JWST/NIRSpec Measurements of the Relationships between Nebular Emission-line Ratios and Stellar Mass at z 3–6

Shapley, Alice E.; Reddy, Naveen A.; Sanders, Ryan L.; Topping, Michael W.; Brammer, Gabriel B.

Published in:
Astrophysics Journal Letters

DOI:
10.3847/2041-8213/acd939

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
JWST/NIRSpec Measurements of the Relationships between Nebular Emission-line Ratios and Stellar Mass at $z \sim 3–6$

Alice E. Shapley\textsuperscript{1}, Naveen A. Reddy\textsuperscript{2}, Ryan L. Sanders\textsuperscript{3,7}, Michael W. Topping\textsuperscript{4}, and Gabriel B. Brammer\textsuperscript{5,6} \textsuperscript{7} NHFP Hubble Fellow.

\textsuperscript{1} Department of Physics & Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA; aes@astro.ucla.edu
\textsuperscript{2} Department of Physics & Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
\textsuperscript{3} Department of Physics and Astronomy, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA
\textsuperscript{4} Steward Observatory, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA
\textsuperscript{5} Cosmic Dawn Center (DAWN), Denmark
\textsuperscript{6} Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen Ø, Denmark
\textsuperscript{7} CEERS Early Release Science program

\textit{Abstract}

We analyze the rest-optical emission-line ratios of star-forming galaxies at $2.7 \lesssim z < 6.5$ drawn from the Cosmic Evolution Early Release Science (CEERS) Survey and their relationships with stellar mass ($M_*$). Our analysis includes both line ratios based on the [N\textsc{ii}] $\lambda 6583$ feature ([N\textsc{ii}] $\lambda 6583$/$H\alpha$, ([O\textsc{iii}] $\lambda 5007$/$H\beta$)/([N\textsc{ii}] $\lambda 6583$/$H\alpha$) (O3N2), and [N\textsc{ii}] $\lambda 6583$/[O\textsc{ii}] $\lambda 3727$) and those featuring $\alpha$-elements ([O\textsc{iii}] $\lambda 5007$/$H\beta$, [O\textsc{iii}] $\lambda 5007$/[O\textsc{ii}] $\lambda 3727$ (O32), ([O\textsc{iii}] $\lambda 3727$), and [Ne\textsc{iii}] $\lambda 3869$/[O\textsc{ii}] $\lambda 3727$). Given the typical flux levels of [N\textsc{ii}] $\lambda 6583$ and [Ne\textsc{iii}] $\lambda 3869$, which are undetected in the majority of individual CEERS galaxies at $2.7 \lesssim z < 6.5$, we construct composite spectra in bins of $M_*$ and redshift. Using these composite spectra, we compare the relationships between emission-line ratios and $M_*$ at $2.7 \lesssim z < 6.5$ with those observed at lower redshift. While there is a significant evolution toward higher excitation (e.g., higher [O\textsc{iii}] $\lambda 5007$/$H\beta$, O32, O3N2) and weaker nitrogen emission (e.g., lower [N\textsc{ii}] $\lambda 6583$/$H\alpha$ and [N\textsc{ii}] $\lambda 6583$/[O\textsc{ii}] $\lambda 3727$) between $z \sim 0$ and $z \sim 3$, we find in most cases that there is no significant evolution in the relationship between line ratio and $M_*$ beyond $z \sim 3$. The [Ne\textsc{iii}] $\lambda 3869$/[O\textsc{ii}] $\lambda 3727$ ratio is anomalous in showing evidence for significant evolution at $4.0 \lesssim z < 6.5$ at fixed mass, relative to $z \sim 3.3$. Collectively, however, our empirical results suggest no significant evolution in the mass–metallicity relationship at $2.7 \lesssim z < 6.5$. Representative galaxy samples and metallicity calibrations based on existing and upcoming JWST/NIRSpec observations will be required to translate these empirical scaling relations into ones tracing chemical enrichment and gas cycling and to distinguish among descriptions of feedback in galaxy formation simulations at $z > 3$.

\textit{Unified Astronomy Thesaurus concepts:} Galaxy formation (595); Galaxy evolution (594); High-redshift galaxies (734)

\section{1. Introduction}

The rest-frame optical nebular emission-line spectrum of star-forming galaxies is rich with information probing their gas, heavy elements, dust, and massive stars. The pattern of rest-optical emission lines is also known to vary systematically as a function of stellar mass ($M_*$). This variation occurs primarily as a result of the dependence of galaxy metallicity on galaxy stellar mass. This so-called “mass–metallicity relation” (MZR) has been traced using vast samples in the local universe drawn from the Sloan Digital Sky Survey (SDSS; e.g., Tremonti et al. 2004; Andrews & Martini 2013) and with smaller, yet still statistical power, all the way out to $z \sim 3$ (e.g., Onodera et al. 2016; Kashino et al. 2017; Sanders et al. 2021; Topping et al. 2021). A secondary, additional dependence of metallicity on star formation rate (SFR) has been discovered as well (e.g., Ellison et al. 2008; Mannucci et al. 2010), the “fundamental metallicity relation” (FMR), which appears not to evolve significantly between $z \sim 0$ and $z \sim 3$ (Sanders et al. 2021).

Previously restricted to $z \lesssim 3.5$, studies of the rest-optical emission-line properties of star-forming galaxies can now be extended deep into the reionization epoch with JWST. Even the very first NIRSpec spectroscopic data released to the public revealed the promise of JWST for estimating the gas-phase chemical abundances of star-forming galaxies at $z \sim 7.5–8.5$ (e.g., Arelanno-Córdova et al. 2022; Curti et al. 2023). Subsequent remarkable NIRSpec spectra have revealed the rest-optical nebular properties of individual galaxies at $z = 9.5$ (Williams et al. 2023) and, now, $z = 10.6$ (Bunker et al. 2023). The low metallicities of the $z > 8$ targets observed by JWST, given their stellar masses and SFRs, suggest that they do not fall on the FMR that describes galaxies at $z \lesssim 3$ (Langeroodi et al. 2022; Williams et al. 2023; Curti et al. 2023). However, it is challenging to draw broad conclusions from such a small sample, which may or may not be representative of the full population at these redshifts, and questions remain about which calibrations can be used to accurately infer metallicity at such high redshifts.

Larger samples are clearly needed to place detailed, single-object results in context. The CEERS Early Release Science program (Finkelstein et al. 2022, 2023; S. Finkelstein et al. 2023, in preparation) provides one of the first such opportunities (see also Matthee et al. 2022; Cameron et al. 2023; Mascia et al. 2023), including six NIRSpec pointings in the EGS field using the medium-resolution ($R \sim 1000$) grating and...
six pointings using the $R \sim 100$ prism, targeting collectively $\sim 1000$ galaxies with photometric redshifts over the range $z \sim 0.5$–12 (Fujimoto et al. 2023). Already, CEERS is providing a window into the excitation properties of star-forming galaxies out to $z \sim 9$ (Sanders et al. 2023a; Tang et al. 2023). Even $z \sim 3.5$–6.5 represents uncharted territory for rest-optical spectroscopic emission-line studies of star-forming galaxies. Furthermore, JWST/NIRSpec provides access to a broader complement of rest-optical emission lines at $z \sim 3.5$–6.5 than those covered for the most distant galaxies with JWST spectra (Shapley et al. 2023). Based on its richness, with simultaneous coverage of [O II] $\lambda 3727$, [Ne III] $\lambda 3869$, H$\beta$, [O III] $\lambda 4959$, 5007, H$\alpha$, and [N II] $\lambda 6583$ in CEERS/NIRSpec medium-resolution spectra, we focus in this work on the redshift range $z = 2.7$–6.5. Robust stellar mass estimates for galaxies in this redshift range enable us to analyze the relationships among rest-optical emission-line ratios and stellar masses. Tracing these relationships for both $\alpha$-elements (e.g., O, Ne) and nitrogen is a first crucial step toward establishing metallicity scaling relationships at $z > 3$, which also requires a robust calibration between emission-line properties and oxygen abundance (Bian et al. 2018; Sanders et al. 2020, 2021; Nakajima et al. 2023).

In Section 2, we describe observations and samples analyzed here. In Section 3, we present results on the observed relationships between rest-optical emission-line ratios and stellar mass, measured continuously from $z \sim 3$ to 6. In Section 4, we compare with other recent work and consider the implications of these new measurements for the evolution of metallicity scaling relations among galaxies. Throughout we adopt cosmological parameters of $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$ and a Chabrier (2003) initial mass function.

## 2. Observations, Sample, and Measurements

### 2.1. Observations

Our analysis is based on the public NIRSpec Micro-Shutter Assembly (MSA) data from the CEERS program (Program ID:1345; Finkelstein et al. 2022, 2023; P. Arrabal Haro et al. 2023, in preparation). We analyze data from six medium-resolution ($R \sim 1000$) NIRSpec pointings in the AEGIS field, spanning 1–5 $\mu$m with the grating/filter combination of G140M/F100LP, G235M/F170LP, and G395M/F290LP. These six pointings included a total sample of 318 distinct targets, observed for 3107 s in each grating/filter combination. Two-dimensional (2D) data processing steps, one-dimensional extraction, slit-loss corrections, band-to-band flux calibration, and the measurement of emission-line fluxes are all described in detail in Shapley et al. (2023), Sanders et al. (2023a), and Reddy et al. (2023).

We extracted spectra for 252 CEERS targets and measured spectroscopic redshifts for 231. As described in Shapley et al. (2023), we modeled existing Hubble Space Telescope (HST)/ACS and WFC3, JWST/NIRCam, Spitzer/IRAC, and ground-based photometry for CEERS galaxy targets using the FAST program (Kriek et al. 2009), assuming the stellar population synthesis models of Conroy et al. (2009). We adopted delayed-$\tau$ star formation histories, where SFR$(t) \propto t \times \exp(-t/\tau)$, with $\tau$ the time since the onset of star formation. Robust spectral energy distributions (SEDs), corrected for emission-line fluxes, were determined for 210 of the 231 galaxies with spectroscopic redshifts, for which we accordingly obtained stellar mass estimates.

### 2.2. Sample

We restricted the current analysis to CEERS galaxies at $2.7 \leq z < 6.5$, which spans in redshift from the current ground-based high-redshift limit for $H_\alpha$ measurements up to the limit at which $H_\alpha$ and [N II] $\lambda 6583$ can be measured with the NIRSpec G395M/F290LP setup. We also require an estimate of stellar mass and a lack of spectroscopic indication of active galactic nucleus (AGN) activity, such as log([N II] $\lambda 6583$/H$\alpha$) $\gtrsim -0.3$ or broad $H_\alpha$ emission. These criteria yield a sample of 94 galaxies. In order to search for redshift evolution, we construct three redshift subsamples at $2.7 \leq z < 4$ (27 galaxies), $4.0 \leq z < 5.0$ (32 galaxies), and $5.0 \leq z < 6.5$ (35 galaxies). In Figure 1, we plot the stellar masses and redshifts for our sample, color-coded according to redshift. In Shapley et al. (2023), we showed that the $2.7 \leq z < 4.0$ and $4.0 \leq z < 5.0$ CEERS samples are representative of main-sequence star-forming galaxies (Speagle et al. 2014), whereas the $5.0 \leq z < 6.5$ may represent galaxies with higher-than-average specific SFRs.

### 2.3. Measurements

Based on the emission-line measurements from CEERS galaxy spectra, we estimated several line ratios for each galaxy, which form the basis of our analysis. These line ratios include the following:

1. [N II] $\lambda 6583$/H$\alpha$;
2. ([O III] $\lambda 5007$)/([N II] $\lambda 6583$/H$\alpha$) (hereafter O3N2);
3. [N II] $\lambda 6583$/[O II] $\lambda 3727$;
4. [O III] $\times 5007$/H$\beta$;
5. [O III] $\lambda 5007$/[O II] $\lambda 3727$ (hereafter O32);
6. ([O III] $\lambda 5007 + [O II]$ $\lambda 3727$/H$\beta$ (hereafter R23); and
7. [Ne III] $\lambda 3869$/[O II] $\lambda 3727$. 

---

**Figure 1.** Stellar mass vs. redshift for all 94 CEERS galaxies at $2.7 \leq z \leq 6.5$ analyzed in this work. Galaxies at $z = 2.7$–4.0 are color-coded green; those at $z = 4.0$–5.0 are shown in blue; finally, those at $z = 5.0$–6.5 are indicated in magenta.
Ratios between lines with small wavelength separation (i.e., $[\text{N II}] \lambda 6583/\text{H}\alpha$, $[\text{O III}] \lambda 5007/\text{H}\beta$, $[\text{Ne III}] \lambda 3869/\text{O II} \lambda 3727$) do not require corrections for dust attenuation and are presented as is, while $[\text{N II}] \lambda 6583/\text{O II} \lambda 3727$, $\text{O}3\alpha$, and $R23$ are corrected for dust attenuation based on the $\text{H}\alpha/\text{H}\beta$ Balmer decrement, assuming an intrinsic $\text{H}\alpha/\text{H}\beta$ ratio of 2.79 and the Cardelli et al. (1989) dust law, which has been shown to be appropriate for high-redshift star-forming galaxies (Reddy et al. 2020).

While the majority of our sample has individual detections of $[\text{O III}] \lambda 5007/\text{H}\beta$, $\text{O}3\alpha$, and $R23$, any line ratio including either $[\text{N II}]$ or $[\text{Ne III}]$ results in a majority of limits in either one or more required emission lines. In order to utilize the full sample, including limits, and obtain results that are representative of the relationships between emission-line ratios and stellar mass, we constructed composite spectra in two roughly equal-sized bins of stellar mass for each of the three redshift bins. The 82 (out of 94) CEERS galaxies at $2.7 \leq z < 6.5$ with detections of $\text{H}\alpha$ were included in the sample for stacking. $\text{H}\alpha$ detections were required since individual spectra were normalized to a common $\text{H}\alpha$ luminosity before averaging. Emission-line fluxes, ratios, and uncertainties were measured from each composite spectrum (Reddy et al. 2023). Figure 2 shows an example of a CEERS composite spectrum, representing the high-mass bin within the redshift range $z = 4.0–5.0$. All key emission lines used in our analysis are labeled.

3. Results

We present the relations between emission-line ratios and stellar mass in each of the three CEERS redshift bins spanning $2.7 \leq z < 6.5$. These values are reported in Table 1. We plot individual CEERS galaxies when a detection or meaningful $3\sigma$ upper or lower limit can be derived. Given the low detection rate for some line ratios, we also plot values measured from composite spectra in two bins of stellar mass for each redshift bin. For comparison, we indicate the analogous relations for lower-redshift samples. Specifically, in all panels we plot the distribution of SDSS $z \sim 0$ star-forming galaxies as a gray 2D histogram (Abazajian et al. 2009). We also plot measurements from stacked spectra of star-forming galaxies drawn from the MOSDEF survey. These include $z \sim 1.5$ measurements from Topping et al. (2021) and $z \sim 2.3$ and $z \sim 3.3$ measurements from Sanders et al. (2021). The $z \sim 2.3$ measurements span $[\text{O II}] \lambda 3727$, $[\text{Ne III}] \lambda 3869$, $\text{H}\beta$, $[\text{O III}] \lambda 5007$, $\text{H}\alpha$, and $[\text{N II}] \lambda 6583$, whereas the $z \sim 1.5$ measurements lack the bluest
features ([O II] $\lambda 3727$, [Ne III] $\lambda 3869$); conversely, the $z \sim 3.3$ measurements lack the reddest features (H$\alpha$ and [N II] $\lambda 6583$).

3.1. Line Ratios Including Nitrogen

We begin with scaling relations including the [N II] $\lambda 6583$ line: [N II] $\lambda 6583$/H$\alpha$, O3N2, and [N II] $\lambda 6583$/[O II] $\lambda 3727$. One basic result concerns the relative faintness of [N II] $\lambda 6583$. Out of 82 galaxies with coverage of both [N II] $\lambda 6583$ and H$\alpha$ at $z = 2.7$–6.5, only 31 have 3$\sigma$ detections of [N II] $\lambda 6583$. The CEERS composite spectra show that the typical ratio of [N II] $\lambda 6583$/H$\alpha$ in this redshift range is 0.04–0.06, except at log($M_*/M_\odot$) $\geq$ 10, where it rises to 0.10–0.15 in the lowest-redshift bin. Indeed, based on both individual data points and stacked measurements, there is evidence in Figure 3 (left) for a scaling between [N II] $\lambda 6583$/H$\alpha$ and $M_*$ within the CEERS $z = 2.7$–4.0 bin, which is consistent with the scaling observed by Sanders et al. (2021) for $z \sim 2.3$ star-forming galaxies in the MOSDEF survey. On the other hand, MOSDEF star-forming galaxies at $z \sim 1.5$ show a progression toward higher [N II] $\lambda 6583$/H$\alpha$ at fixed stellar mass, which continues to $z \sim 0$. At $z > 4.0$, no strong trends are recovered between [N II] $\lambda 6583$/H$\alpha$ and stellar mass. In addition, the stellar mass range probed at $z > 5.0$ extends toward significantly lower values (log($M_*/M_\odot$) $\sim$ 8.0). At this stellar mass, given the existence of the MZR, nitrogen production falls within the primary regime, as discussed below (and see Figure 14 of Andrews & Martini 2013).

O3N2 (Figure 3, middle) and [N II] $\lambda 6583$/[O II] $\lambda 3727$ (Figure 3, right) tell a similar story, with the CEERS $z = 2.7$–4.0 sample overlapping the MOSDEF $z \sim 2.3$ scaling relations and no strong trends within the CEERS $z > 4.0$ galaxies. One possible reason for the lack of strong scaling between [N II] $\lambda 6583$/[O II] $\lambda 3727$ and mass at $z > 4.0$ is that we are probing the low-metallicity regime of primary nitrogen
production, where the N/O ratio is independent of metallicity (e.g., Pilyugin et al. 2012). Along these lines, in the panel displaying [N II] $\lambda 6583$/[O II] $\lambda 3727$ versus $M_*$ (Figure 3, right), we also plot low-mass galaxies from the Local Volume Legacy (LVL) survey (Berg et al. 2012) that span the mass range probed by the $z = 5.0–6.5$ CEERS sample. At \( M_*/M_* \leq 9.0 \) and the corresponding $12 + \log(O/H)$ values, there is no scaling between N/O and $M_*$ in the $z = 0$ dwarf galaxies. Given the evolution of the MZR toward lower metallicity at fixed mass as redshift increases (e.g., Steidel et al. 2014; Kashino et al. 2017; Sanders et al. 2021), the CEERS $z = 5.0–6.5$ sample must represent an even lower average metallicity at the same mass and fall even more firmly within the primary nitrogen regime. More generally, if the $z > 4$ CEERS sample is indeed in the primary nitrogen regime while the $z < 3$ samples are not, this difference at least partially explains why [N II] $\lambda 6583$/H$\alpha$ and O3N2 have a flatter dependence on $M_*$ in the $z > 4$ samples. However, such an explanation may not apply to the high-mass $z = 4.0–5.0$ bin, which is still significantly lower (higher) in [N II] $\lambda 6583$/H$\alpha$ (O3N2) than an interpolation of the $z = 2.7–4.0$ CEERS sample at the same $M_*$.

3.2. Line Ratios Including $\alpha$-elements

We also consider several commonly studied ratios based on $\alpha$-elements (e.g., O, Ne): [O III] $\lambda 5007$/H$\beta$, O32, R23, and [Ne III] $\lambda 3869$/[O II] $\lambda 3727$. The majority of galaxies in the CEERS $z = 2.7–6.5$ sample are detected in [O III] $\lambda 5007$/H$\beta$, O32, and R23, although individual detections of the fainter [Ne III] $\lambda 3869$ line are achieved for only $\sim 40\%$ of the sample. As in earlier works (e.g., Juneau et al. 2014; Coil et al. 2015; Holden et al. 2016), we find an elevated [O III] $\lambda 5007$/H$\beta$ at fixed stellar mass, relative to the local relation (Figure 4, top left). We also find a lack of significant evolution in [O III] $\lambda 5007$/H$\beta$ at fixed stellar mass beyond $z \sim 3$. Specifically (with the exception of the high-mass bin at $z = 4.0–5.0$), the CEERS emission-line measurements at $z > 2.7$ are consistent with those from the MOSDEF $z \sim 3$ sample where they overlap in stellar mass.

While the bulk of $z > 1$ measurements shown here follow a trend of increasing [O III] $\lambda 5007$/H$\beta$ as stellar mass decreases, at the lowest masses probed by individual galaxies in the CEERS $z = 5.0–6.5$ sample we find evidence of a turnover toward lower [O III] $\lambda 5007$/H$\beta$. Specifically, several galaxies at the low-mass end of the CEERS sample fall significantly below the SDSS distribution in [O III] $\lambda 5007$/H$\beta$ versus $M_*$, whereas above $\log(M_*/M_*) \sim 9.0$ we find only one such galaxy. The turnover is also suggested by the trends in the stacked data points from both CEERS and MOSDEF. Together, the CEERS and MOSDEF stacked data points trace a monotonically decreasing sequence in [O III] $\lambda 5007$/H$\beta$ as mass increases above $\log(M_*/M_*) \sim 9.0$. The CEERS low-mass $z = 5.0–6.5$ stack is characterized by a significantly lower mass (log($M_*/M_*) = 7.98$), and yet its [O III] $\lambda 5007$/H$\beta$ value is identical to that of the CEERS low-mass stacks at $z = 2.7–4.0$ and $z = 4.0–5.0$. Such a turnover toward decreasing [O III] $\lambda 5007$/H$\beta$ with decreasing mass at stellar masses below $\log(M_*/M_*) \sim 9.0$ is expected, given the relationship between [O III] $\lambda 5007$/H$\beta$ and $12 + \log(O/H)$ in this regime (Curti et al. 2020; Sanders et al. 2020; Nakajima et al. 2022). This turnover in [O III]/H$\beta$ at low stellar mass is also detected by Mathee et al. (2022) in JWST/NIRCam grism observations of a sample of $z \sim 6$ [O III]-emitting galaxies.

Like [O III] $\lambda 5007$/H$\beta$, the ratio O32 shows no significant evolution at $z \sim 3$ and beyond (Figure 4, top right). O32 is most directly sensitive to ionization parameter and, through the anticorrelation between ionization parameter and oxygen abundance (Pérez-Montero 2014), indirectly traces metallicity (Sanders et al. 2016, 2018). As in the case of [O III] $\lambda 5007$/H$\beta$, we find that CEERS galaxies at $z = 2.7–6.5$ generally have similar O32 to those in the $z \sim 3.3$ MOSDEF sample at fixed stellar mass.

As in the [N II] $\lambda 6583$/H$\alpha$ and O3N2 diagrams, the high-mass $z = 4.0–5.0$ bin presents as an outlier toward significantly higher [O III] $\lambda 5007$/H$\beta$ and O32 at fixed stellar mass relative to the lower-redshift samples (there is no $z = 5.0–6.5$ data point in a similar mass range). We require a larger sample, deeper spectroscopy, and robustly calibrated direct metallicity measurements at $z = 4.0–5.0$ (Sanders et al. 2023b) to determine whether this offset is truly representative of star-forming galaxies at this redshift and reflective of higher-excitation physical conditions or else individually undetected AGN activity in some sources.

In the space of R23 versus stellar mass (Figure 4, bottom left), galaxies at $z \geq 2.3$ are offset from the local sequence toward higher R23 at fixed stellar mass, yet there is no evolution from $z \sim 2.3$ to $z = 6.5$ given that the MOSDEF $z \sim 2.3$ and $z \sim 3.3$ stacks follow the same relation. This behavior differs slightly from what is observed for [O III] $\lambda 5007$/H$\beta$ (and O32), where there is slight evolution toward higher line ratio ($\sim 0.1–0.2$ dex) at fixed mass between $z \sim 2.3$ and $z \sim 3.3$, and then no further evolution at higher redshift.

Finally, we find that [Ne III] $\lambda 3869$/[O II] $\lambda 3727$ shows evidence of being significantly higher at fixed stellar mass for CEERS galaxies at $z \geq 4$ than MOSDEF galaxies at $z \sim 3.3$ (and $z \sim 2.3$). Where we can most robustly gauge this offset, i.e., at $\log(M_*/M_*) \sim 9.0$ and excluding the CEERS high-mass $z = 4.0–5.0$ bin, we find that the average offset in [Ne III] $\lambda 3869$/[O II] $\lambda 3727$ between CEERS $z \geq 4.0$ galaxies and MOSDEF $z \sim 3.3$ galaxies is $+0.2$ dex. At lower masses, there are no average measurements from the MOSDEF survey, so it is not possible to perform a similar comparison. We defer further interpretation of the behavior of [Ne III] $\lambda 3869$/[O II] $\lambda 3727$ at these low masses, as we require larger samples at both $z \geq 4$ and $z \sim 2–3$. We do note that, given the close proximity of the constituent lines in the [Ne III] $\lambda 3869$/[O II] $\lambda 3727$ ratio, the estimate of [Ne III] $\lambda 3869$/[O II] $\lambda 3727$ is immune to the uncertainties in dust correction and wavelength-dependent flux calibration associated with line ratios widely spaced in wavelength such as O32 and R23. At the same time, the divergent behavior of O32 and [Ne III] $\lambda 3869$/[O II] $\lambda 3727$ as a function of redshift will require further investigation. Indeed, both ratios are sensitive to the ionization parameter, and both trace $\alpha$-elements (Strom et al. 2017; Witstok et al. 2021), so similar redshift evolution might have been expected. An elevated ratio of [Ne III] $\lambda 3869$ to [O III] $\lambda 5007$ at $z \sim 4$ may be indicative of a harder ionizing spectrum (Jeong et al. 2020) at fixed nebular metallicity, or higher abundance ratio of neon to oxygen. The former possibility, however, is inconsistent with the conclusions of Sanders et al. (2023a) that the ionization conditions do not strongly evolve in CEERS galaxies from $z \sim 2$ to $z \sim 6$. The latter possibility also appears unlikely given the similar enrichment channels of neon and oxygen and the
lack of significant variation in Ne/O detected in local star-forming galaxies as a function of either metallicity or specific SFR (Izotov et al. 2006).

4. Discussion

We have presented a systematic analysis of the empirical relationships between nebular emission-line ratios and stellar mass at \( z = 2.7\)–\( 6.5 \), based on both individual and stacked composite JWST/NIRSpec spectra from the CEERS program (Finkelstein et al. 2022, 2023). We analyze common rest-optical line ratios (but some observed for the very first time at \( z \geq 3 \)) based on nitrogen, \( \alpha \)-elements (oxygen, neon), and hydrogen. Though \([\text{N II}] \) \( \lambda 6583 \) proves challenging to detect in individual galaxies at \( z \geq 3 \), we find, based on stacked spectra, that there is no strong evolution in the scaling of \([\text{N II}] \) \( \lambda 6583 / \text{H} \beta \), \( \text{O} III \), or \([\text{N II}] \) \( \lambda 6583 / [\text{O II}] \) \( \lambda 3727 \) with stellar mass, relative to measurements from the MOSDEF survey at \( z \sim 3.3 \) (Sanders et al. 2021) are shown with large green triangles.
metallicity scaling relations (e.g., Tremonti et al. 2004; Sanders et al. 2021), and yet care must be taken when doing so. Specifically, robust calibrations are required for the translation between emission-line ratio and metallicity, and the calibrations adopted for local star-forming galaxies need to be updated for distant galaxies (e.g., Bian et al. 2018; Sanders et al. 2023b). We defer the actual translation between line ratio and metallicity to future work, but we here review some recent attempts to use JWST to infer distant galaxy chemical abundances and the reasons why such analyses are so important for constraining models of galaxy formation.

In addition to detailed investigations of remarkable individual NIRSpec spectra at high redshift (e.g., Williams et al. 2023; Bunker et al. 2023; Curti et al. 2023), JWST spectra are also starting to be used for analysis of the chemical abundances of samples of distant galaxies. Matthee et al. (2022) present NIRCam grism observations of [OIII]/Hβ in concert with stellar mass estimates for a sample of 117 [OIII]-emitting galaxies at 5.3 ≤ z ≤ 6.9. Metallicities are estimated from stacked spectra of galaxies in four bins of stellar mass using a single metallicity indicator, the [OIII]/Hβ ratio, because of the limited wavelength coverage of the NIRCam grism (i.e., 3–4 μm). It is worth noting both that there is very little variation in average [OIII]/Hβ across the four stellar mass bins, given that the galaxies in the Matthee et al. (2022) sample probe the peak of the [OIII]/Hβ distribution, and, further, that the metallicities inferred from [OIII]/Hβ are significantly higher at fixed mass than that inferred from a composite spectrum of z = 6.25–6.90 [OIII] emitters in which the auroral [OIII] λ4363 line is detected. It will be important to estimate metallicity for this redshift range based on multiple emission-line ratios like those featured in the current work, including those that vary monotonically with metallicity (e.g., O32, [Ne III] λ3869/[O III] λ3727, O3N2), and using calibrations based on direct metallicities at high redshift.

More directly related to the current work, Nakajima et al. (2023) attempt to constrain the MZR at z ~ 4–9 with JWST/NIRSpec observations of a sample of 135 galaxies drawn from the publicly available CEERS, GLASS, and ERO programs. Gas-phase metallicities are primarily based on R_{23} (78 galaxies) or [OIII] / Hβ (49 galaxies), with an additional 8 galaxies having direct T_e-based metallicities. Consistent with our empirical results, Nakajima et al. (2023) find only weak evolution in the MZR between z ~ 2–3 and z ~ 4–9. However, we caution that Nakajima et al. (2023) have SED-based stellar mass estimates for 81 galaxies in their JWST metallicity sample. The remaining 54 galaxies have stellar mass estimates (31 galaxies) or upper limits (23 galaxies) inferred from UV luminosities alone, which carry significant uncertainties due to variations in the mass-to-light ratio. Furthermore, additional work is needed to establish robust metallicity calibrations at z > 3 based on a larger sample of direct T_e-based metallicities at high redshift.

With proper calibration from direct metallicities (e.g., Curti et al. 2023; Sanders et al. 2023a), JWST promises to provide excellent observational constraints on the evolution of the MZR at z > 3. With a large enough sample and robust SFR estimates, it will also be possible to determine the nature of the FMR among oxygen abundance, M*, and SFR and whether it remains as invariant at higher redshift as it does from z ~ 0 to z ~ 3.3 (Sanders et al. 2021). In order to determine the evolution of the MZR, which is meant to represent the average properties of star-forming galaxies at different cosmic epochs, it will be necessary to assemble not only large but also representative galaxy samples. As discussed in Shapley et al. (2023), the CEERS galaxies we analyze here appear representative of the star-forming main sequence at z = 2.7–5.0, but with higher-than-average sSFR at z > 5.0. Given the connections among oxygen abundance, M*, and SFR, we also require star-forming-galaxy samples at z > 5.0 that are representative in SFR at fixed M*, if we aim to trace MZR evolution at z > 5.0.

Observational MZR constraints can be compared with theoretical predictions for the evolution of the MZR at the earliest times. For example, predictions in the literature for this evolution vary considerably. Both Torrey et al. (2019), based on the IllustrisTNG simulations, and Ma et al. (2016), based on the FIRE simulations, predict evolution of z ~ −0.3 dex in gas-phase metallicity at fixed stellar mass from z = 3 to z ~ 6 as galaxy gas fractions increase, though the normalization of the MZR is significantly higher in Torrey et al. (2019). Furthermore, Torrey et al. (2019) predict a linear decline in 12 + log(O/H) with increasing redshift, whereas Ma et al. (2016) predict a flattening in the evolution of the MZR with increasing redshift. In contrast, at slightly higher redshift (z = 5–8) and based on the FirstLight simulations, Langan et al. (2020) predict weak evolution in the normalization of the MZR. The evolution detected in this model is actually toward increasing metallicity with increasing redshift, since galaxy gas reservoirs are described to build up with time (decreasing redshift) over this interval, as gas accretion outpaces star formation and outflows. Furthermore, while the predicted MZR evolution from Ma et al. (2016) is also very shallow within this higher redshift range, the MZR normalization from FIRE is ~0.3 dex lower than in the FirstLight simulations.

Differences among galaxy formation model predictions arise because of how the processes of gas inflow, buildup, and outflow are described (Langan et al. 2020). Accordingly, robust measurements of the evolution of the MZR with JWST will be able to distinguish among the different prescriptions for baryon cycling at early times. In its first months, JWST has demonstrated that it is capable of returning the necessary data to trace the MZR at z > 3. We now need to assemble sufficient sample sizes (a factor of several larger than the CEERS/NIRSpec z > 3 sample) with both strong emission-line and stellar mass measurements and an adequate direct T_e-based metallicity calibration sample for translating emission-line ratios into metallicities.

Acknowledgments

We acknowledge the entire CEERS team for their effort to design and execute this Early Release Science observational program, especially the work to design the MSA observations. This work is based on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-03127 for JWST. The specific observations analyzed can be accessed via doi:10.17909/z7p0-8481. We also acknowledge support from NASA grant JWST-GO-01914. Support for this work was also provided through the NASA Hubble Fellowship grant #HST-HF2-51469.001-A awarded by the Space Telescope Science Institute, which is
operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555.

**ORCID iDs**

Alice E. Shapley https://orcid.org/0000-0003-3509-4855
Naveen A. Reddy https://orcid.org/0000-0001-9687-4973
Ryan L. Sanders https://orcid.org/0000-0003-4792-9119
Michael W. Topping https://orcid.org/0000-0001-8426-1141
Gabriel B. Brammer https://orcid.org/0000-0003-2680-005X

**References**

*ApJS*, 182, 543