

# Testing International

Volume 51, June 2024

Editor: Nicky Hayes



## International Test Commission

### PRESIDENT

Stephen G. Sireci, University of Massachusetts Amherst, USA

### PRESIDENT-ELECT

Kadriye Ercikan, ETS, USA

### SECRETARY-GENERAL

April Zenisky, University of Massachusetts Amherst, USA

### TREASURER

Anita M. Hubley, University of British Columbia, Canada

### COUNCIL MEMBERS

#### Elected Members

Wayne Camara Law School Admissions Council, USA

Jon Twing, Pearson USA

Peter Macqueen, Compass Consulting, Australia

Rainer Kurz, HUCAMA, UK

### CO-OPTED MEMBERS

Aletta Odendaal, University of Stellenbosch, SA

Paula Elosua, University of the Basque Country, Spain

Steve Dept, cApStAn, Belgium

Felipe Valentini, Sao Francisco University, Brazil

### OBSERVER

Jose Luis Padilla, University of Granada, Spain

### REPRESENTATIVES

IUPsyS Representative Ann Watts, IUPsyS

IAAP Representative Kurt Geisinger,

Buros Center for Testing / University of Nebraska, USA

### EDITORS

*International Journal of Testing*

Chris Nye, University of Michigan, USA

*Testing International*

Nicky Hayes, University of Suffolk, United Kingdom

## Contents

ITC President's Message	2
From the Editor	4
The ITC Conference in Granada	5
Welcome to Granada!	6
The Contributions of Jerry Garcia to Criterion-Referenced Testing: <i>A Cautionary Tale of Chat GPT</i>	11
AI regulation & the implications on Test Use in Europe & beyond	13
Personality inventory contextualized in work	14

Testing International is a publication of the  
International Test Commission



# President's Report

**Stephen G. Sireci**  
ITC President



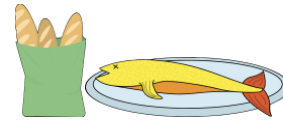
## **Five Loaves, Two Fish, and the International Test Commission**

It is with great pride and a hint of sadness I write my last article for *Testing International* as President of the International Test Commission (ITC). It has been a pleasure serving as the leader of the ITC these past two years, and I look forward to the culmination of my term as President at the ITC 2024 conference in Granada. It is going to be an outstanding conference that will be remembered as one of the most enjoyable and successful in the history of the ITC!

Speaking of history, you may have noticed the historical and religious reference in the title of this article. What does the ITC have in common with loaves and fish? “Loaves and fish” is in reference to one of Jesus’s miracles in the Bible. It is in fact, the only miracle aside from Jesus’ resurrection that is recorded in all four gospels, which gives it particular significance. Do not worry—this will not be a religious article and I am not trying to convert anyone to Christianity. My entire professional life has only been trying to convert us all to fair, efficient, and valid assessment practices!

The miracle of loaves and fish tells the story of how a huge crowd of 5,000 or so gathered to hear Jesus

preach and he decided to feed them. His apostles reported they could only gather five loaves of bread and two fish and so could not feed everyone. Jesus blessed the loaves and fish and when they were passed around to the crowd, everyone had enough to eat, and there were even leftovers!



To me, and I am by no way a Christian theologian or authority, this is the most explainable of all miracles. It could be as the baskets were passed around people started putting food *into* the basket, sharing the food they had with others. In that sense, the miracle could be the incredible sense of humanity and compassion that led everyone there to share what they had with each other for the benefit of all.

Perhaps without further explanation you can see why this miracle reminds me of the ITC. When I took over as President of the ITC in July 2022, we were still coming out of COVID, had our planned conference cancelled, and were experiencing a budget crisis. There appeared to be few resources to maintain our organization, not to mention trying to nourish it. However, since its inception in 1976, the ITC has been run by individuals who keep putting “food in the basket” to not only keep the ITC running, but also to allow it to flourish. I have been amazed by, and so grateful for, all the time and talent selflessly devoted by volunteers throughout the ITC to benefit our organization and our mission.



**Thank you to all ITC Council members, committee members, and other volunteers for putting your food in the basket and allowing the miracle of the ITC to continue! Not many people realize *the ITC has only one employee—and she is part-time!***

Of course we can consider our Office Manager, Ananda van Tonder, to be the greatest miracle of all. Clearly, the ITC could not have succeeded without her; but the real miracle of the ITC is how it was started by a few individuals who cared about the quality of assessment research and practices across the globe; and it continues to be sustained by an small, but amazing group of altruistic measurement researchers and practitioners.



In the past two years alone, the ITC has planned an outstanding conference, developed widely lauded guidelines, published 8 volumes of a journal, hosted several international webinars, updated our website and our mission and vision statements, established organizational goals, added two new committees (Diversity and Justice Committee and Graduate Student Committee), established digital presence on LinkedIn and other platforms, helped establish reporting guidelines for test adaptation research, maintained a successful book series, and contributed to policy debates in the international education community. Not bad for an organization run by a part-time superstar and a handful of volunteers.

As I end my term, it seems insufficient to describe my ITC colleagues as “volunteers.” Clearly you are friends and colleagues, but our relationships seem more like family. There is more than professional admiration that bonds us, the ITC community, together. It is the belief that the values and principles we hold for assessment in our home country transcend the globe. The ITC mission and vision to “promote fair, valid, transparent, and efficient testing, assessment, and reporting practices, guidelines, and policies to benefit individuals, institutions, and societies throughout the world;” is our universal bond. I have greatly enjoyed being part of the ITC and I know, although my term as President is ending, I will continue to be an active member of the ITC. The ITC is an extraordinary organization that has connected me with friends throughout the world and given me more professional development and rewards than any other organization. To connect my feelings of the ITC to the title of this column, and to cement my corniness into the history of the ITC, I will conclude by saying, “I LOAVE the ITC!”

Thank you for allowing me to serve this organization as your President. I look forward to seeing you at our 2024 conference in Granada, and to personally thanking the ITC leadership, and our incredible conference organizers—Jose-Luis Padilla, Luis Manuel Lozano, and Isabell Benítez Baena, for all they have done. I thank God for the miracle of our friendship and I remain eternally grateful to you all. Gracias a todos. Nos vemos en Granada!



**Stephen G. Sireci**  
University of  
Massachusetts  
Amherst, USA

## From the Editor



**Nicky Hayes**  
University of Suffolk, UK

First of all, my apologies to those of you who've been looking forward to them, but this edition doesn't have the usual committee reports. We're so excited about Granada, we wanted to devote most of our space to it! But as well as that, we do have three fascinating articles for you. Our first is an exploration of ChatGPT and the contribution it can make to those struggling to complete book chapters or papers. Caveat Emptor! Or, translated, let the buyer beware! You may get a lot more than you bargained for - and as for fact-checking, best not to think about it! Thanks to our President for that illuminating and amusing contribution.

Following that, Nigel Evans has outlined the content and implications of the new European AI regulations, which promise to become pan-global, not just restricted to Europe. They deal extensively with risk - but sadly, the risks of fabrication don't seem to have been considered. I do worry sometimes that with the growing significance of the virtual world, the concept of truth is simply becoming outdated. It certainly only seems to be important to professionals and academics - people like us, in other words!

Coming back to what is thankfully still the real world, Petteri Nittamo has given us a thoughtful article about the applied value and challenges of the personality inventory at work.

So, happy reading! And see you in Granada!

Nicky



## Announcement of ITC General Meeting

The 2024 General Meeting of the International Test Commission will be held in person at the upcoming ITC conference in Granada, Spain on Thursday the 4th of July from 6.15pm to 7.15pm (Granada Time).

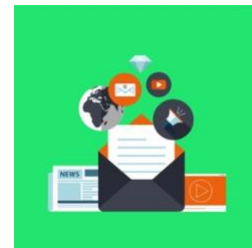
The Agenda for the meeting can be accessed here:

[W ITC Agenda General Meeting\\_2024\\_v1.docx](#)

If you need further information or have questions about the General Meeting, please contact April Zenisky, ITC Secretary-General, on [Secretary@intestcom.org](mailto:Secretary@intestcom.org)

Looking forward to seeing you in Granada!

April Zenisky.



# Don't Miss out!

*It's not too late to register for the ITC Conference in Granada!*

Just click here:

[https://www.itc2024granada.com/programa\\_itcongress\\_a5\\_v5!](https://www.itc2024granada.com/programa_itcongress_a5_v5!)

Imagine, this is what you'd be missing:

## Provisional Programme

### TUESDAY JULY 2<sup>ND</sup>

9.00am - 12.00pm	<b>Short-courses (morning session).</b> <ul style="list-style-type: none"><li>An Introduction to the Theory of Standard Setting. Mark D. Reckase.</li><li>Response Bias in Self-Report Measurements. Ricardo Primi &amp; Felipe Valentini.</li><li>Best Practices for Artificial-Intelligence Scoring of Constructed Responses. Daniel F. McCaffrey.</li></ul>
12.00pm - 12.30pm	<b>Break</b>
12.30pm - 3.30pm	<b>Short-courses (afternoon session).</b> <ul style="list-style-type: none"><li>Questionnaire translation/adaptation: On ensuring equivalence in cross-cultural research. Dorothee Behr &amp; Brita Dorer.</li><li>Cognitive Diagnosis Modeling: Introduction, Recent Developments, and Implementation. R. Jimmy de la Torre &amp; Pablo Nájera.</li><li>Sequence Mining Methods on Process Data in Large-Scale Assessments. Qiwei He, Esther Ulitzsch &amp; Bernard Veldkamp.</li></ul>
3.30pm - 4.00pm	<b>Break</b>
4.00pm - 7.00pm	<b>Short-courses (evening session).</b> <ul style="list-style-type: none"><li>Introduction to AI-based Automated Item Generation and Scoring Practices. Alina Von Davier &amp; Duanli Yan.</li><li>Cross-Cultural Development and Validation of Neuropsychological Tests: Mitigating Cultural Bias. Álvaro Lozano-Ruiz.</li><li>Solving the problems of ipsative data: Designing and scoring forced-choice and other questionnaires in comparative format. Anna Brown.</li><li>A Critical Quantitative measurement perspective: MIMIC models to identify and remediate racial (and other) forms of bias. Matthew Diemer.</li></ul>
7.30pm	<b>Welcome reception:</b> Terrace of the Granada Conference Palace.

### WEDNESDAY JULY 3<sup>RD</sup>

8.30am - 9.45am	<b>Session 1:</b> Parallel sessions from 1.1 to 1.9.
9.45am - 10.45am	<b>Special session:</b> Conference Opening and ITC Presidential Address. Stephen Sireci, <i>ITC President. University of Massachusetts Amherst, USA.</i>
10.45am - 11.15am	<b>Coffee Break</b>
11.15am - 11.45am	<b>Poster session:</b> Poster session 1.
11.45am - 1.00pm	<b>Keynote:</b> Embracing the Future of Remote Proctoring: Ensuring Test Integrity and Accessibility in the Digital Age. Dr. Alina von Davier, <i>Chief of Assessment at Duolingo.</i>
1.00pm - 2.00pm	<b>Conference Lunch &amp; Graduate Student Committee Meeting</b>
2.00pm - 3.15pm	<b>Session 2:</b> Parallel sessions from 2.1 to 2.9.
3.15pm - 4.30pm	<b>Session 3:</b> Parallel sessions from 3.1 to 3.9.
4.30pm - 5.00pm	<b>Coffee Break</b>
5.00pm - 5.30pm	<b>Poster session:</b> Poster session 2.
5.30pm - 6.30pm	<b>Session 4:</b> Parallel sessions from 4.1 to 4.9.
7.15pm - 10.00pm	<b>Alhambra tour:</b> Meeting point: Granada Conference Palace.

### THURSDAY JULY 4<sup>TH</sup>

8.30am - 9.45am	<b>Session 5:</b> Parallel sessions from 5.1 to 5.9.
9.45am - 10.45am	<b>Round Table:</b> Shared Insights and Unique Perspectives: What Can Psychological Testing and Assessment Learn from Innovations in Survey Research and Vice Versa. Beatrice Rammstedt, <i>GESIS - University of Mannheim, Leibniz Institute for the Social Sciences (Germany).</i> Ana Villar, <i>Meta (UK).</i> Bruno D. Zumbo, <i>University of British Columbia (Canada).</i>
10.45am - 11.15am	<b>Coffee Break</b>
11.15am - 11.45am	<b>Poster session:</b> Poster session 3.
11.45am - 1.00pm	<b>Session 6:</b> Parallel sessions from 6.1 to 6.9.
1.00pm - 2.00pm	<b>Conference Lunch &amp; Early Career Scholars Meeting</b>
2.00pm - 3.15pm	<b>Session 7:</b> Parallel sessions from 7.1 to 7.9.
3.15pm - 4.30pm	<b>Session 8:</b> Parallel sessions from 8.1 to 8.9.
4.30pm - 5.00pm	<b>Coffee Break</b>
5.00pm - 5.30pm	<b>Poster session:</b> Poster session 4.
5.30pm - 6.30pm	<b>ITC General Assembly.</b>
8.30pm	Gala Dinner. Carmen de los Mártires.

### FRIDAY JULY 5<sup>TH</sup>

8.30am - 9.45am	<b>Session 9:</b> Parallel sessions from 9.1 to 9.9.
9.45am - 10.45am	<b>Keynote:</b> Seeking Validity Evidence Related to Assessment Justice. Jennifer Randall, <i>University of Michigan (USA).</i>
10.45am - 11.15am	<b>Coffee Break</b>
11.15am - 11.45am	<b>Poster session:</b> Poster session 5.
11.45am - 1.00pm	<b>Session 10:</b> Parallel sessions from 10.1 to 10.9.
1.00pm - 2.00pm	<b>Special session:</b> Closing Ceremony: Awards session & Memorial.

And here's some more of what you would miss!

# ITC Conference 2024

## Granada (Spain) 2024

### WELCOME TO THE 2024 ITC CONFERENCE AND GRANADA CITY!!

Can you believe that the conference is just around the corner? Surely you have already planned your trip, decided what you want to do at the conference, want to talk to, etc. The local committee has planned a program so that your participation in the conference is a before and after in your professional career. You will find the complete program on the conference website, here:

[https://www.itc2024granada.com/programa\\_itcongress\\_a5\\_v5!](https://www.itc2024granada.com/programa_itcongress_a5_v5!)

While contributing to the conference, you surely want to take advantage of the opportunity to enjoy the city. The organizing committee wants to share with you some suggestions for you and your companions. Trust us, we are "locals" and in love with the city.

So that you can plan your trip and what you don't want to miss, we give you two great cultural routes: the Catholic route and the Muslim route. Obviously, it is a simplification because what defines Granada is the mixture of both, but we hope this division helps you.

#### Muslim Granada

We will begin our journey at the Alhambra, where its intricate gardens and fountains will transport us to the splendor of the Nasrid dynasty. From there, we will descend towards the "Paseo de los Tristes" along the Darro River, with impressive views and a unique historical atmosphere. Venturing into the "Albaicín", the ancient Arab v

quarter, we will stroll through its narrow streets and whitewashed houses, discovering the "Placeta del Comino". We will continue towards the "Mirador de San Nicolás", enjoying one of the most iconic views of the Alhambra and Sierra Nevada for unforgettable sunsets. Then, we will climb to the "Mirador de San Miguel Alto" for a complete panoramic view of the city. We know... lots of Spanish words! We want you practice your Spanish...

Do you want to take a flavor of the must-see spot mentioned above? Take a look at the pictures below:

- The Alhambra and Generalife, a UNESCO World Heritage Site since 1984. There is no better example of Muslim architecture with stunning palaces and gardens.



**Don't Forget!!** The 2024 ITC LOC has organized a private visit to the Nasrid palaces of the Alhambra on Wednesday, July 3<sup>rd</sup>, which includes a guided tour, Alhambra tickets, and transport by bus. This is a great opportunity to get to know the monument in a very special way.

- The residential district of the Albaicín is a rich legacy of Moorish town planning and architecture in which Nasrid buildings and constructions of Christian tradition coexist. Within the Albaicín, you can appreciate different viewpoints.



- **Placeta Comino.** From this small square, which surely owes its name to its small size (do some research on the word “comino”), you can contemplate an impressive view of the Alhambra beyond the wall that borders it.



- **Mirador de San Nicolás.** This spot offers one of the best panoramic views of Granada. The set formed by the church of San Nicolás, the cistern, and the small square with its imposing viewpoint defines the relationship of beauty and harmony of the Albaicín neighborhood, the Alhambra, the Darro valley, the old city, and the horizon of the Vega.



- **Mirador de San Miguel Alto.** Situated in front of the “Ermita” of *San Miguel Alto*, you will find the renowned *Mirador de San Miguel Alto*. It is the highest of Granada's views.



- **Mirador de San Cristobal.** The Mirador de San Cristóbal overlooks the ancient citadel of the Albaicín. The buttressed walls were built by the Zirid kings in the 11th century to protect their fortress.



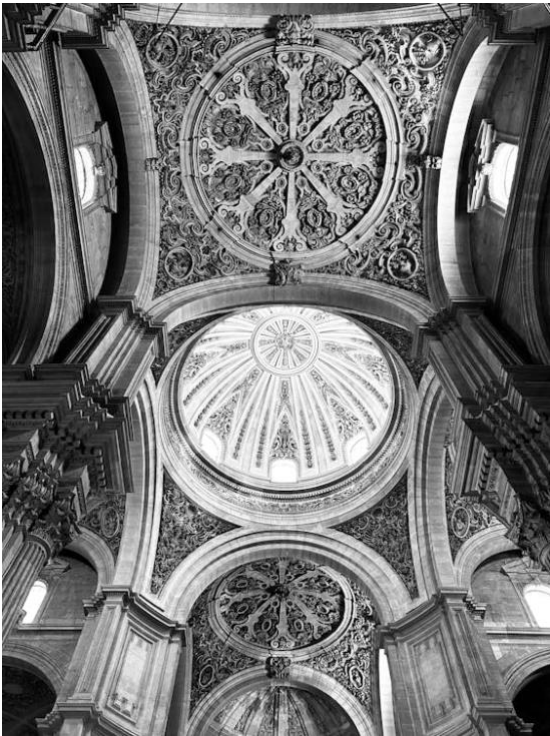
- **Arco de Elvira.** It was built in the 11th century being the main entrance to the city.



### Catholic Granada:

You can begin your day in the heart of Granada, in front of the majesty of the Cathedral of Granada. A few steps away, you will enter the Royal Chapel, where the Catholic Monarchs, Isabel and Fernando, rest in an environment of great artistic beauty. The route will take you to the “Monastery of San Jerónimo”, a masterpiece of the Spanish Renaissance that houses the tomb of Gonzalo Fernández de Córdoba, the Great Captain. This historic place offers you a fascinating insight into the spirituality and devotion of Catholic Granada, as well as the opportunity to admire its magnificent architecture. Finally, the route will take you to the “Monastery of La Cartuja”, a haven of peace and contemplation with elegant cloisters and decorated chapels, where you can immerse yourself in the beauty of devotional art.

- **Granada Cathedral.** Granada Cathedral, a Spanish Renaissance masterpiece commissioned by Isabella the Catholic in 1505. Notable features include the imposing Main Chapel and various chapels in diverse styles, with contributions from architect Diego de Siloé.



- **The Royal Chapel of Granada.** The Royal Chapel, built in Gothic style over 500 years ago by the Catholic Monarchs for their final resting place, houses a rich patrimonial legacy including gold, silver, textiles, and fifteenth-century religious paintings.



- **Royal Monastery of St. Jerome.** It is a Roman Catholic Hieronymite monastery in Granada, Spain. Architecturally, it is in the Renaissance style. The church, famous for its architecture, was the first in the world consecrated to the Immaculate Conception of Mary.



- **Sacromonte Abbey.** The Sacromonte Abbey is a former 17th-century abbey that is currently a residence for canons, a parish, and an ecclesiastical museum located northwest of Granada.



### The two last “tips”

We told you that dividing the two cultural routes is a simplification. There is no better example of this than our last two tips...

- **Paseo de los tristes.** The Paseo de los Tristes offers breathtaking views of the Alhambra. Although its official name is Paseo del Padre Manjón, it's called Paseo de los Tristes for the funeral processions that passed through here en route to the cemetery. Worth seeing on the Paseo de los Tristes is the Renaissance Church of Saints Peter and Paul, with sculptures by Pedro de Mena



- **Camen de los Mártires.** The Carmen of the Martyrs is a famous Carmen and garden located in the Spanish city of Granada. It is situated on the southern slope of the hill of Cerro del Mauror (Hizn Mauror), which supports at its peak the Bermejas Towers, within the precincts of the Alhambra, on the grounds known until the 14th century by the Arabs as Campo de Ahabul and by the Christians as the Campo or corral de los Cautivos.

Maybe you have heard of the “Carmen de los Mártires”... if so, you are a good “2024 ITC student”! It is the breath-taking place where the Gala Dinner will be held. You can learn more... guess where? On the conference website.

[https://www.itc2024granada.com/programa\\_itccongress\\_a5\\_v5!](https://www.itc2024granada.com/programa_itccongress_a5_v5!)

# All About Testing

## The Contributions of Jerry Garcia to Criterion-Referenced Testing: *A Cautionary Tale of Chat GPT*

**Stephen G. Sireci,**  
University of Massachusetts  
Amherst

**Neal Kingston,**  
University of Kansas



In February 2023, the first author of this article invited the second author to collaborate on a chapter to appear in a forthcoming book, *Assessment in the Service of Learning*. The topic of the chapter was new ways of reporting scores on educational tests to better support student learning. The second author eagerly accepted the invitation.

As those of you who work on book projects know, the first step in collaborating on a chapter is to agree on a chapter outline. The first author sent a draft outline to the second author to review, and because they were friends and both music fans, the first author slipped in a satirical topic, just to amuse the second author. See if you can spot the joke in this partial list of chapter topics:

### II. Providing actionable information

- a. Brief intro to NRT and CRT (1 page)
- b. Reporting students' strengths—affirmative assessment, positive assessment, mastery (2 pages)

- c. Influence of Jerry Garcia and Robbie Robertson on CRT (11 pages)
- d. Reporting and discussing assessment results (3 pages)
- e. Dynamic learning maps (1 page)

For those of you who are not familiar with rock music from the 1960s through 1980s, the joke was “c,” the Influence of Jerry Garcia and Robbie Robertson on criterion-referenced testing.” The joke was not lost on the second author. In response to the draft outline, he emailed:

*I did a quick scan and particularly liked IIc... I will think a bit more about it, but it appears to be a solid start. Like everyone else these days, I immediately looked for some artificial assistance..., especially section IIc, which I found insufficient so asked ChatGPT to expand upon it.*

What follows is purely from ChatGPT (including references!). Based on our experience, it is the most authoritative work on the influence of Jerry Garcia and



Robbie Robertson on criterion-referenced testing, and serves as our cautionary tale of using ChatGPT for psychometric research! Enjoy!

### ***Title: The Influence of Jerry Garcia and Robbie Robertson on Criterion-Referenced Testing***

#### ***Introduction***

Criterion-Referenced Testing (CRT) plays a crucial role in modern education, providing a means to assess whether students have achieved specific learning objectives. While the development of CRT is primarily associated with educational and psychometric experts, it is fascinating to explore the unconventional influences that have shaped this field. This essay delves into the unexpected impact of two legendary musicians, Jerry Garcia and Robbie Robertson, on the evolution of Criterion-Referenced Testing.

### ***Jerry Garcia's Contribution to CRT***

Jerry Garcia, best known as the frontman for the iconic rock band the Grateful Dead, may seem an unlikely figure to influence education and testing practices. However, his improvisational approach to music and dedication to innovation left a lasting mark on the world of CRT.

Garcia's influence on CRT can be traced back to the concept of "adaptive testing." In an adaptive test, the difficulty of subsequent questions is determined by the correctness of previous responses. Garcia's musical improvisation paralleled this idea. Just as he adjusted his guitar playing in response to the band's performance, adaptive testing adapts the level of difficulty based on a student's answers, ensuring a more accurate assessment of their abilities (Smith, 2018).

Furthermore, Garcia's emphasis on individual expression and creativity found its way into the philosophy of CRT. In CRT, the focus is on evaluating a student's mastery of specific learning objectives rather



than comparing them to others. This approach encourages students to explore their unique strengths and talents, mirroring Garcia's musical ethos (Thompson, 2019).

### **Robbie Robertson's Influence on CRT**

Robbie Robertson, the lead guitarist and primary songwriter for The Band, also contributed significantly to the development of CRT, albeit indirectly. His storytelling through songwriting and his interest in preserving cultural heritage had a profound impact.

One aspect of CRT influenced by Robertson's work is the importance of context and cultural relevance in assessment. Just as Robertson's songs often narrated stories deeply rooted in history and culture, CRT seeks to create assessments that resonate with students' backgrounds and experiences. By making assessments culturally relevant, educators can engage students more effectively and ensure a fair evaluation (Mason, 2020).

Moreover, Robertson's emphasis on collaboration among diverse talents within The Band parallels the collaborative nature of CRT development. CRT experts often collaborate with educators, psychologists, and content specialists to create assessments that accurately reflect learning objectives across various subjects and grade levels. Robertson's collaborative spirit serves as a reminder of the interdisciplinary nature of education and assessment (Jones, 2021).

### ***Conclusion***

The influence of Jerry Garcia and Robbie Robertson on Criterion-Referenced Testing may not be immediately apparent, but their impact on this field is undeniable. Garcia's improvisational approach and emphasis on individuality find resonance in the adaptive and student-centric aspects of CRT. Meanwhile, Robertson's dedication to storytelling and cultural relevance aligns with the importance of context and collaboration in CRT development.

This unexpected influence of musicians on education reminds us that inspiration can come from the most unlikely sources. It also underscores the need for educators and assessment experts to remain open to diverse perspectives and innovative ideas, as they continue to refine and enhance Criterion-Referenced Testing for the benefit of students and educators alike.



### **References:**

1. Smith, A. (2018). Jerry Garcia and the Art of Adaptive Testing. *Educational Innovations*, 42(3), 112-124.
2. Thompson, B. (2019). The Grateful Test: Exploring Jerry Garcia's Influence on Criterion-Referenced Testing. *Journal of Educational Psychology*, 35(2), 87-101.
3. Mason, R. (2020). Robbie Robertson and Cultural Relevance in Criterion-Referenced Testing. *Educational Assessment*, 28(4), 321-335.
4. Jones, S. (2021). The Band's Collaborative Spirit: Lessons for Criterion-Referenced Test Development. *Educational Research Quarterly*, 44(1), 56-69.

## Testing News

### Artificial Intelligence regulation and the implications on Test Use in Europe & beyond



Nigel Evans, C.Psychol.

On 13<sup>th</sup> March 2024 the European Parliament formally approved the AI Act which sets out clear requirements and obligations on the use of Artificial Intelligence. The EU AI regulation aims to ensure trust and safety in AI by defining levels of risk and banning unacceptable risk applications. This is direct relevance to any test developer and test user who involves AI within their process or system.

It will take two years for the AI Act to be fully implemented, and within that time, rules will be trialled to understand proportionality and effect – including those systems using AI in psychological assessment. *The trial will not be restricted to the European Union.* There is an 'AI Pact' bringing together EU and non-EU industry players in AI. More broadly most other countries are involved in the International Guiding AI Principles linking with G7, G20, OECD and UN initiatives.



In summary the risk categories cover:

**Unacceptable risk:** AI systems considered a clear threat to the fundamental rights of people **will be banned**. This includes AI systems or applications that manipulate human behaviour to circumvent users' free will.

**High-risk:** AI systems that cover: critical infrastructures; medical devices; systems to determine key citizen access; law enforcement and justice processes; biometric categorisation and emotion recognition systems.



**Specific transparency risk:** Users should be aware that they are interacting with a machine. Chatbots, deep fakes and other AI generated content will have to be labelled as such - users need to be informed.

**Minimal risk:** AI systems will fall into this category if they present only minimal or no risk for citizens' rights or safety. Applications such as AI-enabled recommender systems or spam filters will be given a 'free pass'.

Systems to determine access to educational institutions or for recruiting people are prime examples of high risk applications. So too are any system processes that link to determinations for health provision or access to essential services (like utilities, finance).

Designers and operators of these high-risk systems are now required to comply with strict requirements. These include: risk-mitigation systems, high quality of data sets, logging of activity, detailed documentation, clear user information, human oversight, and a high level of robustness, accuracy and cybersecurity.

It may strike Testing International readers that most, if not all, required elements of the new regulation parallel the expected quality methodology behind designing and using a good psychological test, especially when presented in a digital format.

It was this aspect that led the EFPA Board of Assessment (BoA) to respond to EU AI regulatory proposals in Aug 2021. Whilst welcoming more guardrails on AI use in assessment systems, the BoA urged the proposed areas of risk be proportionately categorised for existing psychological assessment across Work, Education and Health domains. For example, the use of a simple summative digital scoring 'algorithm' on a wellbeing screening tool could classify the test as AI driven, and thus high risk.

Some may see AI regulation as stifling innovation. However, with senior big tech figures commenting on being cautious developing AI, some balance needs to be found. No more so when systems significantly affect people lives as employees, students, patients or consumers.

#### Reference links

EU AI Act <https://europa.eu/IWxDJrP>

EFPA Board of Assessment AI responses and recommendations

[Board of Assessment | EFPA](#)

**Nigel Evans, C.Psychol** is Director of NEC Advisory, and Chair of the European Board of Assessment (EFPA). He advises on assessment innovation and strategy with Test Publishers, Consultancies, Growth Companies and Start Ups. [nigel@nigel-evans.com](mailto:nigel@nigel-evans.com)



## Personality inventory essentially contextualized in work



**Petteri Niitamo, PhD.**

Although the Work Personality Inventory (WOPI) was published some time ago (Nederström & Niitamo, 2010; Leung & Zedeck, 2016)) with over 200.000 individuals having taken the test in recruitment and coaching contexts, it remains fairly unknown beyond the Nordic-Baltic region. But more importantly, the test's new 2.5 edition takes a further step toward work contextualization by changing its technical profile terms, such as Focused achievement into personified terms, such as a Quality seeker. The change is presumed to enhance the inventory's utility in applied work settings by arousing individuals' interest through more relatable terms.

### THE PERSON-SITUATION CONTROVERSY

The controversy over persons vs. situations as predictors of behavior took place from the late sixties to early eighties and ended to at least a partial victory for the person viewpoint (Kenrick & Funder, 1988). However, situations are re-entering the scene along the notion that the influence of personality dispositions depends on each situation (Mischel & Shoda, 1995). The renewed interest in situations culminates in the concept of psychological situations presented in a recent handbook (Rauthmann et al., 2020).

The persisting question is how to align persons and situations in a meaningful way. Most personality measures used at work assess global personality constructs which are cross-situationally consistent rather than situationally specific. Organizational behaviors do not easily conceptually align with trait terms as in the Big Five. The issue has been tackled in different ways. The so-called trait activation theory (Tett et al., 2021) presents traits' relevance to situations as the key moderator of personality-performance relationships. Situational relevance is about offering opportunity for expression of personality dispositions. This is not a wholly new conceptualization as Murray's early taxonomy of needs (Murray, 1938) incorporated the concept of environmental "press" arousing needs and consequent behaviors. Another solution for rapprochement of personality dispositions with situations is to combine different scales into work-related composite scales (Ones & Viswesvaran, 2001) for example, in combining abstract perception and intuition for prediction of strategic planning.



The value of aligning personality with contexts receives empirical support from a meta-analysis showing that personality tests with scale items or instructions framed specifically to reference work-specific behaviors reached an average validity of .24 whereas the noncontextualized, global personality tests reached only a .11 level (Shaffer & Postlethwaite, 2012). Namely, context-specific personality items provide respondents with a common frame of reference unlike the traditional, noncontextual personality items. The common frame of reference works toward standardizing item interpretation which reduces measurement error and increases validity over noncontextual items. Similar prediction increments are evidenced in three meta-analyses where vocational interest dispositions were used in predicting

performance in interest-congruent situations (Nye et al., 2017).



In addition to maximizing predictive power, perhaps the more important reason for increased focus on situations is to better understand the evolving of behavior which enables more tangible development efforts. Situation fit has always been targeted by professionals in job analysis, personnel selection, training, leadership and similar endeavors although compelling conceptualizations of situations have been sparse. Conceivably the only globally well-known exception is Holland's RIASEC model (Holland, 1973) where six vocational interest dispositions are linked to corresponding interest-congruent occupations. Unlike clinical psychology where assessment instruments have preponderantly been developed from contexts (e.g., MMPI; Hathaway & McKinley, 1940), work psychology, perhaps also school and sports psychology have settled for global personality constructs instead of conceptualizing their own contexts.

### THINKING DISPOSITIONS TO RE-ENTER PERSONALITY

If personality factors are generally in need of alignment with situations, the ongoing disruption of work calls for an additional reorientation. Namely, "doing" things is increasingly becoming the responsibility of robots, algorithms and AI. The vastly changing information environment poses remarkable challenges to thinking about and understanding the new world of work. Occupations and jobs keep on disappearing and disrupting while new opportunities are arising. The emphasis is shifting toward handling information. The World Economic Forum's list of top ten skills needed in 2025 (WeForum, 2020) has the first five pertaining to thinking processes, from analytical thinking to creativity.

Cognitive styles or individual ways of processing information have led an orphan's life, occasionally seen as part of personality and as part of intelligence. Cognitive or thinking styles reflect the manner of performing, how people use their intelligence in contrast to intellectual abilities referring to individuals' level of performance. Despite that literally all leading scholars of personality, the American Psychological Association included (APA, 2000) incorporate the trio of behavior, thought and feeling as defining markers of personality. Still, the thought element in personality is severely under-represented if not missing in today's mainstream personality psychology.



The heyday of cognitive styles dates back to the period 1950-1970 with well-known exemplars such as personal constructs (Kelly, 1955), locus of control (Rotter, 1966) and field dependence-independence (Witkin & Goodenough, 1981). No obvious reason is available for the post-period waning of research interest in cognitive styles. It appears to relate to the modest empirical results at face of the preceding expectations. But, the criteria at that time were unfavorable for cognitive styles. Salary or position progress, grades or managerial appraisals fail to reflect the real value of the less visible thinking dispositions. Fortunately, the criteria have broadened from old times to include organizational citizenship, integrity, counterproductive behaviors, innovation and the like. In the future cognitive styles should earn their true value as moderator variables toward valuable outcomes such as innovation.

The renewing research interest in cognitive styles has targeted particularly on the dual-process theories (Evans & Stanovich, 2013) made known to the larger public by the Nobel laureate Daniel Kahneman's book on "fast" and "slow" thinking (Kahneman, 2011). The division of intuitive and analytical thinking serves as a

basis for a whole theory and test of personality (Epstein, 2003). Several meta-analyses have been published on analysis and intuition as predictors of performance in decision making (Phillips et al., 2016) and various work related tasks (Alaybek et al., 2022). Given the mainstream personality inventories' meager representation of cognitive styles, it is paradoxical that the allegedly most widely used personality test in the world, the MBTI (Briggs-Myers & Briggs, 1985) measures personality types based essentially on cognitive style constructs.

### THE WORK PERSONALITY INVENTORY

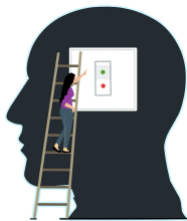
Most, if not all personality inventories used in work settings are global personality tests pitted against performance criteria in all potential applied contexts. Construction of the present Work Personality Inventory (WOPI) took went much further than the aforementioned ways in tackling the person-work alignment issue. Namely, work was taken as the departure point. Work was viewed as occurring in five universal and age-old forms: independent performance, leadership, collaboration, information processing and fit to the environment.

The various forms of work were defined as composed of so-called Basic competencies: independent performance is composed of quality and results seeking, leadership is composed of leading action and thoughts of others and collaboration is composed of communication, advisory and listening to others. Information processing or in more simple term, planning and problem solving was viewed as using two lanes, implementation of existing processes and creation of new processes. The process was defined as occurring along four consecutive steps, each marking a single competency: approaching, perceiving, producing solutions and implementing them. Finally, competent work was defined as occurring in work environments with recurrent and variable processes.



## SCALE CONSTRUCTION

Only after defining the Basic competency constructs, eyes were turned to psychology in order to find and when needed, to construct measurement scales for the constructs. The first three work forms and corresponding constructs pertain to behavior whereupon the choice fell on motives as construct drivers. The decision was also influenced by motivation's context dependency compared to context independent trait constructs (Alston, 1975; Buss & Craik, 1984). Murray's need taxonomy and David McClelland's theory of the "three big" motives (McClelland, 1987) served as the main sources for item writing. Cognitive styles among established constructs were chosen as drivers for the four planning and problem solving competencies. The fifth work form was defined as reflecting the realization of competent work in different environments. The choice here was to write work related items based on the tolerance of ambiguity construct. Finally, scales for optimism and for social desirability were constructed.



Construction of the scales began as a globally leading cell-phone manufacturer and a leading Nordic mobile operator opened their ongoing recruitment and training functions for carrying out the tedious and iterative scale construction process encompassing over five thousand real world subjects and over a thousand candidate items. The final self-report scales include altogether 224 self-descriptive statements. To diminish response style effects, the items are answered on a dichotomous (True/False) scale. The resulting raw scores (0-16) are transformed on the test profiles into Standard-and-ten scores (1-10).

## PSYCHOMETRICS

Reliability estimates were calculated for the final fourteen competency scales both for their internal consistency and retest or stability aspects. All scale reliabilities reached the level commonly required from measures used in individual differences research.



Validity evidence includes criterion studies on peer and spouse ratings, observed behavior in small groups, performance in managerial and non-managerial jobs, salary bonuses among production workers, etc. The criterion validity magnitudes are very much in line with those reported on more established personality inventories. Construct validity has been studied extensively through relations to well-validated neighboring constructs in personality, organizational and health psychology such as the Big Five, RIASEC, organizational culture, life values, learning styles, job burnout, coping with stress etc. The inventory can be responded currently in ten European languages adjusted through several national norms. More detailed empirical accounts are found in WOPI Technical Manual accessible at: <https://wopi.app/wopi/>

## PERSONIFIED COMPETENCY TERMS

The most recent (year 2024) step toward contextualization in work included changing the technical sounding dimension terms into 20 personified terms, "Experts in work" including: Quality vs. Results seeker, Action vs. Thought leader, Socializer, Advisor, Listener to others, Follower of one's own path, Fact-based, Idea generator, viewer of the Practical vs. Complex picture, Analytic vs. Intuitive thinker, Cautious vs. Quick implementer, proponent of Orderly vs. Variable work environment and Realist vs. Optimist, see example profiles: <https://wopi.app/pdf/WOPI-ExampleProfiles.pdf> .

While such personified terms offer a lure similar to personality typologies the scales retain their original dimensional structure. The renewed terms are presumed to boost the inventory's ecological validity, utility in applied settings and particularly in career coaching where the critical momentum is in getting people inspired by the question "Who am I at work"?

### About the author:

Petteri Niitamo

Competence Dimensions Ltd, Aalto University

<https://wopi.app/about/>

Vaakatie 53, 01640 Vantaa, Finland, Tel. +358 408364505,

[petteri.niitamo@wopi.net](mailto:petteri.niitamo@wopi.net)

### References

- Alston, W.P. (1975). Traits, consistency, and conceptual alternatives for personality theory. *Journal for the Theory of Social Behaviour*, 5, 17-48.
- Alaybek, B., Wang, Y., Dalal, R.S., Dubrow, S., Boemerman, L.S.G. (2022). The Relations of reflective and intuitive thinking styles with task performance: A meta-analysis. *Personnel Psychology*, Vol. 75, 2, 295-319.
- APA (2000). <https://www.apa.org/topics/personality>.
- Briggs-Myers, I., & Briggs, K.C. (1985). *Myers-Briggs Type Indicator (MBTI)*. Palo Alto, CA: Cons. Psych. Press.
- Buss, D.M. & Craik, K.H. (1984). Acts, dispositions and personality. In B.A. Maher & W.B. Maher (Eds.), *Progress in experimental personality research: normal personality processes* (pp. 241-301). New York, NY: Academic Press.
- Epstein, S. (2003). Cognitive-experiential theory of personality. In T. Millon & M.J. Lerner (Eds.), *Handbook of psychology: Personality and Social psychology*. (pp. 159-184). Hoboken, NJ: Wiley & Sons.
- Evans, J. St. B. T., & Stanovich, K.E. (2013). Dual process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8, 223-241.
- Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology*, 10, 249-254.
- Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Prentice Hall.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus & Giroux.
- Kelly, G. (1955). *The psychology of personal constructs*. New York: W. W. Norton & Company. OCLC 217761
- Kenrick, D.T., & Funder, D.C. (1988). Profiting from controversy: Lessons from the Person-Situation Controversy. *American Psychologist*, Vol. 43, No. 1, 23-34.
- Leung, S.A. | Zedeck, S. (2016). Reviews on the Work Personality Inventory. In Carlson, J.F., Geisinger, K.F & Jonson, J.L. (Eds.) *The Twentieth Mental Measurements Yearbook*. Buros: University of Nebraska Press, Nebraska. <https://marketplace.unl.edu/buros/work-personality-inventory-edition-1-0.html>
- McClelland, D.C. (1987). *Human motivation*. New York: Cambridge University Press.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102(2), 246-268. <https://doi.org/10.1037/0033-295X.102.2.246>
- Murray, H.A. (1938). *Explorations in personality*. New York, Oxford University Press.

### References ctd.

- Myers, I. B. (1962). *The Myers-Briggs Type Indicator: Manual*. Consulting Psychologists Press. <https://doi.org/10.1037/14404-000>
- Nederström, M. & Niitamo, P. (2010). Construction and validation of a work personality inventory. *Research Reports in Work Psychology and Leadership*, No. 2010/1. Helsinki University of Technology, Department of Industrial Engineering and Management. <https://aaltodoc.aalto.fi/server/api/core/bitstreams/032fe2fe-5543-45a0-9da6-507aba08ba27/content>
- Nye, C.D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior*, 98, 138-151.
- Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (COPS) used in personnel selection. *International Journal of Selection and Assessment*, 9(1-2), 31-39. <https://doi.org/10.1111/1468-2389.00161>
- Phillips, W.J., Fletcher, J.M, Marks, A.D.G., Hine, D.W. (2016). Thinking styles and decision making: A Meta-analysis. *Psychological Bulletin*, Mar; 142(3): 260-290.
- Rauthmann, J.F., Sherman, R.A. & Funder, D.C. (Eds.). (2020). *The Oxford Handbook of Psychological Situations*. Oxford University Press, New York, NY.
- Rotter, J.B. (1966). Generalized expectancies of internal versus external control of reinforcements. *Psychological Monographs*. 80 (whole no. 609).
- Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, 65(3), 445-494.
- Tett, R. P., Toich, M. J., & Ozkum, S. B. (2021). Trait activation theory: A review of the literature and applications to five lines of personality dynamics research. *Annual Review of Organizational Psychology and Organizational Behavior*, 8(1), 199-233. <https://doi.org/10.1146/annurev-orgpsych-012420-062228>
- WeForum (2020) <https://www.weforum.org/agenda/2020/10/top-10-work-skills-of-tomorrow-how-long-it-takes-to-learn-them/>
- Witkin, Herman A., and Donald Goodenough. *Cognitive Styles: Essence and Origins*. New York: International Universities Press, 1981.
- WOPI Technical Manual (2010). Competence Dimensions. [www.wopi.app](http://www.wopi.app)

**If you have something you'd like to discuss with other test professionals, or anything else you'd like to share, do get in touch.**

This is your newsletter, and what you think is always interesting! Just send your article or idea to: [newsletter@intestcom.org](mailto:newsletter@intestcom.org)

## The ITC publishes several Guidelines For Professional Practice.

International Test Commission. (2018). ITC Guidelines for Translating and Adapting Tests (2<sup>nd</sup> Ed.). *International Journal of Testing*, 18 (2), 101-134.

[[https://www.intestcom.org/files/guideline\\_test\\_adaptation\\_2ed.pdf](https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf)]

International Test Commission (2001). International Guidelines for Test Use, *International Journal of Testing*, 1(2), 93-114.

[[https://www.intestcom.org/files/guideline\\_test\\_use.pdf](https://www.intestcom.org/files/guideline_test_use.pdf)]

International Test Commission. (2014). ITC Guidelines on Quality Control in Scoring, Test Analysis, and Reporting of Test Scores. *International Journal of Testing*, 14 (3), 195-217,

[[https://www.intestcom.org/files/guideline\\_quality\\_control.pdf](https://www.intestcom.org/files/guideline_quality_control.pdf)]

International Test Commission. (2015). International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal.

[[https://www.intestcom.org/files/guideline\\_test\\_disposal.pdf](https://www.intestcom.org/files/guideline_test_disposal.pdf)]

International Test Commission. (2016). The International Test Commission Guidelines on the Security of Tests, Examinations, and Other Assessments. *International Journal of Testing*, 16 (3), 181-204.

[[https://www.intestcom.org/files/guideline\\_test\\_security.pdf](https://www.intestcom.org/files/guideline_test_security.pdf)]

International Test Commission. (2019). ITC Guidelines for the Large-Scale Assessment of Linguistically and Culturally Diverse Populations. *International Journal of Testing*, 19 (4), 301-336.

[[https://www.intestcom.org/files/guideline\\_diverse\\_populations.pdf](https://www.intestcom.org/files/guideline_diverse_populations.pdf)]

International Test Commission and Association of Test Publishers (2022). Guidelines for technology-based assessment. <https://www.intestcom.org/page/28> and <https://www.testpublishers.org/white-papers>

If you need further information, please contact the Secretary-General of the ITC (Dr. April Zenisky) at: [Secretary@intestcom.org](mailto:Secretary@intestcom.org)