



Relevance and Coherence as Measures of Quality in Forensic Psychological Reports

João da Silva Guerreiro, Dianne Casoni & Jorge Costa Santos

To cite this article: João da Silva Guerreiro, Dianne Casoni & Jorge Costa Santos (2014) Relevance and Coherence as Measures of Quality in Forensic Psychological Reports, *Psychiatry, Psychology and Law*, 21:6, 890-902, DOI: [10.1080/13218719.2014.918077](https://doi.org/10.1080/13218719.2014.918077)

To link to this article: <https://doi.org/10.1080/13218719.2014.918077>



Published online: 24 Jul 2014.



[Submit your article to this journal](#)



Article views: 166



[View Crossmark data](#)

Relevance and Coherence as Measures of Quality in Forensic Psychological Reports

João da Silva Guerreiro^{a,c}, Dianne Casoni^a and Jorge Costa Santos^{b,c}

^a*Université de Montréal, École de criminologie, Montreal, Canada;* ^b*Universidade de Lisboa, Faculdade de Medicina, Lisbon, Portugal;* ^c*Centro de Ciências Forenses, Coimbra, Portugal*

The goal of this research is to provide a general portrait of forensic psychological reports under the Portuguese justice system. An analysis of 106 reports in terms of their relevance and coherence was conducted since these two dimensions appear to be key to better understanding some of the specific characteristics related to the overall quality of forensic psychological reports. A grid was constructed to identify elements related to these two dimensions and its application show that, while mostly meeting formal characteristics such as adequate organization, most of the reports surveyed fell short of meeting the criteria defining relevance and coherence. Results suggest that research on the quality of forensic psychological reports should direct more attention to the internal coherence of the reports and to the importance of reporting on assessment findings that regard the uniqueness of assessees in view of optimizing reports' role as informational support for legal decision-making.

Key words: coherence; forensic psychological report; qualitative analysis; quality; relevance.

The last two decades have given rise to a number of interesting studies concerning forensic psychological reports. Many of these address ethical issues or questions related to professional standards (Conroy, 2006; Grisso 2010; Heilbrun, 2001; Heilbrun & Collins, 1995; Lander & Heilbrun, 2009; Michaels, 2006; Nguyen, Acklin, Fuger, Gowensmith, & Ignacio, 2011). These studies have led authors to a level of consensus regarding what characterizes a high quality forensic psychological report, which is determined, in large part, by the formal characteristics such reports present (Wettstein, 2005, 2010; Nicholson & Norwood, 2000). As a consequence, a number of guidelines have been suggested for forensic report writing in psychology (Ackerman, 2006; APA, 2013;

Grisso, 2010; Heilbrun, 2001; Lander & Heilbrun, 2009; Witt, 2010). These guidelines include: an optimal organization of the content, quality of writing and of the language used, depth of the information conveyed, and quality of the links established between the source material and its interpretation (Nicholson & Norwood, 2000; Melton, Petrila, Poythress, & Slobogin, 2007; Wettstein, 2005).

When these formal characteristics are not met, shortcomings ensue; hence failure to establish links between different aspects reported on, a lack of logical organization of the contents of the report, superficiality or inadequate use of language and poor writing skills constitute the types of weaknesses that are mostly noted when examining the quality

Correspondence: João da Silva Guerreiro, Université de Montréal - C.P. 6128, Succ. Centreville, Montréal (Québec) H3C 3J, Canada. Email: joao.da.silva.guerreiro@gmail.com

of forensic psychological reports (Grisso, 2010; Lander & Heilbrun, 2009; Nicholson & Norwood, 2000; Skeem & Golding, 1998; Skeem, Golding, Cohn, & Berge, 1998; Wettstein, 2005). Wettstein (2005, 2010) notes however that most studies of forensic reports draw on data from restricted geographical areas and have focused only on their formal characteristics, which might not always be generalizable to other social contexts or to different justice systems around the world, as other authors have also noted (Duits, van der Horn, Wiznitzer, Wettstein, & Beurs, 2012; Lander & Heilbrun, 2009). Thus the conformity to guidelines in forensic report writing appears insufficient to guarantee the quality of forensic psychological reports (Wettstein, 2005).

The present study seeks to address these observations. Its main objective is to provide a general portrait of forensic psychological report writing under the Portuguese criminal justice system. Further, in order to determine their general quality, their analysis in terms of relevance and coherence will be presented. A short theoretical context describing the main concepts used to frame this study will be given, then a methodological section will follow, after which the results will be presented and discussed. Concluding remarks will follow.

Theoretical Context

The choice of relevance and coherence as criteria to evaluate the quality of reports stems from Wettstein's (2005) meta-analysis on the nature and quality of forensic report writing. In research that aimed to identify the strengths and weaknesses of forensic mental health evaluation, Wettstein (2005) argued that one of the main shortcomings of forensic psychological reports is the failure to establish links between the data collected during the forensic assessment and the psycho-legal issues that experts are called on to address, whether competency to stand trial, criminal responsibility, or child custody issues. More specifically, in four out of six studies

considered, psychologists struggled to establish logical connections between the clinical data they presented and their conclusions (Wettstein, 2005). Moreover, the reasoning behind these experts' opinion was not only absent, it was not even implicit in many of the reports analysed. For instance, in a study conducted by Skeem et al. (1998), experts provided data or reasoning to describe how defendants' psychopathology compromised their competency to stand trial in only ten of the 100 reports analysed. In contrast, experts in the same sample typically presented sufficient reasoning to support their clinical conclusions ($n = 87$). Wettstein (2005) argued that forensic psychologists must be able to show that the findings of their psychological assessments are relevant to the judicial file under scrutiny and that their methods are relevant according to judicial, as well as to clinical criteria. Otherwise, their reports might fail to play the role of informational support they should in the judicial decision-making process. Thus a relevant report in the context of this study is one where the various sources of information are weighed in view of portraying the distinctive characteristics of the assessee that are pertinent for the assessment goals. To achieve this ideal of relevance, it is important to identify clearly in the report which method was used in the course of the assessment, that is what guided the expert in the weighing of the various sources of information and in selecting what is specific to the assessee. Relevance in this study is hence defined both in relation to the legal criteria that frame the assessment mandate, and in relation to the degree to which the information reported is unique to the individual assessed.

With regard to coherence, Wettstein (2005) suggests that forensic psychologists must not only take into account all of the information they have at their disposal, but that they must also use the data in their understanding of the person, the situation and the legal issues at hand. Difficulties arise when data are presented but left uninterpreted in a

report. Indeed, when no logical connection is made between different parts of a report or when various data are presented without interpretation of its meaning, there is a risk that an incorrect interpretation be made by one or another of the many actors who eventually get to read the report, be it the state prosecutor, the judge, the attorneys or other clinical and penal staff who look upon such reports for guidance in their interventions once the judicial decision has been rendered. Thus a coherent report in the context of this study is one that integrates information about the assessee from different sources in a logical manner, and if incongruent information appears, this is discussed in the report. Implied in this definition of coherence is a logical sequence in the way the information is presented in that the explanations offered in the report for assessee's behaviour, for instance, follow from the assessment material such as interview material or from results of psychological tests.

This study addresses for the first time the dimensions of relevance and coherence as possible determinants of quality forensic psychological report writing. Indeed, these two dimensions have been alluded to in the literature (Grisso, 2010; Lander & Heilbrun, 2009; Nicholson & Norwood, 2000; Wettstein, 2005), but have not been specifically addressed. A quality forensic report has been determined in large part by criteria based on the formal characteristics of the reports, such as the organization of content or the clarity of the language employed. In order to understand how these formal characteristics relate to the dimensions of relevance and coherence, information about formal characteristics were also collected in this study. This consisted of examining whether the content in the report was structured in well-identified sections, whether experts attribute information to sources, or define psychological terms, when these were used in the report. The information collected about these formal characteristics was used to see in which way they relate to coherent and relevant reports as

defined above. Furthermore, this allowed the comparison between the dataset of reports used in this study and previous ones documented in the literature.

Method

Dataset

Data for this study were collected from archival records in the three main branches of the Portuguese National Institute of Legal Medicine and Forensic Sciences (NILMFS), which is the forensic state institution responsible for the production of forensic psychological assessments in Portugal. The dataset consists of all the forensic psychological reports written between 2006 and 2011 in those three main branches of the NILMFS and concern individuals charged with criminal offences for whom either a judge or a state prosecutor had asked for a forensic assessment. All reports were produced by a psychologist licensed to practise in Portugal and who is affiliated with one of the three main branches of the NILMFS; forensic psychologists affiliated with the NILMFS usually have between five and 20 years of experience. A total of 142 reports were produced during the time frame set for this study. After having been scanned, the names of the author and of the assessee were erased from each report and a code number was assigned to each in order to render them anonymous. The NILMFS' Board granted access to its archives further to the approval of the research protocol by the Research Ethics Review Board of the Université de Montréal.

Judicial Context

Forensic psychological assessments in criminal law are produced in Portugal mainly under two articles of the Portuguese Criminal Procedural Code (PCPC), namely articles 159 and 160 (Carmo, 2005, J.M.P. Silva, 1993). Although both types of forensic reports used in this study are most often requested at the pre-trial stage

of judicial procedures, they may also be required for trial (Carmo, 2005). Article 159 of the PCPC refers to *forensic psychiatric assessments* aimed at determining issues of criminal responsibility (Antunes, 2011). In these assessments, psychologists mainly assist psychiatrists as co-authors of psychiatric assessments. They usually sign their own assessment report, which is joined to the forensic psychiatrist's report (J. M.P. Silva, 1993). As for the assessments produced under article 160, these are said to evaluate the "non-pathological psychological features [and] degree of socialization" of alleged offenders by describing issues of "personality and dangerousness" (Antunes, 2011, p. 80). The opinions presented in these reports may be used to decide upon issues of preventive detention, guilt adjudication and/or the severity of a sentence (Antunes, 2011; Carmo, 2005). It is important to add that, under the PCPC, the results of these assessments are not only considered as evidence, but furthermore the judge is bound to their results, when the assessment report is duly validated as evidence before the court (G.M. Silva, 2002).

Of the 142 reports of this study, 74 were written under article 159 (forensic psychiatric assessment) and 62 were produced in accordance with article 160 (personality assessment).

Research Grid

A Coding Grid was developed to study relevance and coherence as these dimensions manifest in forensic psychological reports. Its description will be followed by a brief overview of the preliminary studies aimed at testing its construct validity and its ease of use by coders. In the following sub-section inter-coder reliability will be evaluated. The grid consists of statements describing elements associated to coherence and relevance. The choice of presenting these elements in short sentences was made to ensure ease of use for the coders whose task it was to determine the presence or the absence of each element as it was described in the related sentence. Three

criteria were used to examine relevance as a dimension of quality; these three criteria were further detailed through seven elements, all related to relevance. The same method was used to examine the dimension of coherence, which was divided into two criteria, which were further specified into five elements. These will be described at length in the results section.

In addition to the elements above, the coders were asked to collect data on the formal characteristics of reports. This included whether the report was structured in well-identified sections, whether experts attribute information to sources; whether they employed a clear language, or psychological jargon, for instance. These formal features are associated to the quality of forensic reports according to many authors (Grisso, 2010; Heilbrun, 2001; Lander & Heilbrun, 2009; Norwood & Nicholson, 2000; Witt, 2010).

Preliminary Study

A preliminary study was done based on six forensic reports randomly taken from the dataset in order to verify the construct validity, and to establish if the grid was easy to use by coders. This preliminary study led to a number of observations, notably that the criteria needed to be further defined, which led to the development of *elements* describing more precisely the features associated to the criteria used to define relevance and coherence. As concerns relevance, it was deemed important to be able to distinguish between reports wherein only general descriptions provided by test manuals are reproduced from those wherein the meaning of test scores are elaborated on specifically for the individual assessed. This distinction was translated into the grid in the form of the two following elements: 1) "Test data are presented in reference to their normative meaning" (R1.3), and 2) "Test data are discussed in relation to the individual assessed" (R1.4). Another example of an observation that prompted adjustments to the grid concerned coherence, more

specifically, the presence of incongruent data that was not elaborated nor discussed in the report. This was translated into the grid in the form of the following element: "Incongruent data are discussed" (C1.2).

A glossary was constructed, once all modifications to the grid have been made, to ensure that coders shared their understanding of the dimensions, criteria and of the elements used to evaluate relevance and coherence. This glossary contains the definition of all the terms used in the grid; it also provided coders with examples taken from the preliminary study for the coding of the 12 elements described in the grid. The coder's task was to evaluate the presence, the case being, of each of the elements that were comprised in the grid. These elements were presented as sentences written in the form of *statements concerning the presence* of each element analysed. Coders were asked to confirm the presence of each element (yes) or state its absence (no). The six reports used in the preliminary study were not included in the final dataset.

Inter-coder Reliability

Further to the preliminary study, the inter-coder reliability was evaluated. The two coders who participated in the analysis of the reports were two licensed clinical psychologists from Portugal, respectively with four and 10 years of experience in psychological assessments. They were trained by the first author to use the grid and become familiar with the coding manual containing the glossary of terms, illustrations of criteria and examples of certain elements. The first author also acted as the third coder for the remainder of the research once all three agreed with the meaning of each dimension, criteria and element and each felt familiar with the grid.

Inter-coder reliability was first calculated based on the results of a sample of 15 reports randomly selected. The inter-coder reliability was found to be unsatisfactory in this first attempt (i.e. a minimum Cohen's kappa value of .6 per element; Sim & Wright, 2005). An

analysis of the coded reports was thus undertaken by the three coders with the collaboration of the second author in order to identify the problems hindering the reliability of coding. This analysis led to the establishment of clearer definitions of the coding criteria, the creation of a glossary of meanings that was completed by examples and illustrations taken from other reports than the ones comprised in the preliminary study. As a result, a better operationalization of the grid was achieved and the working definition of each element of the grid proved satisfactory to each coder.

Following the revision of the coding grid and the supplemental training of the coders, a second sample of 15 randomly selected reports was coded by all three coders. Inter-coder reliability then proved satisfactory for all the elements on the grid. Out of the 15 reports analyzed, the average inter-coder reliability was 74.7 per cent, with a range between 65.1 to 100 per cent. The 30 reports used to establish inter-coder reliability were not included in the final dataset.

Procedures

Once this preparatory stage was completed, the remaining 106 reports were randomly assigned to each of the three coders who then coded them independently one of the other. This dataset was comprised of 57 reports produced according to article 159, and 49 according to article 160. All research material was identified with a code name during the coding process and the reports with code names are kept in a locked file cabinet accessible only to the principal researchers. To assure the anonymity of the dataset, all identifying information was deleted from the material given to coders.

Data Analysis

The presence, or absence, of each element comprised in the grid was assessed by each of the three coders for the entire dataset. The frequencies for each of the seven elements

associated to the presence of relevance were calculated, then the overall score for each of the three criteria was determined. The same type of analysis was followed for coherence, which was assessed by two criteria that were further divided into five elements, the frequency of each will be presented in the next section.

Results

First, the results concerning the formal characteristics of the reports will be presented; second, the results pertaining to the two main dimensions analysed, relevance and coherence will be presented and commented upon.

Formal Characteristics

Most reports had a median length of seven pages, although a few reports were long — the longest being 69 pages — and one was exceptionally short, containing only one page. The majority of reports concerned individuals accused of crimes against persons ($n = 73$); a good proportion were about crimes against property ($n = 25$), and the remaining concerned crimes of various nature. Reports were written in a clear language for the most part (48.1%; $n = 51$) and the results were presented in well-identified sections (57.5%; $n = 61$). When psychological terms were used, they were defined in close to a quarter of the reports (23.6%; $n = 25$); and in about a fifth of them the information used was clearly associated to its source (21.7%; $n = 23$). In more than a third of all reports, the criteria suggested by Grisso (2010, p. 108) concerning the formal presentation of reports were met (38%). These formal characteristics will be discussed later in the general discussion.

Relevance

A summary for each criterion used to code the presence of the dimension of relevance will be presented, followed by a more

detailed description of the seven elements these criteria were further divided into. As pertains to relevance, the coders looked for the following three criteria in the reports: R1) “a clear methodology is employed;” R2) “various sources of information are used and their relative importance is taken into account” and R3) “assessment goals [associated to articles 159 or 160 of the PCPC] are addressed.” The overall scores for each of these three criteria are displayed in Table 1.

An overview of the results for this dimension shows that a clear methodology was used in about half of the reports (47.9%; R1); that in about a tenth of them the results were presented in a manner that shows that a certain hierarchy of importance was given to different sources of information (13.2%; R2), and finally that nearly all reports present opinions that are related to the assessment goal (90.5%; R3.1b) in the case of said *personality assessments* (article 160). However, in the case of reports produced under article 159, which are psychological assessments contained within psychiatric ones, opinions related to the aim of the assessment are not presented (0.9%; R3.1a). A detailed analysis of the elements that are comprised in these three criteria follows.

As table 1 shows, most reports present the methodology used in the assessment at the beginning of the report, (85.8%; $n = 91$; R1.1), furthermore, about a fourth of the reports (26.4%; $n = 28$; R1.2) evidence methodological consistency (R1.2) as measured by the presence of material taken from interviews or test results. Methodological consistency, as an element related to relevance is rather difficult to assess since the absence of explicit references to interview or test material does not necessarily mean that such information was not used in the production of the report. This means that the result obtained for this element is probably much lower than what the actual methodological consistency of these reports would suggest.

Explanations concerning standardized tests results, nomothetic and/or idiographic

Table 1. Results for relevance.

<i>Criteria and elements related to criteria</i>	Percentage observed	<i>N</i> observed (<i>n</i> =106)
<i>R1. Clear methodology employed</i>	47.9	
R1.1. A clear methodology is employed in the assessment;	85.8	91
R1.2. Presence of methodological consistency;	26.4	28
R1.3. Test data is presented in reference to their normative meaning;	69.8	74
R1.4. Test data is discussed in relation to the individual assessed.	9.4	10
<i>R2. Various sources of information used and their relative importance is taken into account</i>	13.2	
R2.1. Observations distinguished from interpretative hypotheses;	21.7	23
R2.2. Qualitative material is integrated and discussed along with quantitative material.	4.7	5
<i>R3.1a. Assessment goals associated to article 159 are addressed (n=57)</i>	.9	
R3.1a. Assessment goals associated to article 159 are met according to two criteria:		
i) capacity to assess the situation that originated the judicial file and capacity to determine oneself accordingly;	1.8 (0) ¹	1(0)
ii) capacity to appreciate the wrongfulness of the offense.	0 (0)	0(0)
<i>R3.1b. Assessment goals associated to article 160 are addressed (n=49)</i>	90.5	
R3.1b. Assessment goals associated to article 160 are met according to three criteria:		
i) personality assessment;	100 (100)	49(49)
ii) dangerousness;	87.8 (53.5)	43(23)
iii) degree of socialization.	83.7 (46.3)	41(19)

¹The assessment goals associated to articles 159 and 160 were coded as present by the coders if they detected information regarding the criteria defined in PCPC for each element. However, not all reports addressed the assessment goals in the same manner. While in some reports the goals were addressed without any information to support the opinion rendered, in other reports such opinions appeared supported. To make the distinction between these two ways of addressing the criteria previewed for articles 159 and 160, the number of reports and its percentage where expert opinions appear supported are displayed in brackets.

results, or still about categorizations associated to test results are also associated to relevance, as coded in two elements (R1.3. and R1.4) associated to relevance through criterion 1, “a clear methodology is employed.” Indeed, most reports contextualized the scores presented in reference to types or other categorizations, for example, to a type of intelligence, or a type of disorder (69.8%, *n* = 74; R1.3). In a few reports, the meaning of a particular test score for the person evaluated was elaborated on (9.4%, *n* = 10; R1.4),

which refers to a more highly developed ability to explain test results. In such instances, particular aspects of the person’s trajectory, or of his or her narrative might be used as examples of particular ways of thinking, or of being, or of acting that stood out during the assessment or in test results. The capacity to communicate what is specific and personal about a given individual is expected from *personality assessments* not only in the Portuguese context, but also in other assessment contexts around the world (Duits et al., 2012;

Grisso, 2010; J.M.P. Silva, 1993). Indeed, relevance in a written report, in its ideal form, might consist in successfully bridging a person's idiosyncrasies and personality traits with his test results and his behaviour.

The second criterion used to identify relevance as a dimension concerns how information from different sources is weighed and if its relative importance is taken into account in the report (R2). This criterion was assessed in the grid by the use of two elements wherein a distinction was established between observations about individuals, through interviews or psychological test results, and the presence of interpretations about these observations (R2.1) on the one hand, and material from quantitative and qualitative sources which was integrated to other results and discussed on the other (R2.2). In more than a fifth of reports (21.7%; $n = 23$; R2.1), clear distinctions were made between observations about the assessee and interpretative hypotheses concerning him, which means that it was clear to the coders which observations led to which hypotheses.

The results show that few reports show links between quantitative and qualitative sources of data (4.7%; $n = 5$; R2.2). Indeed, data collected from different sources, such as interview material and test scores were mostly reported as such, without links being established between them.

The third criterion used to address the dimension of relevance consisted of evaluating whether the information presented in the reports corresponded to the goals set for these assessments according to either article 159 or article 160 of the PCPC. A total of 57 reports were conducted according to article 159 amongst which only one report corresponded to the goal set for the assessment. This result was expected considering that in assessments produced under article 159, the psychologist's role consists in assisting the psychiatrist who is the main author of the report (Pais, 2004; Silva, 1993).

Nearly all the reports (90.5%; R3.1b) addressed the three assessment aims related

to article 160; which are personality aspects, degree of socialization and dangerousness. Concerning personality aspects, all the reports (100%; $n = 49$; R3.1bi) presented arguments taken from the information gathered during the assessment process, most commonly from test material. In the case of dangerousness, more than half the reports (53.5%; $n = 23$; R3.1bii) contained arguments that sustained the assessments made of dangerousness. Finally, close to half of the opinions presented about the degree of socialization were based on arguments taken from the data gathered during the assessment process (46.3%; $n = 19$; R3.1biii).

Coherence

A summary for each criterion used to code the presence of coherence will be presented, followed by a more detailed description of the five elements these criteria were further divided into. As pertains to coherence, the coders looked for the following two criteria in the reports: C1) "the presence of coherent information about the assessee across different sections of the report" and C2) "Explanations for behaviour follow from information presumably reported." The overall scores for each of these two criteria are displayed in Table 2.

The results presented in table 2 show that information about the assessee was coherent across different sections of the report in only 27.9% (C1) of the reports, and relatively few reports (7.5%; C2) offered explanatory hypotheses for assessee's behaviour. As for the elements related to the first criterion, three elements were coded, the first considered whether information about the assessee is presented in a logical way, that is, not contradictory or incoherent. Incoherent and contradictory information about the assessee may be reported in a forensic psychological report without this resulting in an incoherent report. What is critical according to the literature on forensic report writing is that this incoherence is acknowledged and discussed in the report

Table 2. Results for coherence.

<i>Criteria and elements related to criteria</i>	Percentage observed	<i>N</i> observed (<i>n</i> =106)
<i>C1. Presence of coherent information about the assessee across different sections of the report</i>	27.9	
C1.1. Information about the assessee is articulated in a logical way (i.e. not contradictory or incoherent);	48.1	51
C1.2. Incongruent data is discussed;	16.0	17
C1.3. All information presented in the report is taken into account when interpreting test results.	19.8	21
<i>C2. Explanations for behaviour follow from information previously reported</i>	7.5	
C2.1. Presence of hypotheses explaining behaviour;	10.4	11
C2.2. Use of the results of the assessment material (interview material, psychological tests, etc.) to develop hypotheses about behaviour.	4.7	5

(Grisso, 2010; Ogloff & Douglas, 2003). The distinction between incoherence in the information reported on, on the one hand, and the discussion of incoherent information, on the other, was taken into consideration in the first criteria used to study coherence. In about half of the reports, logical connections were found between various sources of information about the assessee (48.1%; $n = 51$; C1.1.), and it appears that less than a fifth of reports presented a discussion of incongruent data (16%; $n = 17$; C1.2.). The third element pertaining to the first criterion to study coherence consisted in ascertaining whether there was a logical sequence between report sections. Results show that in close to 20% of reports, previous information was taken into account in the interpretation of test results (19.8%; $n = 21$; C1.3). As for the second criterion concerning coherence, two elements were coded; the first identified the presence of explanatory hypotheses (10.4%; $n = 11$; C2.1), whereas the second relates to the use of assessment material to offer explanations of the alleged criminal behaviour. Thus, few reports evidenced such a usage of assessment

results (4.7%; $n = 5$; C2.2) that it might not be a very robust indicator of coherence.

Discussion

The observations that emanate from this research will be discussed firstly in terms of the main objects of this research, namely relevance and coherence as indicators of overall quality in light of existing literature. Secondly, hypotheses for future research will be suggested, followed by closing remarks.

Relevance in forensic psychological reports was defined in this exploratory study in relation to the legal criteria that frame Portuguese assessment mandates as well as to existing literature on the subject. Although most of the reports in the present dataset met the requirements inherent to their mandate since they contained conclusions pertaining to each of the three elements comprised in article 160 of PCPC, which are assessing *personality* aspects, *dangerousness* and *degree of socialization*, it appears that some specific elements were lacking. For example, many reports did not propose arguments explaining

their assessments of the individual's *dangerousness*, or in other cases, the individual's *degree of socialization*. However, it appears that in this sense, the results obtained from this study are quite in line with those of Lander and Heilbrun (2009) who have found that in about 60% of the reports analysed, the reasoning explaining the connection between the psycho-legal issues that assessments are expected to address, and the data collected during the forensic assessment were not provided in the report. A similar finding was observed in Grisso's (2010) sample ($n = 62$) where the experts' opinions about key psycho-legal issues were absent in 56% of the reports. Indeed, the ability to present the source or sources in the material for one's opinions enhances both the report's credibility before the justice system as well as showing the usefulness of psychologists as forensic experts (Skeem et al., 1998; Wettstein, 2005).

Another facet to relevance explored in this study consisted in examining whether the various sources of information were weighed in the reports, and whether the information provided therein was unique to the individual assessed. Results show that in only a minority of reports considered, the data from various sources of information appear weighed (4.7%; $n = 5$; R2.2), and in only 10% of the reports (9.4%; $n = 10$; R1.4), the information reported goes beyond the normative meaning of test scores. Relying mainly on the results of standardized tests, as it was often observed in the dataset of reports used in this study may amount to results that are more general and in which the individuality of the assessee might be lost (APA, 2013; Griffith, Stankovic, & Baranoski, 2010; Griffith & Baranoski, 2007). This observation is unfortunate in the Portuguese context, since the aim of these judicial mandates is specifically to provide legal decision-makers with highly specific information concerning the individual so as to render the judicial decision-making process more individualized (Dias, 1983). The results of this study regarding the way

the various sources of information are weighed, mimic those of Lander and Heilbrun (2009) who found that in about 60% of the 125 reports analysed, experts presented their opinions based on one sole data of information. In the sample of 62 reports analysed by Grisso (2010), 22% fail to use more than one source of information to sustain expert opinions. Furthermore, experts' over-reliance on one single source of data was singled out by Grisso (2010) as one of the ten most frequent faults in forensic report writing.

Heilbrun (1990) recommends treating test results as hypotheses subject to verification through other sources of information such as an individual's history, his medical records, as well as with the help of third party observation. Some professional associations and licensing boards, both in North America and in Portugal, have supported that same recommendation (APA, 2013; Ordem dos Psicólogos Portugueses, 2011; Ordre des psychologues du Québec, 2002a, 2002b).

The dimension of coherence was explored notably by assessing whether the information presented in different sections of the report fit together in a logical fashion, without appearing incongruent. Results of this exploratory study suggest that in about half of the reports considered (51.9%; $n = 55$), information about the assessee is incoherent across the reports sections and when incongruence was detected, for instance, between information reported on the assessee based on different data sources, incongruence was rarely discussed in the report (16%; $n = 17$). To the contrary, reports where information about the assessee was presented in a coherent fashion and incongruence discussed, invited the reader to reason along with the report author and amounted in that sense to a much more coherent rendition of the assessee's characteristics. This observation replicates that of Grisso (2010) who found that in a third of his sample (30%; $n = 62$), although the data presented in the report allowed for alternative explanations, this was not discussed in the report. Grisso recommends in line with this

observation that the meaning of inconsistencies in the information provided about the assessee should be carefully addressed by the expert, for instance, resorting to results from different sources of data to provide a more solid basis for the results' discussion. Rendering forensic reports more coherent is of importance as this may minimize the deleterious effects of data misinterpretation as information that is duly integrated avoids the possibility of leaving information open for interpretation by other judicial actors that look upon reports for informational support for their decisions. Indeed, it has been documented that legal decision-makers' judgments are as exposed to errors and biases as those of other people and that some of these errors may be prompted by the way in which information is presented or emphasized in forensic reports (Dhami & Ayton, 2001; Goodman-Delahunty & Dhami, 2013). If information is duly integrated, potential misinterpretations can thus be prevented which contributes to a more significant participation of psychologists in the criminal justice system.

This exploratory study sought to fill the need identified in the literature for more knowledge on civil-law countries. The analysis of the relevance and coherence as a way of examining the quality of reports suggest that the characteristics identified in the dataset of reports written under the Portuguese criminal law are consistent with previous report surveys in North America. Amongst the areas that appear deficient, the need to better link the test results to observations associated to interview material and /or information from other sources stands out, as is the need to link the results of the assessment to the forensic issues that form the basis of the referral (Lander & Heilbrun, 2009; Nicholson & Norwood, 2000; Wettstein, 2005).

The literature about forensic report writing has given much emphasis to the formal aspects of reports as markers of their general quality (Griffith et al., 2010; Nicholson & Norwood, 2000; Wettstein, 2005, 2010). It was interesting to note in this exploratory

study that, although formal characteristics such as clarity of the language used, presentation of assessment findings in well-identified sections and correct attribution of information to their sources were manifest in a significant proportion of the reports evaluated, these qualities were not necessarily equated with the general coherence and relevance of their content in our study. As some authors have suggested, such formal characteristics while indicative of quality of a report do not amount to its relevance (Griffith et al., 2010; Griffith & Baranoski, 2007; Melton et al., 2007; Wettstein, 2010). To that effect, Griffith et al. (2010) argue that formal characteristics provide a structure, which is very important, but "insufficient for delineating what is necessary to create a persuasive and relevant product" (p. 33).

The survey of the reports' characteristics presented in this study generated ideas and raised questions that could be addressed in future research in view of developing forensic psychology in Portugal. The development of a more nuanced grid based on the one used in this study is one possible research avenue. Relevance seems to have been better operationalised through the seven elements, when compared to the dimension of coherence which will require further elaboration in order to develop a better operationalization of what, at face value, appears well suited to measure quality. The use of the grid in other settings, forensic or therapeutic, may help to better operationalise the dimensions of relevance and coherence as quality indicators of forensic psychological reports.

The findings of this study also raise the question of what are the psychologists' views on what would constitute a quality report? How do they see relevance and coherence as quality indicators of forensic report writing? Enquiring of the professionals who regularly conduct forensic psychological assessments may offer new insights into whether relevance and coherence are the best concepts to study quality in forensic psychological reports and to better operationalize these two concepts.

The question of whether these two concepts are valid measures of quality in forensic report writing is important for the intellectual debate about the quality of forensic psychological assessments because it goes beyond the focus on the mechanics and organization of a report that has mainly been at the centre of the literature in forensic report-writing rather than considering more abstract and conceptual issues such as the requirements of forensic reports which, when compared to clinical reports, need to address very specific issues (Griffith et al., 2010). If the concepts of relevance and coherence used in this exploratory study are proven valid, then one can imagine possible avenues for continuing professional training of forensic psychologists drawing on the findings of this first study, notably concerning the integration of diverse sources of knowledge about assesses, and how to coherently weigh such information.

Concluding Remark

This exploratory research sought to contribute to the ongoing discussion about forensic psychological reports by analysing issues of relevance and coherence as indicators of the quality of forensic psychological reports as suggested by Wettstein (2005). Research that succeeds in better understanding the role played by these dimensions in the quality of reports will contribute to the improvement of professional training of forensic psychologists and other mental health experts.

Acknowledgements

We wish to express our gratitude to Ms. Eliana Vilaça and Mr. João Ferreira who agreed to participate in the preparatory work involved in this research and for coding the reports.

Funding

This work was supported by Fundação para a Ciência e a Tecnologia, IP [grant number SFRH/BD/60286/2009].

References

- Ackerman, M.J. (2006). Forensic report writing. *Journal of Clinical Psychology, 62*, 59–72.
- American Psychological Association. (2013). Specialty guidelines for forensic psychology. *American Psychologist, 68*, 7–19.
- Antunes, M.J. (2011). *Código de Processo Penal: Anotado e legislação complementar* [Penal Procedural Code: noted and supplementary legislation]. Coimbra, Portugal: Coimbra Editora.
- Carmo, R. (2005). A prova pericial: Enquadramento legal [Forensic assessment: legal framework]. In R.A. Gonçalves & C. Machado (Eds.), *Psicologia Forense* (pp. 33–54). Coimbra, Portugal: Quarteto.
- Conroy, M.A. (2006). Report writing and testimony. *Applied Psychology in Criminal Justice, 2*, 237–260.
- Dhami, M.K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making, 14*, 141–168.
- Dias, J.F. (1983). *Liberdade, culpa, direito penal* [Freedom, guilt and penal law] (2nd ed.). Coimbra, Portugal: Coimbra Editora.
- Duits, S., van der Horn, S., Wiznitzer, M., Wettstein, R.M., & Beurs, E. (2012). Quality improvement of forensic mental health evaluations and reports of youth in the Netherlands. *International Journal of Law and Psychiatry, 35*, 440–444.
- Goodman-Delahunty, J., & Dhami, M.K. (2013). A forensic examination of court reports. *Australian Psychologist 48*, 1, 32–40. doi: 10.1111/j.17429544.2012.00082.x
- Griffith, E.E.H., & Baranoski, M.V. (2007). Commentary: The place of performative writing in forensic psychiatry. *Journal of the American Academy of Psychiatry and Law, 35*, 27–31.
- Griffith, E.E., Stankovic, A., & Baranoski, M. (2010). Conceptualizing the forensic psychiatry report as performative narrative. *Journal of the American Academy of Psychiatry and Law, 38*, 32–42.
- Grisso, T. (2010). Guidance for improving forensic reports: A review of common errors. *Open Access Journal of Forensic Psychology, 2*, 102–115.
- Heilbrun, K. (1990). Response style, situation, third-party information, and competency to stand trial: Research issues in practice. *Law and Human Behavior, 14*, 193–196.
- Heilbrun, K. (2001). *Principles of forensic mental health assessment*. New York: Kluwer Academic/Plenum.
- Heilbrun, K., & Collins, S. (1995). Evaluations of trial competency and mental state at the time of

- offense: Report characteristics. *Professional Psychology: Research and Practice*, 26, 61–67.
- Lander, T., & Heilbrun, K. (2009). The content and quality of forensic mental health assessment: Validation of a principles-based approach. *International Journal of Forensic Mental Health*, 8, 115–121.
- Melton, G., Petrila, J., Poythress, J., & Slobogin, C. (2007). *Psychological evaluations for the courts* (3rd ed.). New York: Guilford.
- Michaels, M.H. (2006). Ethical considerations in writing psychological assessment reports. *Journal of Clinical Psychology*, 62, 47–58.
- Nguyen, A.H., Acklin, M.W., Fuger, K., Gowensmith, W.N., & Ignacio, L.A. (2011). Freedom in paradise: Quality of conditional release reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry*, 34, 341–348.
- Nicholson, R., & Norwood, S. (2000). The quality of forensic psychological assessments, reports, and testimony: Acknowledging the gap between promise and practice. *Law and Human Behavior*, 24, 9–44.
- Ogloff, J.R.P., & Douglas, K.S. (2003). Psychological assessment in forensic settings. In J.R. Graham, J.A. Naglieri, & I.B. Weiner (Eds.), *Handbook of psychology, volume 10: Assessment psychology* (pp. 345–363). New York: Wiley.
- Ordem dos Psicólogos Portugueses. (2011). *Código deontológico da ordem dos psicólogos portugueses* [Code of Ethics of the Portuguese Psychologists' Professional Association]. Lisbon, Portugal: Ordem dos Psicólogos Portugueses.
- Ordre des psychologues du Québec. (2002a). L'expertise psycho-légale (partie 1). *Fiche déontologique*, 3(5), septembre 2002. *Psychologie Québec*, 19(6), 8–19.
- Ordre des psychologues du Québec. (2002b). L'expertise psycholégale (partie 2). *Fiche déontologique*, 3(5). Retrieved from www.ordrepsy.qc.ca/pdf/1Fiche_Deonto_Nov02.pdf
- Pais, L.G. (2004). *Uma história das ligações entre a psicologia e o direito em Portugal: Perícias psiquiátricas médico-legais e perícias sobre a personalidade como analisadores* [A history of connections between psychology and law in Portugal: Forensic psychiatric assessment and personality assessment as analysers] (Doctoral dissertation, Universidade do Porto).
- Ryba, N.L., Cooper, V.G., & Zapf, P.A. (2003). Juvenile competence to stand trial evaluations: A survey of current practices and test usage among psychologists. *Professional Psychology*, 34, 499–507.
- Silva, G.M. (2002). *Curso de processo penal II* [Course on procedural law, Vol. II] (3rd ed.). Lisbon, Portugal: Verbo.
- Silva, J.M.P. (1993). A propósito do exame psicológico no âmbito penal [About the forensic psychological examination in penal law]. *Análise Psicológica*, 1(11), 29–36.
- Sim, J., & Wright C.C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical Therapy*, 85, 257–268.
- Skeem J.L., & Golding, S.L. (1998). Community examiners' evaluations of competence to stand trial: Common problems and suggestions for improvement. *Professional Psychology*, 29, 357–367.
- Skeem, J.L., Golding, S., Cohn, N., & Berge, G. (1998). Logic and reliability of evaluations of competence to stand trial. *Law and Human Behavior*, 22, 519–547.
- Wettstein, R. (2005). Quality and quality improvement in forensic mental health evaluations. *The Journal of the American Academy of Psychiatry and the Law*, 33, 158–175.
- Wettstein, R. (2010). Commentary: Conceptualizing the forensic psychiatry report. *Journal of the American Academy of Psychiatry and Law*, 38, 46–48.
- Witt, P.H. (2010). Forensic report checklist. *Open Access Journal of Forensic Psychology*, 2, 233–240.