

FOR PUBLICATION

**UNITED STATES COURT OF APPEALS
FOR THE NINTH CIRCUIT**

STEPHANIE GARCIA,
Plaintiff-Appellant,

v.

COMMISSIONER OF SOCIAL
SECURITY,

Defendant-Appellee.

No. 12-15686

D.C. No.
1:10-cv-01997-
JLT

OPINION

Appeal from the United States District Court
for the Eastern District of California
Jennifer L. Thurston, Magistrate Judge, Presiding

Argued and Submitted
February 14, 2014—San Francisco, California

Filed September 23, 2014

Before: Alex Kozinski, Chief Judge, and Diarmuid F.
O’Scannlain and Mary H. Murguia, Circuit Judges.

Opinion by Judge Murguia;
Dissent by Judge O’Scannlain

SUMMARY*

Social Security

The panel reversed the district court's order affirming the Commissioner of Social Security's denial of benefits on the basis that the claimant was not intellectually disabled.

The panel held that the administrative law judge ("ALJ") had a duty to order further IQ testing, and concluded that the ALJ's failure to do so was error that cannot be considered harmless. The panel remanded to the district court with instructions to reverse the final decision of the Commissioner and to order the Commissioner to develop the record through further IQ testing.

Judge O'Scannlain dissented because he believed that the majority erroneously presumed that the Commissioner's ostensible error prejudiced the claimant. Judge O'Scannlain also was unconvinced that the ALJ erred by not ordering a new and complete round of IQ tests.

COUNSEL

Lawrence David Rohlfig (argued), Law Offices of Lawrence D. Rohlfig, Santa Fe Springs, California, and Cyrus Safa, Grancell, Stander, Reubens, Thomas and Kinsey, San Diego, California, for Plaintiff-Appellant.

* This summary constitutes no part of the opinion of the court. It has been prepared by court staff for the convenience of the reader.

Donna Wade Anderson (argued), Supervisory Attorney, and Patrick William Snyder, Special Assistant United States Attorney, Social Security Administration Office of General Counsel, San Francisco, California, for Defendant-Appellee.

OPINION

MURGUIA, Circuit Judge:

Stephanie Garcia appeals from the district court's order affirming the Commissioner of Social Security's (the "Commissioner") denial of benefits on the basis that she was not intellectually disabled. Garcia argues that the administrative law judge (ALJ) who determined that she was not disabled had a duty to develop the record because that record did not include a complete set of valid IQ scores. We agree that the ALJ had a duty to order further IQ testing, and we further conclude that the ALJ's failure to do so was an error that cannot be considered harmless. We therefore reverse the district court and remand for further proceedings.

I

As a minor, Stephanie Garcia received social security benefits because of her intellectual disability. After she reached the age of 18 in 2007, the Social Security Administration (SSA or the "Administration") concluded that she no longer qualified as disabled and was therefore not entitled to further benefits. Garcia sought review by an ALJ, before whom she had a hearing on April 8, 2010. At the time of her hearing, Garcia lived with her mother and two siblings, as well as her own disabled daughter. Although she had learned some skills for caring for herself through an

independent living program, Garcia was dependent on her mother for her own care and for the care of her child. After taking special education classes, Garcia earned a high school diploma, but she was unable to read and did not know the alphabet.

Garcia worked part-time at a pizza shop for several months in 2008. She testified to having had difficulty with making pizzas, taking orders, and cashiering; as a result, she required constant supervision. She quit because she found the work “too hard.” Garcia was then placed in a clerical job by the California Department of Rehabilitation; her duties included photocopying, alphabetizing files, and removing staples from documents. She worked four or five hours per day, five days per week. She testified at her hearing that she had difficulty understanding how to perform the tasks assigned to her and had to rely on a coworker for help. Garcia also quit this job after two months because “[i]t was too hard.” Vicky Medina, Garcia’s counselor at the Central Valley Regional Center, testified that, based on her observations, Garcia would be unable to “do any job eight hours a day, five days a week as it would be performed in the national economy without extra supervision.” Medina explained that Garcia has difficulty remembering how to perform tasks, and that she needs to be re-taught “on a constant basis.”

Apart from her intellectual disability, Garcia has suffered from depression stemming from having to care for her young daughter, who has Down Syndrome, asthma, and heart and thyroid problems. Garcia has been treated for her depression, and her psychiatric condition has improved.

In evaluating Garcia's disability claim, the ALJ considered the reports of three experts: psychologist Mary K. McDonald, Ph.D., psychologist Allen Middleton, Ph.D., and physician Evangeline Murillo, M.D.

On February 13, 2008, Dr. McDonald evaluated Garcia at the request of the California Department of Social Services. Dr. McDonald administered the Bender Visual Motor Gestalt Test, II Edition; the Wechsler Memory Scale, III Edition; and the Wechsler Adult Intelligence Scale, III Edition ("WAIS-III"). The WAIS-III measures an individual's "intelligence quotient," or "IQ"; IQ is reported as three scores: verbal, performance (non-verbal), and full scale. See 20 C.F.R. § 404, subpt. P, app. 1, listing 12.00 ("Listing 12.00") (D)(6). Garcia's scores on the Motor Gestalt Test were average to low average, and her Memory Scale scores indicated that her "[v]erbal memory is impaired and visual memory is within the low average range."

Dr. McDonald administered only the performance portion of the WAIS-III "[d]ue to the constraints of time and the slowness with which [Garcia] worked."¹ Consequently, Dr.

¹ This is not that first time that Dr. McDonald has given this reason for failing to administer a complete IQ test when evaluating a patient for intellectual disability. See *Andrade v. Comm'r of Soc. Sec.*, No. 1:09-cv-1926 GSA, 2011 WL 864700 (E.D. Cal. Mar. 10, 2011), *aff'd*, 474 F. App'x 642 (9th Cir. 2012) ("Dr. McDonald's report indicates that only the Performance IQ portion of the Wechsler Adult Intelligence Scale was administered 'due to the constraints of time.'"). This excuse is troublesome, and the district court should not have accepted it in the absence of some more compelling reason. The SSA's regulations indicate that potentially disabled individuals may take more time than others to complete an IQ test administration, and the administrator of the test should plan accordingly. See 20 C.F.R. § 416.919n(a).

McDonald did not report a verbal or full-scale score. Garcia's performance IQ score was 77, which is in the "borderline range" for disability. McDonald concluded that Garcia was "capable of employment."

After reviewing Garcia's medical records, including the incomplete IQ test results, Dr. Middleton completed a Mental Residual Functional Capacity Assessment,² Psychiatric Review Technique,³ and Case Analysis.⁴ He determined that Garcia was "moderately limited" in her "ability to [understand, remember, and carry out] detailed instructions." He concluded that Garcia was "able to understand and remember [work] locations [and] procedures of a simple, routine nature involving 1–2 step job tasks [and] instructions."

² Residual Functional Capacity (RFC) is the work that an individual is capable of performing in spite of her limitations. 20 C.F.R. § 416.945(a)(1). The Mental RFC Assessment form used by Dr. Middleton, Form SSA-4734-SUP, requires a reviewing expert to evaluate the degree of the subject's limitations in various aspects of (1) "understanding and memory," (2) "sustained concentration and persistence," (3) "social interaction," and (4) "adaption," such as in responding to workplace hazards or navigating public transportation. Based on the evaluation of the subject's limitations in each category, the reviewing expert then makes a general assessment of the subject's "functional capacity."

³ The Psychiatric Review Technique form used by Dr. Middleton, Form SSA-2506-BK, requires the reviewing expert to (1) summarize relevant documentation, such as IQ test results, (2) rate the subject's "functional limitations," and (3) provide additional notes in narrative form.

⁴ Dr. Middleton used Form SSA-416, on which he listed "significant objective findings," such as Garcia's IQ test scores, her progress in school, and her depression.

Dr. Murillo also reviewed Garcia's medical records, including the incomplete IQ results, and completed a Mental Residual Functioning Capacity Assessment and Case Analysis.⁵ Like Dr. Middleton, Dr. Murillo concluded that Garcia was "moderately limited" in her "ability to [understand, remember, and carry out] detailed instructions." She determined that Garcia could "understand and remember work locations and procedures of a simple, routine nature involving 1–2 step job tasks and instructions" and "maintain concentration and attention for above in 2 hour increments" during "8 hr/40 hr work schedules."

At the hearing, the ALJ also heard testimony from vocational expert Thomas Dachelet. Dachelet testified that the ability to read and write at a basic level is a requirement for even those jobs classified by the Dictionary of Occupational Titles (DOT) as needing the lowest "general educational development." However, he also acknowledged that Garcia had worked at "light unskilled" jobs at which "she didn't read or write." Dachelet testified that in California "there were 1,020,830 persons employed at the light unskilled level." He identified three light unskilled jobs Garcia could perform: (1) a bagger, of which 44,304 were employed in California, (2) a garment sorter, of which 21,179 were

⁵ Dr. Murillo completed the same forms as Dr. Middleton: Mental RFC Assessment Form SSA-4734-SUP and Case Analysis Form SSA-416.

employed in California, and (3) a grader,⁶ of which 20,188 were employed in California.

In a May 18, 2010, decision, the ALJ concluded that Garcia was not disabled as of February 1, 2008, consistent with the SSA's original determination. The ALJ determined that Garcia had the severe impairment of borderline intellectual functioning but that the impairment was not so severe that it met the requirements for intellectual disability; *see* 20 C.F.R. § 404, subpt. P, app. 1, listing 12.05 ("Listing 12.05").

Listing 12.05 lays out four ways in which an individual may qualify as intellectually disabled without requiring any further inquiry into her ability to work: (1) "[m]ental incapacity . . . such that the use of standardized measures of intellectual functioning is precluded"; (2) "[a] valid verbal, performance, or full scale IQ of 59 or less"; (3) "[a] valid verbal, performance, or full scale IQ of 60 through 70 and a physical or other mental impairment imposing an additional and significant work-related limitation of function"; and (4) "[a] valid verbal, performance, or full scale IQ of 60 through 70, resulting in at least two [milder impairments]." *Id.* Each of these alternatives depends on a subject's IQ test performance, unless she is unable to undergo testing.

⁶ Dachelet refers to the DOT listing for a "fruit-grader operator." One employed in this position "[t]ends machine that grades fruit according to size: Changes chains and other driving gear according to type of fruit. Directs workers engaged in loading of elevator belt and removal of graded fruit. Cleans and lubricates chains, bearings, and machine gears, using rags and grease gun. Repairs, replaces, and adjusts malfunctioning parts of machine." DOT 529.665-010, 1991 WL 674628.

Based on Garcia's performance IQ score of 77, the ALJ concluded that Garcia could not meet Listing 12.05. The ALJ further concluded that Garcia had the RFC "to perform a full range of work at all exertional levels but with the following nonexertional limitations: [Garcia] can perform simple repetitive tasks where the jobs can be learned mostly by demonstration, but she cannot perform reading and/or writing as a job task." Based primarily on Dachelet's testimony, the ALJ concluded that Garcia was "capable of making a successful adjustment to other work that exists in significant numbers in the national economy," including the jobs of bagger, garment sorter, and grader. For this reason, the ALJ concluded that Garcia was "not disabled."

Garcia appealed the ALJ's decision to the Social Security Administration Appeals Council, but her appeal was denied, making the ALJ's decision the final decision of the Commissioner. Garcia then sought judicial review of the Commissioner's decision in the district court, arguing in part that the ALJ erred when she failed to develop the record by ordering a new IQ test administration to obtain a complete set of test scores. The district court affirmed the final decision of the Commissioner.

II

We review *de novo* a district court's judgment affirming the denial of social security benefits. *Bray v. Comm'r of Soc. Sec. Admin.*, 554 F.3d 1219, 1222 (9th Cir. 2009). "We may set aside a denial of benefits only if it is not supported by substantial evidence or is based on legal error." *Robbins*, 466 F.3d at 882.

It was legal error for the ALJ not to ensure that the record included a complete set of IQ test results that both the ALJ and the reviewing experts could consider. While it is not certain from the record before us that Garcia would have been determined to be disabled if the record had been properly developed, it is also not “clear from the record that ‘the ALJ’s error was inconsequential to the ultimate nondisability determination.’” *Tommasetti v. Astrue*, 533 F.3d 1035, 1038 (9th Cir. 2008) (quoting *Robbins v. Soc. Sec. Admin.*, 466 F.3d 880, 885 (9th Cir. 2006)). Therefore we reverse the district court and remand with instructions to reverse the final decision of the Commissioner and to order the Commissioner to develop the record through further IQ testing.

III

To be eligible for disability benefits, an individual must be unable “to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months.” 42 U.S.C. § 423(d)(1)(A).

The evaluation of disability in adults is governed by a five-step process, which the ALJ followed in assessing Garcia. 20 C.F.R. § 416.920. The ALJ skipped the first and fourth steps, as they were not applicable to Garcia’s situation.⁷ At the second step, the ALJ determines whether a

⁷ At the first step, the ALJ would have considered Garcia’s present work activity; however, this step does not apply to individuals whose disability determinations are being reevaluated because they turned 18. See 20 C.F.R. § 416.987(b). At the fourth step, the ALJ would have considered Garcia’s past relevant work, *id.* § 920(a)(iv); however, the ALJ skipped

claimant has an impairment or combination of impairments that is medically severe; if not, the claimant is not disabled. *Id.* §§ 416.920(a)(4)(ii), 416.920(c). The ALJ concluded that Garcia had the severe impairment of “borderline intellectual functioning,” and so proceeded to the third step.

At the third step, the ALJ again considers the severity of the impairment or combination of impairments by comparing it to the listings in 20 C.F.R. § 404, subpart P, appendix 1. *Id.* §§ 416.920(a)(4)(iii), 416.920(d). If the impairment or combination of impairments is at least as severe as the relevant listing, and has lasted at least twelve months, then the claimant is deemed disabled, and the inquiry ends; otherwise, the ALJ proceeds to the next step. *Id.* The ALJ concluded that Garcia did not meet Listing 12.05 and so proceeded to step five. At the fifth step, the ALJ considers the claimant’s RFC – that is, her ability to work in spite of her limitations – along with her age, education, and work experience, to determine whether she can make an adjustment to a new kind of work. *Id.* § 416.920(a)(4)(v). The ALJ concluded that Garcia could perform jobs requiring the ability to undertake simple, repetitive tasks, and so found that she was not disabled.

IV

Garcia argues that the ALJ erred by failing to order additional IQ testing and instead relying on the results of the partial examination performed by Dr. McDonald. We agree. “The ALJ always has a ‘special duty to fully and fairly develop the record and to assure that the claimant’s interests

this step because she concluded that Garcia did not have any past relevant work.

are considered.” *Celaya v. Halter*, 332 F.3d 1177, 1183 (9th Cir. 2003) (quoting *Brown v. Heckler*, 713 F.2d 441, 443 (9th Cir. 1983)).

The ALJ is not a mere umpire at such a proceeding . . . : it is incumbent upon the ALJ to scrupulously and conscientiously probe into, inquire of, and explore for all the relevant facts. He must be especially diligent in ensuring that favorable as well as unfavorable facts and circumstances are elicited.

Id. (quoting *Higbee v. Sullivan*, 975 F.2d 558, 561 (9th Cir. 1992)).

In a case, such as this one, that turns on whether a claimant has an intellectual disability and in which IQ scores are relied upon for the purpose of assessing that disability, there is no question that a “fully and fairly develop[ed]” record, *id.*, will include a complete set of IQ scores that report verbal, non-verbal, and full-scale abilities. There are two principal reasons for our conclusion.

First, IQ testing plays a particularly important role in assessing the existence of intellectual disability. Listing 12.00 generally lays out the necessary procedures for evaluating mental disorders, including intellectual disability, and for documenting relevant objective findings. In that listing the SSA has recognized that “[s]tandardized intelligence test results are essential to the adjudication of all cases of intellectual disability,” except where a claimant is unable to complete such testing. Listing 12.00(d)(6)(b). At the third step of the SSA’s five-step process, when a

claimant's impairment is compared to the criteria in Listing 12.05, three of the four criteria for intellectual disability rely in whole or in part on IQ test scores. (The fourth criterion applies when the claimant's incapacity precludes IQ testing.) Because meeting the relevant listing conclusively determines that a claimant is indeed disabled, 20 C.F.R. § 416.920(a)(4)(iii), the claimant's IQ score can be the deciding factor in a determination of intellectual disability.

Further, as was the case with Garcia, IQ test results can play a role in the development of other evidence in the record. For example, Dr. Middleton and Dr. Murillo both reviewed Garcia's IQ results before making their determinations about her ability to work. Thus, as a practical matter, the importance of IQ scores in this case did not end with step three. The partial test results also affected the ALJ's conclusions about Garcia's ability to work, even if less directly.

The second reason for our conclusion is that the regulations promulgated by the SSA demonstrate that the Administration, based on its considerable expertise, has determined that it is essential for complete – rather than partial – sets of IQ scores to be used in evaluating intellectual disability. As a general principle, all reports of test results “must conform to accepted professional standards and practices in the medical field for a complete and competent examination,” 20 C.F.R. § 416.919n(b), and an examination is not complete unless it includes “all the elements of a standard examination in the applicable medical specialty,” *id.* § 416.919n(c).

The regulations specifically identify the “Wechsler series” of IQ tests (of which WAIS-III is a part) as

“customarily” including “verbal, performance, and full scale IQs.” Listing 12.00(D)(6)(c). This characteristic of the Wechsler exam makes it particularly well suited to the assessment of intellectual disability, because “[g]enerally, it is preferable to use IQ measures that are wide in scope and include items that test both verbal and performance abilities.” Listing 12.00(D)(6)(d).

The Commissioner argues that the regulations themselves suggest it is acceptable for an ALJ to rely on partial test results in a situation, such as this one, in which only part of an IQ test was administered. The Commissioner points specifically to a passage in Listing 12.00 providing that “[i]n cases where more than one IQ is customarily derived from the test administered, *e.g.*, where verbal, performance, and full scale IQs are provided in the Wechsler series, we use the lowest of these in conjunction with [Listing] 12.05.” *Id.* at 12.00(D)(6)(c).⁸

However, our reading of this same passage leads us to conclude the opposite: Listing 12.00 strongly disfavors reliance on partial test results. The plain text of the regulation clearly suggests that IQ tests like those in the Wechsler series should be administered and reported in full, because it assumes that the ALJ will have multiple scores – “verbal, performance, and full scale” – from which to “use the lowest.” We also note that the regulations’ insistence that the ALJ look at all three scores in order to identify the lowest among them seems intended to benefit the disability claimant, for whom each test score is an opportunity to demonstrate that she meets one of the IQ-related criteria specified in Listing 12.05 – as well as an opportunity to demonstrate the

⁸ The district court came to the same conclusion.

extent of her impairment to other experts reviewing her IQ as part of their own evaluations of her limitations.

Because the regulations clearly assert the importance of a complete IQ test administration, the ALJ had a duty to develop the record so that it included a complete set of IQ test results. Her failure to do so was legal error.⁹

V

Our conclusion that the ALJ committed legal error is not the end of our inquiry. We will not reverse an ALJ's decision on the basis of a harmless error, "which exists when it is clear from the record that 'the ALJ's error was inconsequential to the ultimate nondisability determination.'" *Tommasetti*, 533 F.3d at 1038 (quoting *Robbins*, 466 F.3d at 885). While the record here may not definitively demonstrate that Garcia would have been adjudicated disabled if the ALJ had ordered that a complete set of IQ tests be administered, it is certainly not clear from the record that Garcia was not harmed by the ALJ's error.¹⁰

⁹ We recognize that our holding here is contrary to *Andrade v. Commissioner of Social Security*, 474 F. App'x 642 (9th Cir. 2012). We are not bound by our earlier decision. See 9th Cir. R. 36-3(a).

¹⁰ The dissent suggests that the harmless standard recognized in *Tommasetti* does not apply to cases in which the legal error at issue is a failure of the duty to develop the record. Citing *McLeod v. Astrue*, 640 F.3d 881 (9th Cir. 2010), the dissent argues that in such cases we should turn our stringent harmless standard on its head and presume any error is harmless until the claimant or record demonstrates otherwise. See Dissent at 20, 25–26. *McLeod* provides no basis for us to create such a peculiar carve-out from our well-established rule. We have consistently treated an ALJ's failure to adequately develop the record as reversible legal error. See *Celaya*, 332 F.3d at 1183. We have never suggested that

Again, we recognize that the importance of IQ test results in adjudicating intellectual disability is not limited to the claimant's ability to meet the listing at step three of the five-step process. Both Dr. Middleton and Dr. Murillo considered Garcia's incomplete IQ test results in assessing her ability to support herself through gainful employment, and the ALJ relied on these experts' findings in assessing Garcia's RFC and ultimately in determining that she was not disabled. The Commissioner points out that neither Dr. Middleton nor Dr.

failure to develop is somehow lesser error, or should be treated differently to other types of legal error. Indeed, often the same error can be characterized as either failure-to-develop or "normal" legal error depending on how it's described. Adopting a separate—and inverted—harmlessness standard for failure-to-develop cases would not only create confusion in our case law, but also hinge a great deal on a nebulous, and often unimportant, distinction.

McLeod concerned a disability claim by a veteran who argued on appeal that the ALJ had failed adequately to develop the record. We observed that there may be situations in which "further administrative review is needed to determine whether there was prejudice from the error [of not developing the record]." 640 F.3d at 888. However, contrary to the dissent's assertion, we explicitly recognized that "it is quite clear that no presumptions operate, and we must exercise judgment in light of the circumstances of the case." *Id.* We remanded to the ALJ for a harmlessness determination, even though it was not clear from the record that the potentially omitted evidence—a VA disability rating—even existed. *McLeod* is limited to situations where the record is insufficient for the court to make its own prejudice determination, and remand is required for the ALJ to determine the harmfulness of the omission in the first instance. It makes good sense that, in such a situation, "mere probability" that hypothetical new evidence—like the potential disability certificate—may be influential is insufficient to support a remand. Because, here, we know precisely which evidence was omitted from the record and have no doubts about its significance in reaching an intellectual disability determination, we see no reason to depart from the harmlessness standard articulated in *Tommasetti*.

Murillo “expressed any concerns about the adequacy of Dr. McDonald’s psychological testing,” but that does not necessarily mean that neither would have reached a different conclusion or offered other findings beneficial to Garcia based on a complete set of scores. Such an outcome seems particularly plausible where, as here, Garcia’s testing history as a juvenile strongly suggests that her verbal and full-range IQ scores would be considerably lower than the performance score of 77 obtained by Dr. McDonald. In a December 2004 test administration, Garcia was assessed with a verbal score of 61, a performance score of 74, and a full-scale score of 66. In June 2005, she received a full-scale score of 44 and a verbal score of 53. Further, the testimony of Garcia’s counselor Vicky Medina also suggests that verbal functioning was a particular weakness for Garcia.

In this case, there is a genuine probability that, had a complete set of valid IQ test scores been included in the record, the opinions of the reviewing experts might have been different, or Garcia might have had an additional factual basis for challenging their opinions. This is especially true when, just three years earlier, Garcia’s full-scale test score was dramatically below the threshold for establishing disability even on the basis of just the score by itself. *See* Listing 12.05(B) (providing that intellectual disability may be established by “[a] valid verbal, performance, or full-scale IQ of 59 or less”). The fact that IQ test results may be considered by multiple reviewing experts, as well as by the ALJ, makes it particularly difficult to conclude that any error affecting the quality of those results is “inconsequential to [an] ultimate nondisability determination,” let alone to conclude that such harmlessness is “clear from the record.” *Tommasetti*, 533 F.3d at 1038.

Perhaps even more significantly, Garcia may have been able to meet Listing 12.05(B),¹¹ under which she would have been adjudicated disabled if she had scored below 60 on either the verbal, performance, or full-scale portion of an IQ test. Given that Garcia had previously received a childhood Wechsler full-scale score of 44 and a verbal score of 55, and that she tended to score lower on the verbal component than on the performance component, it appears likely that Garcia could have met Listing 12.05(B) at step three of the evaluation process. Based on that evidence alone, it cannot be “clear from the record” that failure to obtain those two tests was “inconsequential.” *Tommasetti*, 533 F.3d at 1038.

VI

The ALJ’s failure to develop the record to include a complete set of IQ scores was legal error. Because we cannot conclude that the error was harmless, we **REVERSE** the judgment of the district court and **REMAND** with instructions to remand to the Commissioner for further proceedings.

¹¹ The dissent argues we should ignore Listing 12.05(B) when reviewing for harmless error because Garcia “*never* claimed on appeal that she would have qualified under Listing 12.05 B.” Dissent at 25 (emphasis in original). Garcia’s opening brief, however, clearly raised the issue. Garcia argued that “[b]ased on the high correlation between the tests, the expected verbal IQ score supports the contention that the complete IQ test would result in IQ scores sufficient to meet or equal the Listing . . . 12.00.” Listing 12.00 describes the evaluation process to determine whether an applicant’s impairment is a “mental disorder.” It expressly states: “If your impairment satisfies the diagnostic description in the introductory paragraph [of Listing 12.05] and *any one of the four sets of criteria*, we will find that your impairment meets the listing.” Listing 12.00 (emphasis added).

O'SCANNLAIN, Circuit Judge, dissenting:

The panel majority, eager to reprimand the Commissioner of Social Security for what it deems to be inexcusably sloppy practices, disregards—I suggest, respectfully—the deference we owe under law to the agency's determinations. Rather than observing the standard for harmless error that our precedents have previously prescribed, the majority has erroneously *presumed* that the Commissioner's ostensible error has prejudiced Stephanie Garcia, the claimant in this case. I respectfully dissent from this regrettable exaggeration of our Court's properly limited role in the adjudication of Social Security disability benefits claims.

I

Congress has carefully prescribed a minimal role for the Federal courts in adjudicating claims of disability under the Social Security Act. *See* 42 U.S.C. § 405(g). Accordingly, we have only limited authority to nullify the decisions of the agency and its administrative law judges with which we disagree. As the majority opinion correctly notes, we may not disturb an ALJ's denial of benefits unless “it is not supported by substantial evidence or is based on legal error.” *Robbins v. Soc. Sec. Admin.*, 466 F.3d 880, 882 (9th Cir. 2006). Legal error alone, furthermore, is not sufficient to warrant our interference: for example, we generally must stay our hand if it is “clear from the record” that any ostensible error “was inconsequential to the ultimate nondisability determination.” *Tommasetti v. Astrue*, 533 F.3d 1035, 1038 (9th Cir. 2008) (internal quotation marks omitted).

Indeed, one such error that we have identified in past cases has been an ALJ's failure “fully and fairly [to] develop

the record and to assure that the claimant's interests are considered." *Celaya v. Halter*, 332 F.3d 1177, 1183 (9th Cir. 2003). This "special" and "independent" duty of the ALJ exists in all circumstances, although, when the applicant is uncounseled, the responsibility to ensure an adequate record is heightened. *See Tonapetyan v. Halter*, 242 F.3d 1144, 1150 (9th Cir. 2001); *Smolen v. Chater*, 80 F.3d 1273, 1288 (9th Cir. 1996). Despite our solicitude in this regard, we have nevertheless clearly limned the outer boundaries of such responsibility. "An ALJ's duty to develop the record further is triggered *only* when there is ambiguous evidence or when the record is inadequate to allow for proper evaluation of the evidence." *Mayes v. Massanari*, 276 F.3d 453, 459–60 (9th Cir. 2001) (emphasis added).

More recently, we have refined—in the context of the ALJ's duty to develop the record—the standard by which we appraise whether any such error prejudiced the claimant. In *McLeod v. Astrue*, the unsuccessful applicant for disability benefits contended that the "ALJ erred by failing to develop the record adequately," specifically by not "request[ing] more explanation from two of his treating physicians" and by not obtaining "whatever VA disability rating" he may have had. 640 F.3d 881, 884 (9th Cir. 2011). We determined that the ALJ had shirked this duty to develop the record, but nevertheless that this dereliction was not alone sufficient warrant for reversal. Rather, we explained that "the burden is on the party attacking the agency's determination to show that prejudice resulted from the error." *Id.* at 887. But "where the circumstances of the case show a substantial likelihood of prejudice," the reviewing court can remand the case so the agency may reconsider the claimant's eligibility for benefits. *Id.* at 888. We emphasized, nevertheless, that a "mere probability" of prejudice "is not enough." *Id.* Either

the claimant must himself shoulder the burden of demonstrating prejudice, or otherwise such prejudice must be apparent on the face of the record or the “circumstances of the case.”

II

The majority’s opinion turns this duty-to-develop doctrine on its head. Even assuming, *arguendo*, that the ALJ committed legal error by not ordering Dr. McDonald to perform another round of IQ tests on Miss Garcia,¹ the

¹ I remain unconvinced that, at least in the circumstances of this case, the ALJ erred by not ordering a new, and complete, round of IQ tests. The majority opinion does not assert that the partial test scores constitute “ambiguous evidence” or make the “record . . . inadequate” for the purposes of assessing residual functional capacity. *See Mayes*, 276 F.3d at 460.

At most, the majority opinion gleans from the regulations an expectation of or a preference for “multiple scores” from a Wechsler series IQ test, *maj. op.* at 14. Whether such regulatory intimations can “trigger[]” the ALJ’s duty further to develop the record, 276 F.3d at 459, does not appear compelled by our precedents. And the majority does not pause to explain why.

Furthermore, the majority scarcely indicates what countervailing constraints—if any—may defeat the regulations’ preference for or expectation of multiple test scores. Dr. McDonald’s purported reasons for not administering the complete Wechsler series IQ test were “the constraints of time and the slowness with which [Miss Garcia] worked.” The majority simply deems this explanation an “excuse,” dismissing it as “troublesome” and scolding the district court, which in its judgment “should not have accepted it in the absence of some more compelling reason.” *Maj. op.* at 5–6 & n.1.

I strongly resist this lecture to medical practitioners. Not only does the record lack any clear implication of either excuse-making or duty-

majority misstates—and misapplies—the proper standard for assessing any prejudice such error caused.

In the first place, the majority correctly acknowledges that “[w]e will not reverse an ALJ’s decision on the basis of a harmless error,” which occurs “when it is clear from the record that the ALJ’s error was inconsequential to the ultimate nondisability determination,” maj. op. at 15 (internal quotation marks omitted). Although the majority does not expressly state that such rule is the *exclusive* standard by which to assess the harm caused by an error, its reasoning assumes so. For the majority detects prejudice in “a genuine probability” that a complete set of IQ test scores may have altered the medical reports or provided another basis for Miss Garcia to challenge the ALJ’s determination. *Id.* at 17. *McLeod*, however, specifically forecloses this basis for reversing a denial of benefits: a “mere probability,” no matter how “genuine,” simply does not suffice. 640 F.3d at 888. The majority articulates an exclusive standard for harmless error that *presumes* prejudice unless such error appear “inconsequential” on the face of the record. Such may be the ordinary analysis for determining the prejudice caused by legal error. In the special context of the ALJ’s duty to develop the record, however, our Court has already clearly explained that we cannot find prejudice unless and until

shirking, but also it is not self-evident that the time Dr. McDonald did devote to administering the tests and interviewing Miss Garcia was insufficient or otherwise imprudent. We should be reticent to craft, in footnotes to our opinions, legal rules governing the minutiae of medical practice—such as how and when to schedule tests and interviews—where Congress has not legislated and where the agency has not regulated. And especially not where the record and the parties’ briefings do not present an adequate basis for determining which sort of constraints are reasonable and which are merely “excuses.”

demonstrated by the claimant or the record and circumstances of the case.

Furthermore, the majority offers no basis, either in law or in fact, for simply *asserting* that the absence of a full set of IQ test scores would have had any likely effect on the ALJ's disability determination. The majority first observes that “[b]oth Dr. Middleton and Dr. Murillo considered Garcia’s incomplete IQ test results in assessing her ability to support herself through gainful employment.” Maj. op. at 16. Indeed, the medical experts considered the test scores—but they also considered sundry other relevant data, such as her employment history, educational and recreational activities, financial independence, grooming, and the cooperation and comprehension she displayed during her clinical evaluation. The majority does not indicate any basis from these experts’ reports that the partial test scores figured decisively in their recommendations. Nor does the majority opinion advert to any item in the record or the “circumstances of the case” that suggests the slightest chance—let alone a “genuine probability”—the ALJ would have concluded differently had he seen a full set of IQ test scores.

Even Miss Garcia’s own briefing does not attempt such an argument. In her opening brief, she emphasizes only that, deprived of a full battery of test scores, she lost the opportunity to qualify for automatic disability benefits under Listing 12.05 C or D, *see* 20 C.F.R. § 404, subpt. P, app. 1. She does not, however, attempt affirmatively to link the incomplete IQ tests with the medical reports and the ALJ’s determination of her residual functional capacity. Only in her supplemental brief does Miss Garcia clearly assert such a connection—and, even there, she does not offer any reason why we may expect the medical experts would have

substantively revised their reports in light of complete test results.

The majority assures us, however, that an alternative finding by the ALJ “seems particularly plausible” based on Miss Garcia’s “considerably lower” test results as a juvenile. Maj. op. at 16–17. But this is a *non sequitur*. The ALJ determined Miss Garcia not to be disabled in light of her record as a whole: he did not explain that the partial IQ test score carried dispositive weight. Nothing in the record to which either Miss Garcia or the majority point suggests a necessary connection between marginally lower IQ scores and a RFC finding that would prevent her from procuring and performing gainful employment. This “genuine probability” of a different outcome that the majority identifies, accordingly, appears little more than an unsubstantiated hunch.

In addition, Listings 12.05 C and D require not only a sufficiently low IQ test score, but also additional impairments, before the applicant may qualify for disability benefits thereunder. Miss Garcia does not, before this court, argue that she may have qualified under Listing 12.05 B, which she would satisfy simply by scoring below 60 on any of her tests without presenting any other additional impairments.² Nevertheless, the majority, pointing to her substantially lower testing results as a juvenile, predicts that Miss Garcia may have scored low enough to qualify as disabled under Listing 12.05 B. For such reason, the majority finds prejudice in Dr. McDonald’s failure to administer the

² In her opening brief, Miss Garcia specifically argued that “a valid IQ score on one of the two missing IQ tests may provide satisfaction of the Listing at § 12.05(C) or (D).”

entire battery of IQ tests and in the ALJ's acceptance of these partial scores. In effect, this reasoning says—bizarrely—that Miss Garcia wins an argument she does not make. Since she *never* claimed on appeal that she would have qualified under Listing 12.05 B, the *possibility* that she could have so qualified should not be a grounds that she suffered prejudice.

III

The majority's reasoning, furthermore, threatens to undermine the highly deferential standard under which we review the Commissioner's decisions. When presented with an appeal from an unsuccessful applicant, we may not second-guess the Commissioner's determination or reverse him simply because we disagree with the result. Our authority to order relief is more limited: if substantial evidence exists in the record to support the agency's fact-bound conclusions, our analysis must generally come to an end. Here the majority opinion does not suggest an absence of substantial evidence to ballast the ALJ's nondisability finding; rather, it *posits* that, despite any such substantial evidence, the ALJ *might* have reached an alternative conclusion if the record had contained a full set of IQ scores.

Such holding opens a potentially fatal breach in the substantial-evidence framework. Indeed, the majority determines that the ALJ committed legal error by not developing the record to include a full set of test scores; and, indeed, "legal error" is a basis distinct from the lack of substantial evidence for reversal. Nevertheless the relationship between these two standards, in the context of the ALJ's legal duty to develop the record, should be apparent enough. Claimants previously required to disprove the existence of substantial evidence will now plead an

incomplete record and, citing the majority opinion, will assert that the outcome of their case “might have been different,” maj. op. at 17. Seldom will be the occasion where the ALJ could not have examined more reports or ordered more tests. In *Mayer*, we specifically rejected a challenge from a claimant who contended, in effect, that substantial evidence did not support the ALJ’s denial *because* he did not adequately develop the record. 276 F.3d at 459. The substantial-evidence standard protects against precisely such attacks on the administrative process: the courts may not overturn the agency’s findings, substantiated by sufficient data, even in the presence of compelling countervailing evidence. Claimants ought not be able to circumvent this standard by invoking *hypothetical* evidence that the ALJ could have but neglected for one reason or another to consider. *Id.* Our procedure, elucidated in *McLeod*, for assessing the prejudice caused by an inadequately developed record reinforces these principles. The ALJ’s duty to develop “is triggered only” in certain circumstances, *Mayer*, 276 F.3d at 459, and, unlike other contexts, we do *not* presume prejudice until the claimant or the record demonstrates otherwise, *see McLeod*, 640 F.3d at 887–88.

The majority’s doctrinal innovation destabilizes this framework, substantially lowering the burden for plaintiffs seeking the intervention of the Federal courts in the Commissioner’s decision-making processes and portending to make substantial-evidence review a dead letter. Such result contravenes the precedents of this Court, the intent of Congress, and the separation of powers.

IV

For the foregoing reasons, I respectfully dissent.