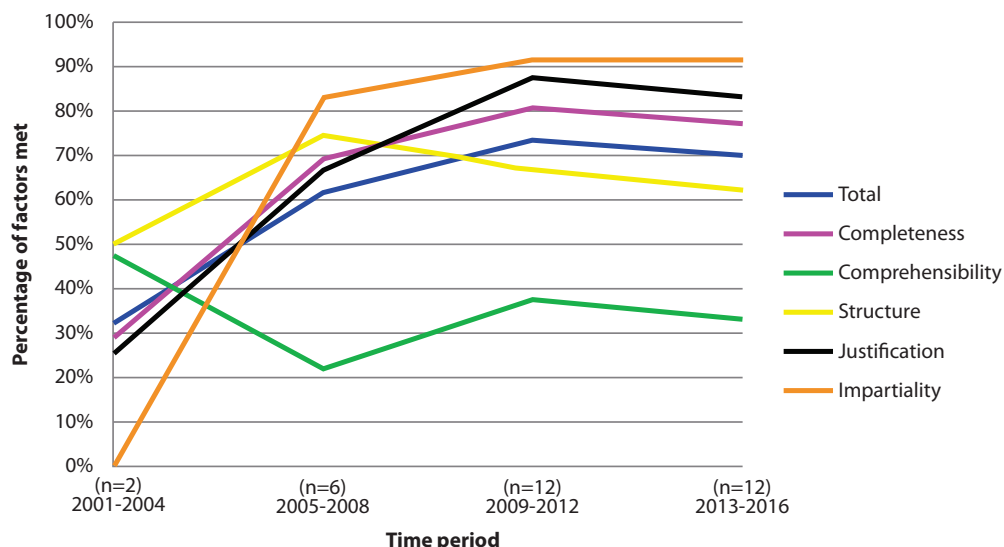
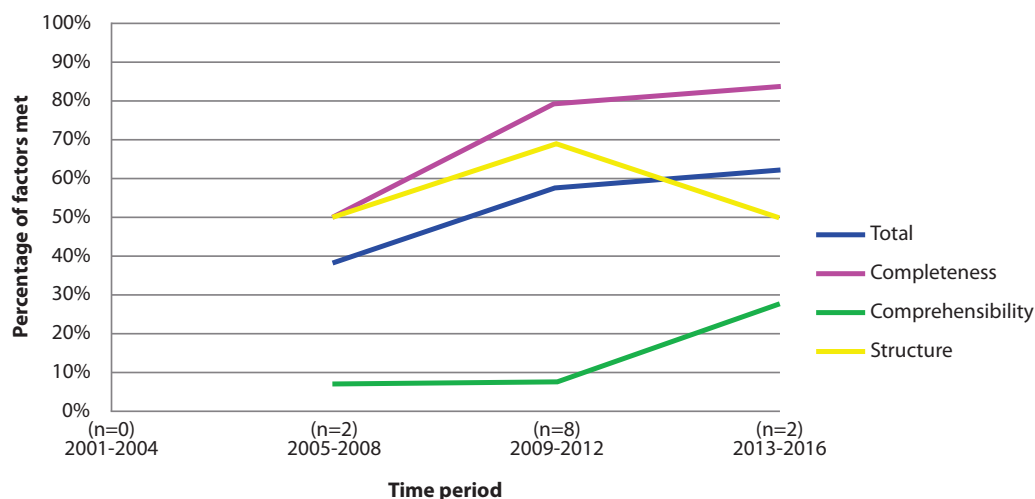


Figure 3 Level of legitimacy reports for UK NCP



years, the UK NCP increased and stabilised its scores on almost all factors.

The sole exception concerns the comprehensibility of the UK NCP’s (final) statements. This factor has always received a score lower than 50%. The comprehensibility scores during the first period (2001–2004) were higher when compared to later periods, chiefly because in all cases (in the period 2001–2004) the NCP did not use any jargon and hence received a full score on this indicator.<sup>62</sup>

During the ensuing years, jargon was used,<sup>63</sup> but only explained in seven cases. For example, NCP jargon such as ‘specific instance’ and ‘due diligence’ were often used without any further explanation.<sup>64</sup> During 2005–2008, the low comprehensibility scores were caused by the lack of explanation of jargon and were compounded by a very low number of cases wherein the NCP explained

the provisions of the MNE Guidelines.<sup>65</sup> Lower readability scores indicate that the decisions also have become more difficult to read. The three indicators taken together explain why comprehensibility scores plummeted during 2005–2008.

The lowest as well as highest scores reached by the UK NCP pertain to the level of impartiality. At the very outset, the UK NCP’s final decision reflected the position of the MNE and not that of the other party.<sup>66</sup> As a result, according to the selected indicators, complete impartiality could not be ensured. As the years progressed, the UK NCP developed the habit to include the arguments of both sides of the dispute in its final decision. In some cases, the UK NCP decided in favour of the MNE, but still accepted a number of arguments of the other party.<sup>67</sup> In other cases, the NCP did not decide in favour of the MNE, but still acknowledged the actions the MNE took to mitigate its adverse impacts.<sup>68</sup>

62. UN Expert Panel v. De Beers (2004); UN Expert Panel v. Avient (2004).

63. One exception is CBE v. National Grid Transco (2005).

64. See, e.g. in International trade union v. UK registered multinational (2010) and RAID ACIDH v. ENRC (2016).

65. Only one case was recorded (Amicus and T&G v. PSA Peugeot Citroen (2008)).

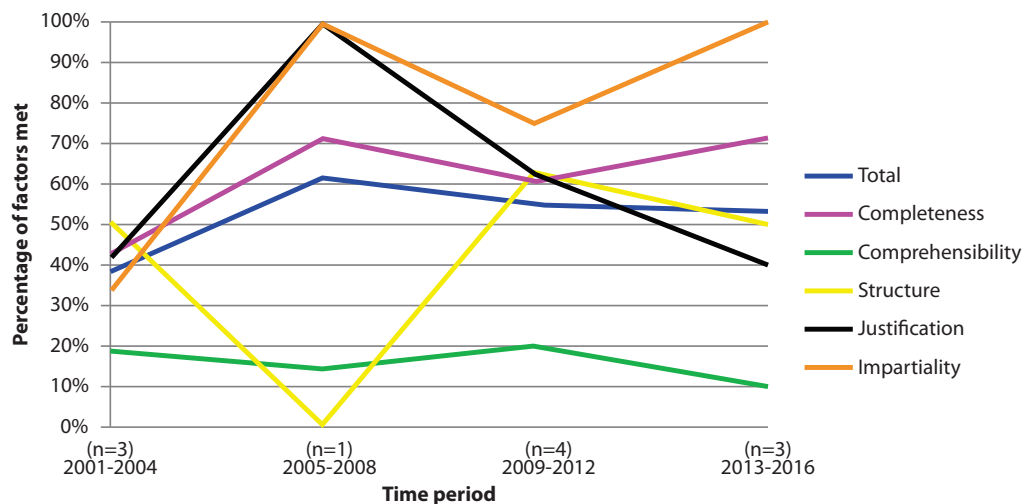
66. UN Expert Panel v. De Beers (2004); UN Expert Panel v. Avient (2004).

67. For example in Individual in India v. UK Company UAE (2013).

68. For example in LPHR v. G4S (2015).



Figure 4 Level of legitimacy (final) statements for Dutch NCP



In both cases the NCP received a full impartiality score, because it reflected on the arguments of both sides of the dispute in its decision.

The completeness, structure and justification scores remained rather stable during the last two periods. The slightly declining completeness trend can be explained by a lower amount of cases wherein the UK NCP includes information about the parties involved and the response of the MNE on the allegations raised.<sup>69</sup> An increase in the number of cases wherein facts were not distinguishable led to a lower structure score, even though the UK NCP started to include a heading called ‘fact finding’ from 2015.<sup>70</sup> During the last period, the UK NCP to a lesser extent used evidence to justify its decision or just mentioned that it ‘used evidence’ without specifying its sources.<sup>71</sup> Taken together with a lower number of cases wherein the different considerations involved were weighed, the lack of supporting evidence led to a lower justification score.

### 5.1.2 Reports

In Figure 3 the first period is left blank, because no reports were published during this period. The first report of the UK NCP was published in 2007.<sup>72</sup> From this moment onwards, the completeness and comprehensibility scores of the UK NCP have increased, while the structure scores have decreased. The total lack of explanation of the MNE Guidelines in all decisions contributed to the low comprehensibility scores as well as the usage of jargon without any explanation. In only one case all jargon was explained.<sup>73</sup> The lower scores for the decision’s structure can be attributed to the decreased level of distinguishability between procedure, facts, considerations of the parties and the results of the negotiations. Especially the considerations of the parties and

facts were not distinguishable.<sup>74</sup> The UK NCP could have received a full completeness score, if it had included information about the parties involved during the last period.

## 5.2 Dutch NCP

### 5.2.1 Final Statements

A total of 10 publicly available (final) statements were issued by the Dutch NCP. Figure 4 shows the level of legitimacy. Over the years, the Dutch NCP increased and stabilised its scores on almost all factors. However, a sudden change in the numbers occurred in the period of 2005–2008. Since only one decision<sup>75</sup> was delivered by the NCP during that period, the sudden increase and decrease of certain factors has to be put into perspective. With regard to the factor structure, the main concern was the distinguishability of the facts and the procedure. The Dutch NCP introduced a new layout in 2012, which, although appealing, did not improve the actual structure (the graph of the [final] statements actually shows a decrease in structure from 2012 onwards). With regard to completeness, this factor gradually improved. Noticeable is that the Dutch NCP often neglected to include the response of the MNE on the allegations; this improved slightly over the years.<sup>76</sup>

Another matter is the comprehensibility of the Dutch NCP’s decisions. The Dutch NCP scored relatively low on this factor when compared to the other factors. For instance, the Dutch NCP did not meet the indicator ‘use of NCP jargon’ in any of its statements. This means that in all of the decisions NCP jargon was used but not explained. Especially the words ‘specific instance’ and ‘investment nexus’ re-occurred over the years without explanation. Also the indicator ‘explanation of the pro-

69. For example in *NGO v. six UK-based telecommunication companies* (2014) and *RAID and ACIDH v. ENRC* (2016).

70. *LPHR v. G4S* (2015).

71. For example in *Individual in India v. UK Company UAE* (2013).

72. *Anonymous v. international manufacturer* (2007).

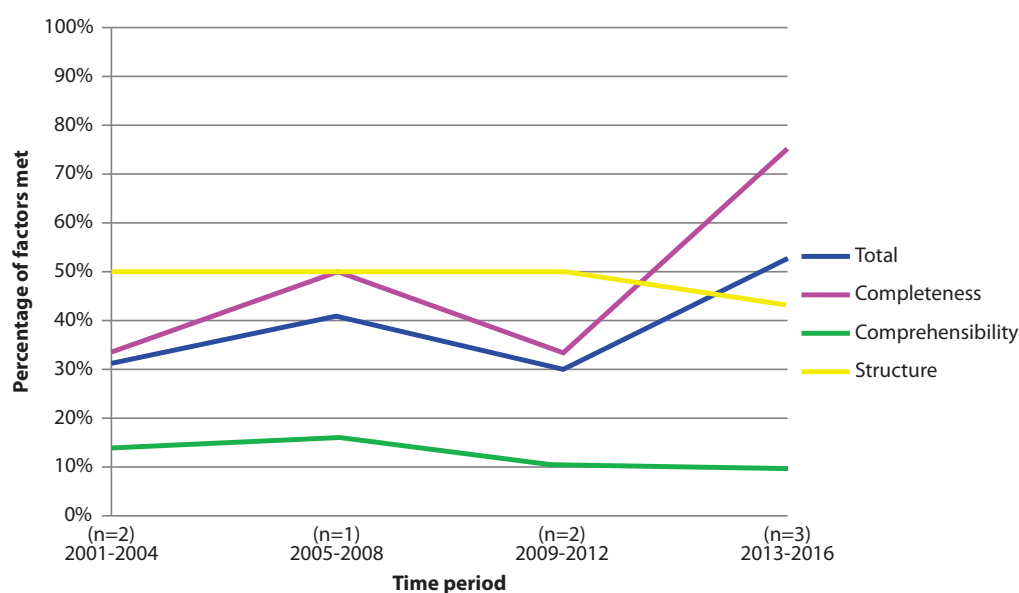
73. *WWF v. SOCO* (2015).

74. For example in *ECCHR v. ICT Cotton* (2011) and *The LEAD Group Inc. v. Xstrata Plc.* (2012).

75. *Plaid v. FNV* (2005).

76. Only in the cases *Plaid v. FNV* (2005), *SHV v. Shehri-CBE* (2010), *Shell v. Friends of the Earth (FoE) and Amnesty International (Amnesty Int.) I* (2013), *Shell v. FoE Amnesty Int. II* (2014) the Dutch NCP included the response of the MNE.

Figure 5 Level of legitimacy reports for Dutch NCP



visions of the MNE Guidelines' was only met twice.<sup>77</sup> In these cases, the NCP did not explain the provisions of the MNE Guidelines, which were brought forwards by the parties. The readability scores of the Dutch NCP's (final) statements have varied between 0.27 and 0.44. These scores indicate that all decisions were difficult or very difficult to read and best understood by people with a university degree.

Overall, the level of impartiality has increased the most. The highest scores reached by the Dutch NCP were on this factor. Increasingly, the Dutch NCP has managed to ground its decision on reasons that reflect both sides of the dispute. The factor impartiality was met in all of the decisions from 2012 onwards. With regard to the factor justification, when compared to the NCP's first decisions, the Dutch NCP improved the level of justification considerably in the period 2005–2008. Nevertheless, as stated earlier, only one decision was delivered by the NCP during that period. Overall, the level of justification stayed relatively the same over the years. In six cases<sup>78</sup> the NCP did not support its decision with multiple arguments.

### 5.2.2 Reports

Nine of the 19 publicly available decisions delivered by the Dutch NCP were reports. A certain pattern emerges when examining the factor completeness: the level of completeness has increased noticeably over the years. From the beginning and in almost all of the reports<sup>79</sup> the content of the agreement reached or the results of the agreement were included. Conversely, in the latest decision, the 2016 Mylan case, this factor was not met since the report only stated that 'the parties agreed to

continue a constructive dialogue on this matter'.<sup>80</sup> As with the (final) statements, the biggest issue with the factor structure is the distinguishability of the facts and the procedure. With regard to the comprehensibility of the Dutch NCP's decisions, the reports scored lower than the (final) statements, since the reports never explained the provisions of the MNE Guidelines.

## 5.3 US NCP

### 5.3.1 Final Statements

A total of 17 publicly available (final) statements were issued by the US NCP. Figure 6 displays the scores of the (final) statements of the US NCP during the period 2011–2016. When looking at this graph, no general trend or pattern emerges with regard to the development of the level of legitimacy of the factors.

The US NCP's highest score is on the factor completeness (55%). However, as the graph shows, no improvement on this factor occurred over the years. Only in three of the 17 cases, the US NCP gave insight in the provisions of the MNE Guidelines that were brought forwards by the parties.<sup>81</sup> This score is also influenced by the many cases in which no information is provided about the parties.

It furthermore appears that the US NCP frequently neglects to make recommendations to the parties. Although the US NCP explicitly mentions in its decisions that, where appropriate, it makes recommendations as to how the enterprise might make its business practices more consistent with the MNE Guidelines, they only did so in five cases.<sup>82</sup> The US NCP also

77. ICN v. Kubbinga (2003) and Shell Philippines v. NGOs (2009).

78. ICN v. Kubbinga (2003); Touroperators v. FNV/CNV (2004); Shell Philippines v. NGOs (2009); Dutch company v. Local community (2010); Shell v. FoE Amnesty Int. I (2013); Shell v. FoE Amnesty Int. II (2014).

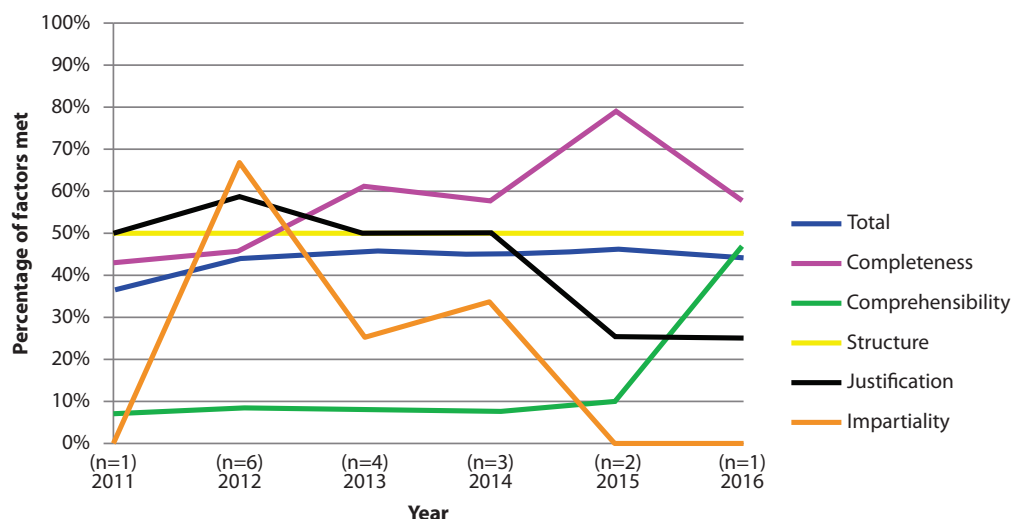
79. Not in the cases Nidera v. SOMO et al. (2012) and Mylan v. Stapert (2016).

80. Mylan v. Stapert (2016), p. 4.

81. IUF v. Mondelez International (2014); IUF v. Pepsi Co. (2014); UAW and IndustriALL v. Nissan North America Inc. (2015).

82. CLEC and ERI v. American Sugar Refining Inc. (2013); IUF v. Mondelez International (2013); International Union v. Crown Holdings (2015); UAW and IndustriALL v. Nissan North America Inc. (2015); IUF v. PepsiCo II (2016).

Figure 6 Level of legitimacy (final) statements for US NCP



neglected to make a determination on the adherence of the MNE to the MNE Guidelines in nearly all cases.<sup>83</sup> Firstly, the US NCP did not make a determination in case it offered its good offices, but the parties nonetheless failed to reach an agreement, or when one of the parties refused to cooperate.<sup>84</sup> Secondly, the US NCP refused to offer its good offices in several cases after it found that it would not be appropriate to offer its good offices<sup>85</sup> or when its involvement would not further the effectiveness of the Guidelines.<sup>86</sup> In these cases, the US NCP did not make a determination as well, and consequently left the parties empty-handed. The reason why the US NCP does not make a determination in case parties fail to reach an agreement is because the US NCP states in its decisions that its role is ‘to provide a neutral, third-party facilitated dialogue, and not make a determination whether a violation of the MNE Guidelines has occurred, nor to adjudicate disputes submitted under the process’.<sup>87</sup>

With a score of 35%, the US NCP’s lowest score is on impartiality. Strikingly, this factor further decreased over the past 5 years. This is caused by the fact that in most of the cases the position of the non-governmental organisation (NGO) and/or trade union is not reflected in the decision. Consequently, the US NCP did not ground its decision on reasons that reflect both sides of the dispute.<sup>88</sup>

The justification of the decision scores 49%. No weighing of the various considerations has taken place in 15

cases.<sup>89</sup> This is, among other things, caused by the fact that in several cases one of the parties did not want to engage in the negotiations. As a result, the US NCP ended the specific instance procedure and did not give a determination on the case.<sup>90</sup> Consequently, no weighing of the different considerations has taken place. Moreover, the US NCP does not give multiple reasons, which caused a downfall in the score. Just like the factor impartiality, this factor therefore declined over the years.

With regard to the comprehensibility of the US NCP’s decisions, the US NCP did not meet the indicator ‘use of NCP jargon’ and ‘explanation of the provisions of the MNE Guidelines’ in any of its statements. In all of its decisions, NCP jargon was used but not explained. Furthermore, the US NCP did only explain the provisions that were brought forwards by the parties in one of its cases.<sup>91</sup> The average text complexity was 0.25. This means that decisions were very difficult to read.

The only score that remained constant over the years is that of the factor structure (50%/55%). This lack of improvement is caused by the fact that in none of these cases the different aspects of the decisions were distinguishable. Specifically, the facts and considerations of the parties were not mentioned by the US NCP.<sup>92</sup>

### 5.3.2 Reports

Two reports of the US NCP are publicly available. Despite these low numbers, the outcomes could give an indication about the level of legitimacy of the US NCP reports (see Figure 7). It must be noted that, contrary to the US NCP’s (final) statements, the factor completeness increased over the years. Whereas the US NCP first did not include the allegations raised, the information about the parties and the content of the agreement

83. One notable exception is *Edouard Teumagnie v. AES Corporation* (2012).

84. *CLEC and ERI v. American Sugar Refining Inc.* (2013).

85. *ILRF v. Dole* (2011).

86. *Edouard Teumagnie v. AES Corporation* (2012) and *CWA, ver.di, UNI Global Union v. Deutsche Telekom AG* (2013).

87. *The Lead Group v. Innospec* (2012), *CWA, ver.di, UNI Global Union v. Deutsche Telekom* (2013) and more recently *CED v. Herakles* (2015).

88. See, e.g. *IUF v. PepsiCo II* (2016), *UAW-IndustriALL v. Nissan North America Inc.* (2015), *IUF v. PepsiCo Inc. (India) I* (2014), and *IUF v. Mondelez International* (2013).

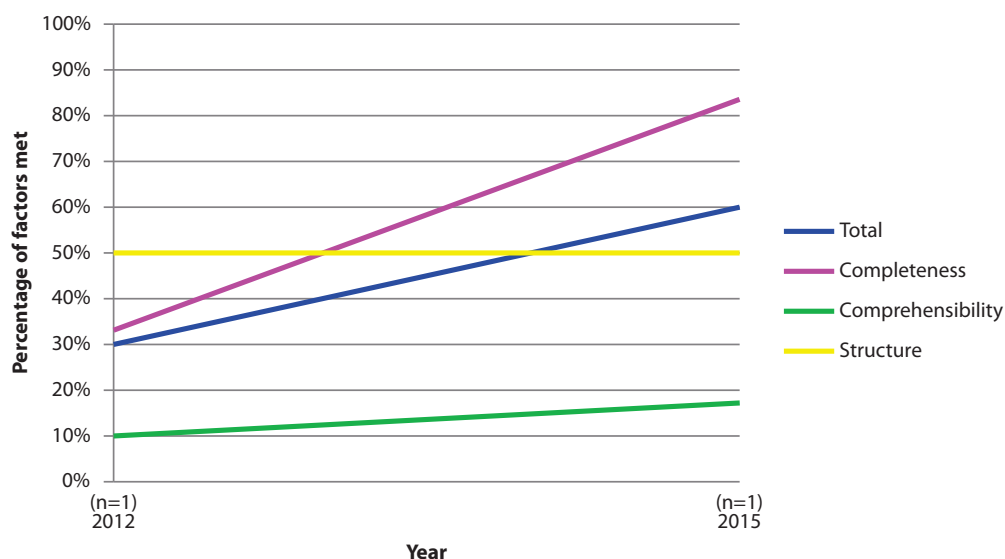
89. Weighing of the considerations did only take place in *IUF v. Mondelez International* (2014) and *FFW v. Johnson & Johnson (Philippines) Inc.* (2013).

90. *CLEC and ERI v. American Sugar Refining Inc.* (2013); *IUF v. Mondelez International* (2013); *IUF v. Mondelez International* (2014).

91. *IUF v. PepsiCo II* (2016).

92. *Individual A cases* (2012).

Figure 7 Level of legitimacy reports for US NCP



reached, they did do so several years later.<sup>93</sup> The factor comprehensibility showed a small increase during the past five years as well. The factor 'structure' remained constant. This has to do with the fact that the different aspects of the reports (procedure, facts, considerations of the parties and the results of the negotiations) were not distinguishable.

## 48 6 Comparative Analysis UK, US and Dutch NCPs

After analysing the 82 decisions of the UK NCP, US NCP and Dutch NCP, it has become apparent that, within the framework of our factors, the UK NCP is the best performing NCP. Figures 8 and 9 display the cumulative legitimacy scores of the three NCPs' (final) statements and reports (see total scores). As the graphs show, the differences between the three NCPs are most evident in (final) statements. With regard to the (final) statements, the UK NCP received the highest rating on all factors, except for impartiality. The UK NCP is also the NCP that improved the most, starting lower than the other two NCPs on almost all points, but ending with higher scores. Especially on the factor justification the UK NCP scored very high. Because the UK NCP received the highest legitimacy scores on almost all factors, the (final) statements and reports of the UK NCP can be regarded as the most legitimate of the three NCPs. The Dutch NCP maintains a position between the UK NCP and US NCP. The US NCP scored lowest on our scale and their decisions must therefore be considered as the least legitimate when compared to the UK NCP and Dutch NCP. Especially on the factors completeness and impartiality the US NCP received a lower rating.

93. UNITE HERE v. LSG Sky Chefs (2012) and CED v. Herakles Farms (2015).

A comparison between the three NCPs shows that the UK NCP's and Dutch NCP's understanding of the role of an NCP differs from that of the US NCP. The US NCP – in contrast to the UK NCP and Dutch NCP – considers its main role is to assist affected parties in their efforts to reach a satisfactory and consensual resolution to the issues raised under the MNE Guidelines. In line with this role, the US NCP does not make a determination whether a violation of the MNE Guidelines has occurred.<sup>94</sup> According to the US NCP, its function is not to make a determination whether a violation of the MNE Guidelines has occurred and it is of the opinion that it does not have legal authority to adjudicate disputes submitted under the specific instance procedure.<sup>95</sup> The UK NCP and Dutch NCP have chosen to make a determination in most of their (final) statements. The UK NCP made a determination in 26 out of 32 (final) statements and the Dutch NCP made a determination in six out of 11 (final) statements. Even though the US NCP states that it does not make any determinations, it still made one determination in the past (see Section 4.3.1).<sup>96</sup> This discrepancy in determinations between the three NCPs resulted in lower completeness and justification scores for the US NCP. The low completeness score of the US NCP was also intensified by the low frequency of recommendations, which, as previously stated, is an important task of an NCP as it improves the legitimacy of an NCP and is enshrined in the MNE guidelines.<sup>97</sup>

For (final) statements as well as reports the comprehensibility scores of each NCP are the lowest of all factors. The low comprehensibility scores signify the low amount of attention paid to explanation. When comparing the different trend graphs of the previous sections

94. See, e.g. The LEAD Group v. Innospec (2012); UNITE HERE v. LSG Sky Chefs (2012).

95. See, e.g. The LEAD Group v. Innospec (2012); UNITE HERE v. LSG Sky Chefs (2012).

96. Edouard Teumagnie v. AES Corporation (2012).

97. OECD (2011), pp. 72-73.

Figure 8 Comparison level of legitimacy (final) statements for UK NCP, Dutch NCP and US NCP (2011–2016)

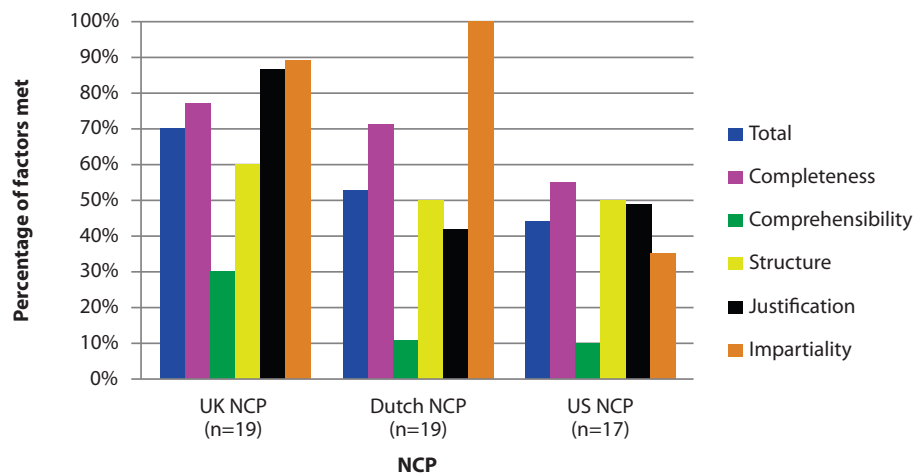
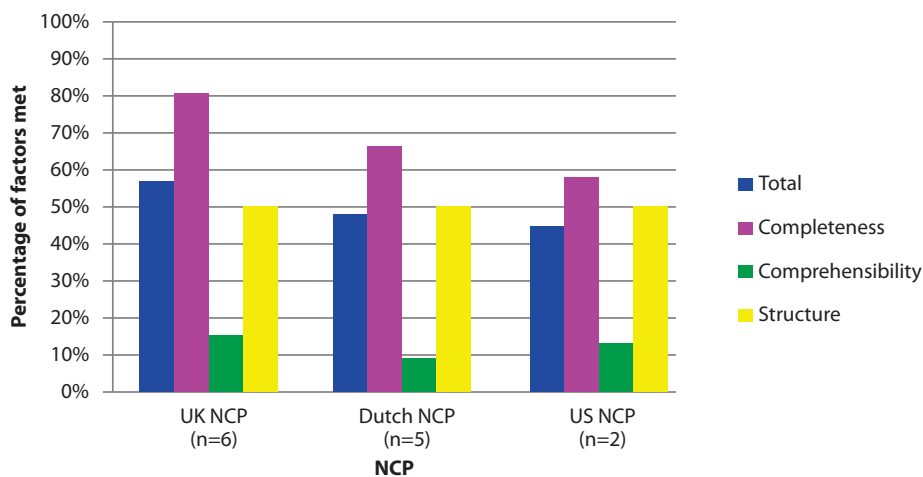


Figure 9 Comparison level of legitimacy reports for UK NCP, Dutch NCP and US NCP (2011–2016)



(see Sections 4.1–4.3), it can be observed that the (final) statements of the UK NCP and Dutch NCP do not show any progress in the level of comprehensibility, while the (final) statements of the US NCP show a sudden increase in the last two years. The highest readability score reached was 0.52 for both (final) statements<sup>98</sup> and reports.<sup>99</sup> This score indicates that decisions were fairly difficult to read. In total, 26 decisions received a readability score of lower than 0.30, which means that these decisions are very difficult to read and can only be well understood by people with a university degree.

## 7 Conclusion and Recommendations

This research analysed the level of legitimacy of (final) statements and reports of the UK NCP, Dutch NCP and US NCPs using five factors: completeness, comprehensibility, structure, justification and impartiality. A total of 82 publicly available decisions were analysed by applying these factors, after which it could be concluded

that the UK NCP is the best performing NCP and can be regarded as delivering the most legitimate (final) statements and reports. The decisions of the Dutch NCP received a lower legitimacy score and the decisions of the US NCP the lowest. Based on these findings, it could be concluded that the US NCP's decisions are the least legitimate. The difference in legitimacy scores between the three NCPs is most evident in (final) statements.

As has become clear from the previous sections, comprehensibility forms the Achilles' heel of NCP decisions. The best-rated decisions were still fairly difficult to read and oftentimes could only be understood by university graduates. By shortening sentences and using words with fewer syllables, the readability score could be further improved. Besides improving the readability of a decision, the NCPs could include an explanation of the relevant provisions of the MNE Guidelines in their decision. A short passage from the commentaries of the MNE Guidelines may be sufficient. The commentaries of the MNE Guidelines explain how a certain provision must be interpreted and may improve the comprehensibility of the decision when included. Jargon can be dealt with in two ways: either by explaining jargon or by omitting it completely. Typical NCP jargons such as 'specific instance' and 'final statement' need to be

98. RAID v. Oryx (2005).

99. Greenpeace v. Herakles Farms/Capital (2014).

explained within the NCP context for the reader to understand the decision. The UK NCP can serve as a best practice here. In every decision, the UK NCP incorporates a section called 'UK NCP Procedure' in which it provides the necessary context and explanation for NCP jargon.

Especially for the US NCP, with the lowest completeness scores, the inclusion of the provisions of the MNE Guidelines and information about the parties may be quick wins that may considerably increase the level of legitimacy. Simply stating that 'an NGO' filed a complaint against 'an MNE' does not suffice. More information about, for instance, the mission of the NGO or history of the company may be of use to the reader. The inclusion of recommendations may also improve the completeness of decisions and may further the effectiveness of the MNE Guidelines when recommendations aim to stimulate responsible business conduct.

Procedure, facts, considerations of the parties and reasons of the NCP were not always distinguishable in decisions. NCPs can improve their decision's structure by introducing headings and separate sections. Some of the latest UK NCP's decisions serve as a good example here.<sup>100</sup> These decisions comprised a section on fact finding and on the NCP procedure, adding to the distinguishability of facts and procedures. Considerations of the parties were often found under the header 'complaint' and 'company's response' and the reasons of the UK NCP were mostly covered by the 'UK NCP conclusions' section. Other NCPs can use the UK NCP's decisions as a reference to improve their own decisions' structure.

The gap between the impartiality scores of the US NCP on the one hand and the UK NCP and Dutch NCP on the other is rather large: the impartiality scores of the UK NCP and Dutch NCP are approximately twice as high as the score of the US NCP. The US NCP can bridge this gap by ensuring that its decisions do reflect the arguments concerning both sides of the dispute. The decisions of the US NCP are too often dominated by the position of the MNE. Impartiality scores could be improved by also taking the arguments of the other party, such as NGOs and trade unions, into consideration. This could be achieved by, for example, arguing why it diverts from the arguments of the other party.

A final recommendation specifically addresses the US NCP. As shown earlier, the US NCP does not make any determination on whether the MNE Guidelines are violated or not. The US NCP could possibly improve the legitimacy of its decisions, when it changes this policy by including determinations. Including determinations may increase the level of completeness and justification and consequently the level of legitimacy of its decisions. In general, NCPs may not default on their obligations and not making a determination may lead to situations wherein complainants are left empty-handed. For instance, in cases that the MNE does not want to engage in mediation and the specific instance is closed as a

result, making a determination may be an interesting avenue to explore in order to further the effectiveness of the MNE Guidelines.<sup>101</sup>

100. See, e.g. LPHR v. G4S (2015) and RAID ACIDH v. ENRC (2016).

101. Van 't Foort S. & Wilde-Ramsing J. (2015). A comparative analysis of the Dutch specific instance procedure, *Nederlands-Vlaams tijdschrift voor Mediation en conflictmanagement*, Aflevering 4, pp. 24-25.

Annex I Operationalisation Legitimate (Final) Statements and Reports

Factor	Indicator	Measurement	Source
Completeness of the decision	The decision includes the provisions of the MNE Guidelines.	This indicator is met when the provisions of the MNE Guidelines that were brought forwards by the parties are included in the decision. A referral to a provision does not suffice.	In an empirical comparative study of 91 court decisions of Western European countries, the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) examined the legitimacy decisions established using a panel of 31 participants. Completeness was one of the factors that was taken into account during the examination. The NSCR argued that decisions need to be 'self-contained', but did not explain what a 'self-contained' decision exactly entails. <sup>102</sup> Hence, we decided to use other sources to operationalise the completeness of a decision. The first indicator (i.e. 'the decision includes the provisions of the MNE Guidelines') is derived from the initial analysis of NCP decisions.
	The decision includes the allegations raised.	This indicator is met when the allegations of the parties are included in the decision. Allegations dismissed during an initial assessment do not have to be included.	The MNE Guidelines stipulate that statements and reports should include the 'issues raised'. <sup>103</sup> The 'issues raised' have been divided into the allegations raised and the MNE's response on the allegations, since both parties are able to raise issues.
	The decision includes the response of the MNE on the allegations.	This indicator is met when the response of the MNE on the allegations is included in the decision.	The MNE Guidelines stipulate that statements and reports should include the 'issues raised'. <sup>104</sup> The 'issues raised' have been divided into the allegations raised and the MNE's response on the allegations, since both parties are able to raise issues.
	The decision includes information about the parties.	This indicator is met when information, such as number of employees, company history, type of sector or affiliated parties, is included in the decision. Mentioning that the party is an MNE, trade union or NGO without any further information does not suffice.	The MNE Guidelines stipulate that (final) statements should 'identify the parties concerned'. <sup>105</sup> We argue that this indicator is equally important for reports in order to identify the parties that reached an agreement.
	The decision includes the date on which the issues were first raised with the NCP.	This indicator is met when the decision includes the date on which the issues were raised with the NCP.	The MNE Guidelines stipulate that (final) statements and reports should include 'the date on which the issues were raised with the NCP'. <sup>106</sup>
	In case of a report: the decision includes the content of the agreement reached or the results of the agreement.	This indicator is met when the decision includes the content of the agreement reached. Whenever confidentiality inhibits the publication of the agreement, or parties agree not to disclose the agreement, the NCP must at least publish the results of the agreement.	The MNE Guidelines stipulate that 'information on the content of the agreement will only be included insofar as the parties involved agree thereto'. <sup>107</sup> Presenting the results without parties' approval is not mandatory. Nevertheless, we argue that without divulging the results of the agreement it is not possible to obtain a complete view of the decision.

(Continued)

Factor	Indicator	Measurement	Source
	In case of a (final) statement: the decision includes (a) recommendation(s).	This indicator is met when the decision includes (a) recommendation(s).	The MNE Guidelines stipulate that (final) statements should include 'recommendations on the implementation of the Guidelines as appropriate'. <sup>108</sup> Making a judgement about the appropriateness of recommendations is a task of an NCP and for this reason we decided to solely judge whether recommendations are included or not.
	In case of a (final) statement: the decision includes (a) determination(s) about adherence to the MNE Guidelines	This indicator is met when the decision includes (a) determination(s).	Despite the fact that a determination is not mandatory according to the MNE Guidelines, OECD Watch argues that 'in situations where no agreement is reached through mediation, NCPs should make a determination on whether the company in question has complied with the Guidelines'. <sup>109</sup> OECD Watch's assertion was based on all NCP cases over the past 15 years. The OECD has aired similar views in its stocktaking report of NCPs. <sup>110</sup> We therefore agreed to include this indicator.
Comprehensibility of the decision	Use of NCP jargon	This indicator is met when jargon is used and explained or when no jargon is used at all. This indicator is not met when jargon is used and not explained, that is, when the following words are not explained: 'initial assessment', 'specific instance', 'good offices', 'final statement', 'initial statement', '(human rights) due diligence' and 'investment nexus'. Jargon mentioned by the parties and not the NCP as well as jargon mentioned in a cited provision of the MNE Guidelines does not have to be explained in order to meet this indicator.	When examining the legitimacy of court decisions, the aforementioned NSCR took into account the comprehensibility of the decision and showed that comprehensibility positively affected the legitimacy of decisions. According to the NSCR, jargon may lead to confusion and adds to the overall complexity of a decision. <sup>111</sup> Hence, the need for explanation.
	The decision explains the provisions of the MNE Guidelines.	This indicator is met when provisions of the MNE Guidelines that are brought forwards by the parties are explained. This indicator is not met when one or more provisions are included, but not explained. Copy-paste of the MNE Guidelines is not accepted as a sufficient explanation.	According to the NSCR including an explanation of provisions in a decision positively contributes to the legitimacy of that decision. <sup>112</sup>
	Readability/text complexity	Application of the Flesch–Kincaid readability formula for text complexity	Research showed that 'taking readability statistics into account is valuable, because legal writing that is easier to read is more comprehensible and may also be perceived by readers as more persuasive, more prestigious and just more likeable than less-readable alternatives'. <sup>113</sup> A method for judging writing objectively is to measure it by using readability statistics, including the 'Flesch Reading Ease' score. <sup>114</sup>



(Continued)

Factor	Indicator	Measurement	Source
Decision's structure	The decision does not repeat sentences.	If one or more sentences are literally repeated, this indicator will not be met. Repetition in a summary or conclusion is allowed.	The NSCR found that decisions with a 'clear structure' contribute to the legitimacy of decisions. Repetitiveness seems to adversely impact the legitimacy of decisions. <sup>115</sup>
	In case of a report: procedure, facts, considerations of the parties and the results of the negotiations are distinguishable.	This indicator is met when the aforementioned aspects of the decision are distinguishable. The usage of headers may enhance the distinguishability.	In their research on the justification of court decisions, Langbroek et al. distinguished between the procedure, the facts, the considerations of the parties and the considerations of the court. <sup>116</sup>
	In case of a final statement: procedure, facts, considerations of the parties, reasons of the NCP are distinguishable.	This indicator is met when the aforementioned aspects of the decision are distinguishable. The usage of headers may enhance the distinguishability.	In their research on the justification of court decisions, Langbroek et al. distinguished between the procedure, the facts, the considerations of the parties and the considerations of the court. <sup>117</sup>
Justification of the decision	The NCP weighs the different considerations involved.	This indicator is met when the various considerations ( <i>i.e.</i> the allegations and the other party's response) are taken into account and weighed. The indicator is not met when the NCP refrains from one of the two.	In line with the arguments of former president of the Dutch Supreme Court, Corstens, the different considerations must be taken into account and weighed for a decision to be legitimate. <sup>118</sup>
	The NCP accompanies its decision with multiple reasons.	This indicator is met when the NCP supports its decision with multiple arguments. This indicator is not met when the NCP provides no reasons at all or one single supporting reason. When the decision of the NCP can be divided into multiple decisions and at least one of the decisions is supported by multiple reasons, this indicator will be met.	Scholars acknowledge the importance of reasoning for the legitimacy of court decisions. <sup>119</sup> Empirical research conducted amongst 700 participants recorded how participants responded to four strings of reasoning applied by arbitrators and judges in their decisions. The researchers found that decisions that were accompanied by multiple reasons received the highest legitimacy scores. <sup>120</sup>
	The NCP includes supporting facts for the decision.	This indicator is met when facts are stated to justify the NCP's conclusion. This indicator is not met whenever no facts are used for a decision's justification.	In line with Corstens, supporting facts must be taken into account when assessing the legitimacy of decisions. <sup>121</sup>

*(Continued)*

Factor	Indicator	Measurement	Source
	The NCP puts forwards supporting evidence for the decision.	This indicator identifies whether evidence is put forwards to justify the NCP's decision. Letters, submissions of the advisory bodies of the OECD, policy documents, newspaper articles, emails, notes, court rulings, results from field visits, laws and regulations, corporate policies, evaluation reports, websites, contracts, annual reports, financial statements, affidavits are all examples of evidence. This indicator is not met when no sources are stated and is met when at least one source is provided.	The NSCR found that the amount of evidence provided can contribute to the legitimacy of the decision, but refrain from giving any examples. <sup>122</sup> The investigated sources of evidence in this research are therefore derived from the initial analysis of NCP decisions.
Impartiality	In case of a final statement: the NCP grounds its decision on reasons which reflect both sides of the dispute.	This indicator is met when an NCP grounds its decision on reasons which reflect both sides of the dispute.	Corstens argues that courts may not have any bias. <sup>123</sup> In the aforementioned empirical research amongst 700 respondents, the researchers found that non-biased decisions received a higher legitimacy rating than biased decisions. <sup>124</sup>

<sup>102</sup> Loth M.A. (2009), pp. 267-288; Malsch *et al* (2005), pp. 9-20. We requested and received a hard copy of the report from M. Malsch.

<sup>103</sup> OECD (2011), p. 73.

<sup>104</sup> *Ibid.*

<sup>105</sup> OECD (2011), p. 85.

<sup>106</sup> OECD (2011), p. 73.

<sup>107</sup> *Ibid.*

<sup>108</sup> *Ibid.*

<sup>109</sup> OECD Watch (2015), p. 44.

<sup>110</sup> OECD (2016). Implementing the OECD Guidelines for Multinational Enterprises: The National Contact Points from 2000 to 2015, p. 29.

<sup>111</sup> Malsch *et al* (2005), pp. 9-20.

<sup>112</sup> *Ibid.*

<sup>113</sup> Murphy Romig (2012), p. 64.

<sup>114</sup> *Ibid.*

<sup>115</sup> Malsch *et al* (2005), pp. 9-20.

<sup>116</sup> Langbroek *et al* (2014), p. 113.

<sup>117</sup> *Ibid.*

<sup>118</sup> Corstens (2013), pp. 6-9.

<sup>119</sup> Hol (2009), p. 78.

<sup>120</sup> Simon & Scurich (2011), pp. 709-727.

<sup>121</sup> Corstens G. (2013), pp. 6-9.

<sup>122</sup> Malsch *et al* (2005), pp. 9-20.

<sup>123</sup> Corstens (2013), pp. 6-9.

<sup>124</sup> Simon & Scurich (2011), pp. 709-727.